



HAL
open science

Le Vivant Rare, Faible et Amorphe - Évolution depuis les Origines jusqu'à la Vie telle qu'elle nous Apparaît

Nicolas Glade

► **To cite this version:**

Nicolas Glade. Le Vivant Rare, Faible et Amorphe - Évolution depuis les Origines jusqu'à la Vie telle qu'elle nous Apparaît. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université Grenoble Alpes, 2022. tel-04075075

HAL Id: tel-04075075

<https://theses.hal.science/tel-04075075>

Submitted on 19 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

MÉMOIRE d'HABILITATION
à DIRIGER des RECHERCHES

à

l'École Doctorale d'Ingénierie pour la Santé et l'Environnement
Université Grenoble Alpes

Le Vivant Rare, Faible et Amorphe
Evolution depuis les Origines jusqu'à la Vie telle
qu'elle nous Apparaît

par

Nicolas GLADE

Manuscrit soumis le 12 avril 2022

Dernière mise à jour le 19 avril 2023

Soutenance le 8 novembre 2022

JURY

Patrick AMAR	MCU, Univ. Paris Saclay, HDR	Rapporteur
Pascal BALLEET	MCU, Univ. Bretagne Occidentale, HDR	Membre invité
Hughes BERRY	DR INRIA Lyon, HDR	Rapporteur
Fabien CRAUSTE	DR CNRS Paris, HDR	Examineur
Franck DELAPLACE	Pr., Univ. Evry val d'Essonne, HDR	Examineur
Francesca MERLIN	CR CNRS, Univ. Paris Sorbonne, HDR	Rapporteuse
Maël MONTEVIL	CR CNRS, ENS Paris	Membre invité
Victor NORRIS	Pr., Univ. Rouen, HDR	Examineur
Angélique STEPHANOU	CR CNRS, Grenoble-Alpes, HDR	Membre invité

Le mot question apparaît 202 fois dans ce document ; le mot réponse, 39 fois seulement !

Ça me paraît bien mieux dans ce sens que dans l'autre.

Le Vivant Rare, Faible et Amorphe

Evolution depuis les Origines jusqu'à la Vie telle qu'elle nous Apparaît

Résumé.

Les origines de la vie, la nature même de la matière vivante et son activité biologique, les représentations théoriques que nous en faisons pour la modéliser, l'appréhender, la contrôler, et les conséquences que l'usage de ces représentations ont sur notre façon de penser le vivant sont autant de considérations théoriques, méthodologiques et philosophiques que j'aborde dans ce manuscrit en m'appuyant sur mes recherches expérimentales et théoriques actuelles et passées concernant l'organisation du vivant, ses origines et son évolution.

Cette réflexion et les recherches effectives que nous menons, mes étudiants, collaborateurs et moi, se placent dans le cadre de l'émergence et la complexification de la vie. Les systèmes biologiques depuis les âges prébiotiques (ex : proto-métabolismes, proto-génétique) jusqu'à nos jours (ex : réseaux de régulation génétique, organisations cellulaires, subcellulaires, réseaux de neurones) ont évolué mais dépendent toutefois de systèmes existants. L'étude des propriétés structurelles et dynamiques (complexité, robustesse) mais également évolutives, de systèmes formels comme les réseaux Booléens ou autres systèmes de vie artificielle, et la recherche d'ensembles de systèmes aux propriétés structurelles ou dynamiques communes, permet d'introduire une vision ensembliste des systèmes biologiques et penser l'évolution ou les variations de fonctionnement d'un système biologique comme des trajectoires dans un paysage morphogénétique, un meta-réseau. Le passage, évolutif ou fonctionnel, d'un système à un autre est permis par leurs proximités structurelles et dynamiques. Notre travail utilise de tels systèmes formels en tant qu'abstractions des systèmes biologiques pour étudier (i) comment ils évoluent tout en préservant des caractères et comportements ancestraux, (ii) ce qui détermine l'évolutivité de ces réseaux, leur respective robustesse vis-à-vis de changements structuraux et leur relation avec la complexité structurelle et fonctionnelle, (iii) comment des trajectoires peuvent exister sous contraintes de viabilité dans de tels paysages morphogénétiques, et (iv) comment des systèmes et leurs comportements associés peuvent se combiner au cours de ces évolutions.

L'essai que je propose ici s'inscrit dans le cadre général de la biologie théorique et de la philosophie des sciences. J'y aborde des questions aussi variées que la méthode scientifique, l'humain dans les sciences, l'organisation et le fonctionnement de la vie, la nature des modèles, le contrôle, la finalité du vivant. J'y réfléchis, sur la base des recherches effectives que nous menons, à comment et pourquoi s'éloigner de la pensée mécaniste pour considérer davantage un vivant dont la nature et l'activité, depuis ses origines jusqu'au vivant actuel, sont davantage fondées sur ce que je désigne par *rare*, *faible* et *amorphe*, des qualificatifs qui s'éloignent de la conception que nous en avons, le *vivant-machine*, et des outils dont nous disposons pour l'étude du monde biologique.

Mots clef. Nature du vivant, Vivant-Machine, Origines du vivant, Evolution, Réseaux Booléens, Philosophie des Sciences, Biologie théorique.

Rare, Weak and Amorphous Life

Evolution of Life from the Origins until Life as it Appears Nowadays

Abstract.

I address in this manuscript theoretical, methodological and philosophical considerations based on my current and past, experimental and theoretical research on the organisation of life, its origins and its evolution. Of particular interest is how the theoretical representations of the very nature of living matter and the origins of life have consequences when using modeling to understand and control living matter.

These reflection and research are carried out by my students, collaborators and myself in the context of the emergence and complexification of life. Biological systems from prebiotic ages (e.g. proto-metabolisms, proto-genetics) to the present day (e.g. gene regulation networks, cellular and subcellular organisations, neural networks) have evolved but continue to depend on current systems. The study of structural and dynamical properties (complexity, robustness) but also evolutionary properties of formal systems such as Boolean networks or other artificial life systems, and the search for sets of systems with common structural or dynamical properties, allows us to introduce a holistic vision of biological systems and to think of their evolution and variations in terms of trajectories in a morphogenetic landscape, *i.e.* a meta-network. An evolutionary or functional path from one system to another is allowed by their structural and dynamical proximities. In our work, using such formal systems as abstractions of biological systems, we study (i) how they evolve while preserving ancestral traits and behaviours, (ii) what determines the evolvability of these networks, their respective robustness to structural changes and their relation to structural and functional complexity, (iii) how trajectories can exist under viability constraints in such morphogenetic landscapes, and (iv) how systems and their associated behaviours can combine during these evolutions.

The essay I propose falls within the general framework of theoretical biology and the philosophy of sciences. I try to address questions as various as the scientific methodology, humans in science, the organisation of life, the nature of models, control, purpose in life and many others. While using machine-based formalisms such as Boolean networks, I aim at legitimate the need to move away from a mechanistic thinking and the usual conception of the biological world, *i.e. Life as a Machine*, and the associated theoretical and experimental tools we use to study it. I actually propose to consider an alternate view of the nature and functioning of living matter - from its origins to the present day - based more on what I call *the rare, the weak and the amorphous*.

Keywords. Nature of living matter, Life as a Machine, Origins of Life, Evolution, Boolean networks, Philosophy of Sciences, Theoretical Biology.

INDEX

Partie introductive	1 - 33
1. Préambule	2 - 7
2. Remerciements	8
3. Le rare, le faible, l'amorphe – Introduction.	9 - 33
I. Dossier Scientifique	34 - 59
4. <i>Curriculum vitae</i>	35 - 38
5. Enseignement, encadrement et animation scientifique	39 - 48
6. Production scientifique	49 - 59
II. Recherche	60 - 370
7. Travaux antérieurs	61 - 91
8. Travaux de recherche actuels	92 – 224
9. Discussions sur l'évolution, la vie ...	225 – 278
10. Conclusion	279 – 284
11. Rapports des évaluateurs et analyse <i>a posteriori</i>	285 – 313
12. Bibliographie	314 - 338
13. Annexes	339 – 370

PARTIE I

PARTIE INTRODUCTIVE

1. PREAMBULE – LA SCIENCE ... ÉMOI

Tu ne peux te figurer le plaisir du compositeur écrivant librement sous l'influence directe de sa seule volonté. Quand j'ai tracé la première accolade de ma partition où sont rangés en bataille mes instruments de différents grades, quand je songe à ce champ d'accords que les préjugés scolastiques ont conservé vierge jusqu'à présent et que, depuis mon émancipation, je regarde comme mon domaine, je m'élançais avec une sorte de fureur pour y fourrager.

Hector BERLIOZ - Lettre à sa sœur

Entre le moment où j'ai envisagé l'idée d'écrire ce document et la période où je m'y suis attelé (donc le moment où la rédaction commence vraiment), il a bien dû s'écouler 8 ans ! La raison : ce document est un conflit intérieur que j'ai difficilement cherché à résoudre : ***quel est mon positionnement actuel en science ?***

Derrière cette question, de nombreuses réponses, mais toutes dotées de forces de rappel extrêmement puissantes, moi écartelé au milieu. La page blanche en pire ! C'est une époque de ma vie apparentée au confluent bouillonnant de plusieurs sources tranquilles. L'esprit trop plein, tout le temps, d'idées souvent floues, interconnectées ; des idées obsédantes dont on se demande au bout d'un moment si elles sont scientifiquement cohérentes, défendables, voire si elles ont vraiment du sens ; des positionnements scientifiquement forts mais fluctuants qui ont mis du temps à trouver un certain équilibre ; un bilan personnel avec des compétences renforcées mais des manques académiques cruels, lourds parfois ; une volonté d'écrire finalement ... vite terrassée par la charge de travail (universitaire, il va sans dire) explosant à tout moment, par le coup bas administratif, la tâche insipide qui réduit à néant tout influx neuronal sensé qu'on a mis si longtemps à stabiliser ... bref, ce fut long et douloureux ! mais intellectuellement passionnant aussi ; cette Habilitation à Diriger des Recherches ... une vraie mauvaise excuse pour prendre son temps pour réfléchir un peu.

Et puis il y a eu les nombreux étudiants que j'ai eu en stage et que je vais tous citer ici, Abdessalem (M2 MOSIG et Thèse EDISCE Psychomotricité, 2010-2013), Anita (L3 Biologie, 2016), Bing Yi (Stage ingénieur Polytech 3I M1, 2013), Clément (Stage IRL ingénieur ENSIMAG, 2017), Hedi (M2 ENSI, Tunisie, et Thèse EDISCE Informatique, 2008-2011), Hessem (Thèse EDISCE Epidémiologie, 2012-2016), Justine (M2 MITI, 2020), Malik (M2 MOSIG, 2011), Marion (Stage ingénieur Polytech RICM4, 2017), Quoc Trung (M1 MOSIG, 2017), Rémi (M1 Informatique, 2018, et thèse 2019-2022), Roselyne (L3 Polytech TIS, 2012), Sundeep (L3 Biotechnologie, Inde, 2008), Urvan (M1/M2 Informatique, 2018), Zhuofan (Stage ingénieur électronique L3, 2013), Bashaer, Clément, Havva, Hakim, Mounir, Marie Djasma, Raphaël, Ryma, (tous les 8 élèves ingénieurs Polytech TIS4 confinés, 2020), Nicolas (L3 MIAGE, 2021), Florian (M1 Phelma, 2021), Remy (L3 ENS Paris, 2021), Vincent (L3 ENS Paris, 2022) tous en formation pour les métiers de la recherche ou d'ingénieur, mais aussi tous ceux qui se dirigeaient vers une formation technique (BTS SIO pour

la plupart) et dont plusieurs ont continué vers les métiers de l'ingénieur, Anthony, Arnaud, Aurélien, Axel, Bastien, Camille, Cyril, Dorian, Estelle, Jonathan, Katell, Maxime, Nicolas, Stéphane, Théo, Thibaut, Xavier et William, qui ont considérablement contribué à maintenir mon activité de développement logiciel à flot.

Tous ces étudiants du BTS informatique à la thèse, venant de plusieurs coins du globe, remplis de fraîcheur et de suffisamment de naïveté salutaires auxquels ils fallait, pour qu'ils puissent supporter l'entrée dans l'esprit particulier du chercheur et faire en sorte d'exprimer tout leur potentiel, régulièrement insuffler cette même étincelle, cette astuce et cette audace qui m'ont été transmises, mais aussi faire accepter, parfois avec difficulté, la rigueur, le cadre, l'effort qu'on m'a appris et que j'ai bien sûr accommodés à ma façon et surtout adaptés à chacun d'entre eux.

Ce sont ces étudiants qui, je le pense, ont le plus contribué à me former. Ils ont été le miroir nécessaire à ma survie en recherche, celui que j'ai besoin de regarder pour me maintenir scientifiquement actif, et souvent à mon épanouissement ; l'épanouissement est fort heureusement souvent réciproque. Ils m'ont empêché de naviguer toutes voiles dehors sans prendre le temps d'accoster pour découvrir une côte inconnue de moi, parfois en poussant plus loin dans les terres, et de risquer de me perdre en vieux loup solitaire dans un océan infini de recherches vide de sens, bercé par le doux tangage de la vie de laboratoire et le roulis de la vie universitaire.

Au fil des années passées à m'aider à tenir le cap dans le gros temps, à hisser les voiles, ils m'ont amené vers les quarantièmes rugissants, là où aucun navire ne s'arrête, où le vent est fort et porteur. La métaphore maritime est bonne : ils ont été mousses et équipiers. Ils ont appris, nous avons fait du chemin ensemble et nous nous sommes mutuellement rassurés devant l'inconnu qui se profilait, quand le doute nous soufflait d'arrêter et lorsque les écueils, gouffres et autres vortex menaçaient d'achever la lente dérive que le capitaine n'a pas senti.

Ils ont été l'élément déclencheur pour que j'écrive (enfin !) ce document. La combinaison de leurs forces a très doucement fini par donner du sens à mes interrogations, parfois à donner quelques éléments de réponse, en tout cas à m'aider à préciser mon propre positionnement. C'est donc à eux que ce document rend hommage.

Ce qui est en jeu, dans un tel document (pour prétendre à l'HDR), c'est la proposition d'un travail scientifique doté d'une profondeur suffisante, c'est la transmission d'un vécu scientifique et d'un savoir faire, une source de *stimuli* scientifiques, la promesse d'un *encadrement doctoral de qualité*.

Qu'est qu'un *encadrement doctoral « de qualité »* ? Tout dépend de ce qu'on entend par *qualité* bien entendu ! Ce serait présomptueux de qualifier de meilleur un encadrement ou un autre. Cela dépend de trop de facteurs liés à un contexte, un étudiant, un encadrant, sans parler de l'évaluation de cette 'qualité' ; parle-t-on de rendement en publications, de l'emploi après thèse, de la profondeur des idées soutenues et de leur impact plus dur à évaluer car s'inscrivant dans une temporalité longue ? En ce qui me concerne, cette notion de qualité se trouve plutôt associée au respect d'un certain nombre de marqueurs de la recherche et de l'encadrement, doublé de la perception propre

qu'auront l'étudiant d'abord, l'encadrant ensuite, de la recherche et de l'encadrement. Remarquons que je ne fais pas entrer le sujet de thèse dans ce critère de qualité. Si un sujet amenant à une réflexion approfondie sur la nature des objets étudiés et à un travelling arrière visant à penser ces objets dans quelque chose de plus global (vision holiste) est quelque chose qui m'attire davantage que l'approche réductionniste de l'étude d'un objet biologique, l'exercice de style qui consiste à appliquer au moins un des canons de la méthode scientifique, à commencer par les bonnes pratiques (lecture, recoupages, relectures, définition des hypothèses, rédaction ...), reste le même.

Quoi qu'il en soit, ma façon de procéder associe ce qui, dès la première lecture que j'en avais faite en 1999, m'avait fasciné chez Buffon à la lecture du chapitre *De la Manière d'Etudier & de Traiter l'Histoire Naturelle* dans son ouvrage *Histoire Naturelle Générale et Particulière* [Buffon, 1746], à savoir un certain empirisme naïf, le romantisme d'une époque d'exploration des sciences et des connaissances du monde encore naissante et de l'enthousiasme à observer le monde et à le raconter, à une bonne part de réflexion générale philosophique et épistémologique, tout ce qui trait à la démarche scientifique [dans le désordre Bachelard, 1989 ; Barrau, 2016 ; Cariou, 2019 ; Chalmer, 1987 ; Descartes, 1930 ; Kuhn, 2008 ; Feyerabend, 2005 ; Morin, 2017 ; Popper, 1991 ; Rousseau, 1971 ; Voltaire, 1757 ; Monod, 1970], toujours hanté pendant mes lectures par l'idée de trouver des éléments de réponses à la question récurrente que je me pose, à savoir la pertinence de s'interroger sur la nature même du vivant et sur les paradigmes biologiques en vigueur. Si je pense que l'interrogation épistémologique fine doit arriver tard dans le développement du chercheur, c'est-à-dire après plusieurs années de pratique de la recherche, pour ne pas freiner celle-ci justement, les questionnements d'ordre philosophique sur la méthodologie scientifique et sur le métier de chercheur pouvant rapidement bloquer dans sa créativité et son enthousiasme celui qui s'interroge, je pense cependant qu'intéresser progressivement le doctorant à ces questions, en particulier les questions déontologiques sur la méthodologie scientifique, est un bon moyen de lui ouvrir l'esprit à ce qui devrait être plus que de la technique et de la connaissance factuelle. Faire de la science c'est aussi s'interroger sur comment et pourquoi on la fait.

Enfin il faut ajouter à ça un minimum de pragmatisme nécessaire à la survie en milieu scientifique à notre époque. Commentons tout de suite ce dernier point, car je ne souhaite plus y revenir : cela doit se traduire par des réflexions, voire des contraintes, sur le sujet d'étude, la méthodologie et la valorisation. Il faut aussi doucement apprendre à évoluer dans l'écosystème de recherche. Il faut accepter le « lâcher prise » nécessaire entre la 2ème et la 3ème année, cette étape finale pendant laquelle le doctorant apprend à assumer son organisation du travail, sa méthodologie, le message qu'il veut défendre. C'est une étape délicate qui requiert, pour l'encadrant, de bien comprendre comment fonctionne son doctorant et d'accepter de perdre une partie du contrôle du travail fait pour que la thèse devienne celle de l'étudiant. Je trouve intéressante cette partie humaine de ce rôle d'encadrant, cette partie qui doit être menée tout en subtilité et qui implique l'encadrant dans la vie du thésard et le responsabilise quant à son avenir.

Le sujet proposé à un doctorant doit avoir une couleur dominante bien identifiable ; l'encadrement doit veiller à ne pas trop la faire déteindre. Fini l'interdisciplinarité affichée ?! De nos jours, on peut

faire des sciences de la complexité et de la science interdisciplinaire, voire transdisciplinaire, mais on défend sa thèse en mathématiques, en informatique, en physique *ou* en biologie. C'est à ce prix que le doctorant pourra espérer un avenir professionnel en recherche. L'interdisciplinarité, mieux la transdisciplinarité, cependant sont fécondes ! La recherche fonctionne beaucoup par analogies, mises en correspondance. C'est le cas quand on modélise : on n'applique pas un modèle comme une vérité d'un système d'intérêt, mais on opère une mise en correspondance entre des éléments factuels issus de l'observation, les plus factuels possibles dirons nous, et des éléments théoriques dont on « sait » qu'ils représentent acceptablement au moins une partie de la réalité observée et pensée. La transdisciplinarité, c'est-à-dire le fait de naviguer tout aussi bien dans la connaissance biologique, mathématique, physique, informatique, philosophique, favorise ces mises en correspondance et analogies. Elle est, selon ceux qui la défendent, moi compris, un excellent moteur pour produire des idées innovantes. Est-elle incompatible avec l'avenir du jeune chercheur ? Non, mais il faut être très prudent : il faut autoriser le travail interdisciplinaire, voire verser dans la transdisciplinarité, mais il faut une force de rappel forte vers une identification sans doute possible de l'appartenance à une discipline précise et reconnue ; cette force de rappel, c'est à l'encadrant de la jouer.

La méthodologie ? D'abord je voudrais faire l'éloge de la « lenteur », loin des jalons imposés par les projets sensés favoriser l'innovation et dynamiser la recherche, à commencer par beaucoup de lecture spécialisée, de culture scientifique, de recoupements, de notes et d'écriture. L'observation et la réalisation passent après ! Elles doivent en effet pouvoir se projeter sur une connaissance vaste qu'il faut d'abord accumuler. Le travail de chercheur n'est pas celui d'un ingénieur ! Pour ce dernier, c'est le pragmatisme qui doit prévaloir, tant que sa réalisation est conforme aux spécifications qu'il aura établies. Ce faisant, il pourra produire une réalisation qui satisfera ces exigences et répondra à un besoin. Or, de mon point de vue du moins, la thèse est la vraie première étape de l'apprentissage du métier de chercheur. Il faut donc apprendre à travailler comme un chercheur. Le chercheur est en quête de compréhension avant tout. Il cherche, observant son sujet, à en peindre un tableau le plus détaillé possible, vérifiant que chaque détail qu'il en a saisi n'est pas issu de sa subjectivité mais est bien un savoir objectif pouvant être partagé par tous. Pour cela il doit d'abord connaître ce que les autres ont cherché à établir. Il doit aussi trouver comment affiner cette image ou la remettre en question, c'est-à-dire qu'il doit trouver des moyens d'observation ou de réflexion qui lui permettront de proposer une image valide, un certain temps seulement, et potentiellement exploitable pendant ce temps, de son objet d'étude.

Mais le scientifique est aussi quelqu'un qui, sur la base d'intuitions fondées sur les nombreux recoupements qu'il fait à partir de ses connaissances, ou par analogies, peut proposer de nouveaux cadres conceptuels (ce que Kuhn nomme *paradigmes* et Lakatos *noyaux durs* [Chalmer, 1987 ; Kuhn, 2008 ; Barrau, 2016]) ou du moins des incursions vers de nouveaux cadres conceptuels possibles. Ce faisant, il prend un risque contrôlé. Il est aidé en cela par le processus d'induction (*NB* : que Popper réfutait cependant en tant que méthode scientifique) qui lui permet de proposer, sur la base d'une collection de faits et théories établis, de nouveaux concepts, voire théories ; ces propositions doivent tout de mêmes rester fortement contraintes par les observables et rapidement éprouvées. En cela, et c'est une particularité de la recherche interdisciplinaire, dans mon cas en

biologie computationnelle, l'usage de l'ordinateur et de la programmation permet à la fois l'induction et la validation. Induction par le biais de l'expérimentation numérique ; validation grâce à la simulation numérique et à la confrontation des résultats des simulations et des données observées. R. Abbott, en s'interrogeant sur la nature du calcul numérique (*computation*) et de sa relation avec les idées, décrit en effet bien les processus computationnels comme systèmes d'aide à la réflexion au même titre que les calepins dans lesquels nous compilons nos pensées [Abbott, 2006].

Apprendre à faire de la recherche, c'est notamment appréhender ces éléments méthodologiques et leurs implications.

Quant à la valorisation par la communication scientifique orale ou écrite, j'en suis un fervent défenseur et maintiens que ce doit être la qualité première du chercheur : les chercheurs ne travaillent pas pour innover, mais pour produire de la connaissance et la transmettre. Certaines aboutissent à des innovations. Quoi qu'il en soit, un chercheur énonce des faits et propose des interprétations du monde qu'il étudie. Il délivre ainsi à l'ensemble de la communauté scientifique (plus raisonnablement à une de ses sous-parties restreinte à une discipline, voire une thématique) un message. Quel que soit le vecteur (publication dans une revue, poster, communication orale), la délivrance efficace d'un message scientifique, parfois complexe, passe par un savoir faire quant à la structuration de ce message, structuration qui doit se répercuter sur la recherche menée. Il est donc indispensable pour moi qu'un doctorant, dès lors qu'il s'est muni d'un bagage bibliographique, entame rapidement la rédaction d'un article scientifique, afin que cela structure sa propre recherche. Pour former son esprit scientifique il faut, j'en suis convaincu, guider son travail par la rédaction d'articles et non faire l'inverse, à savoir voir les publications comme les produits finaux du processus d'étude. Cela est possible, mais sous la supervision forte, très cadrante, du directeur de thèse. Je n'adhère pas à ce dernier procédé qui n'aide pas le doctorant à s'affranchir de cette dépendance. Et puis, quoi de plus gratifiant pour un directeur de thèse que de voir son doctorant prendre rapidement son autonomie en recherche, l'autonomie de publication (certes supervisée) l'amenant à une assurance certaine quant au contrôle qu'il a sur sa recherche et de là favorisant un développement souvent fort intéressant de ses propres idées.

Si j'ai été bon publiant, je ne le suis très malheureusement plus autant qu'avant faute de temps, ma charge d'enseignement et surtout mes responsabilités universitaires lourdes mobilisant une bonne partie de mon esprit et de mon énergie. Pour autant, je n'ai pas oublié ni comment valoriser ainsi une recherche, ni comment apprendre à un étudiant à écrire un article de sciences. C'est ce que j'ai fait pour mes 3 doctorants passés (certains étant plus prolixes que d'autres), mais également pour mes derniers étudiants de M1 (Quoc Trung, Urvan et Rémi) et de L3 ENS/Centrale (Rémy, Florian, Vincent et Ai Ling).

Voilà ce qu'est ma perception de ce que doit être une direction de thèse, certes idéaliste pour certains, mais c'est celle que je me suis efforcé à respecter jusqu'à présent.

Pour le reste, je laisse le lecteur de ce document juge des choix scientifiques, tant thématiques qu'épistémologiques, que je propose et de l'esprit scientifique qui m'anime pour stimuler les futurs

doctorants, développer leur propre esprit scientifique et préparer au mieux leur transition vers le métier de chercheur.

2. REMERCIEMENTS

Remerciements aux membres du jury. Je tiens ici à grandement remercier les personnes qui ont accepté de lire ce long manuscrit (de s'inquiéter à ce stade de la lecture qu'il vont devoir faire), de m'écouter en parler et de contribuer par leurs réflexions à corriger et améliorer les miennes.

Remerciements personnels. Il y a beaucoup de personnes que je pourrais remercier pour l'ensemble des raisons qui m'ont menées à ce stade, à mes recherches, à la rédaction de ce manuscrit. Mais je n'en nommerai que peu par leur nom pour souligner mon sentiment à leur égard et l'importance de leur contribution dans ce qui m'anime à l'écriture de ces lignes.

Si comme je l'ai souligné précédemment, je dois beaucoup l'arrivée à ce point de ma carrière à mes étudiants, je le dois aussi à ma famille, en particulier à mon épouse, Sophie qui, je dois le dire, a les nerfs solides puisqu'elle, ainsi que mes filles Louise et Rose, supportent mes longues absences physiques, avec tout le temps que je passe à l'université et au labo, et mentales surtout, étant donné que, comme beaucoup de chercheurs probablement, « ça continue de mouliner là-haut pratiquement tout le temps ». Je suis reconnaissant aussi envers mes collègues de travail et autres collaborateurs proches avec qui j'ai toujours des échanges passionnants, sinon apaisants.

Je voudrais toutefois souligner l'insistance d'Angélique Stéphanou, Laurent Trilling, de Pascal et Nisma Ballet pour que j'écrive et soutienne mon HDR. Sans eux, pas d'HDR !

Je remercie aussi Angélique, Pascal et Laurent, Arnaud Chauvière pour les pauses salvatrices, ainsi qu'Ibrahim Cheddadi avec qui nous partageons très agréablement un bureau, mais aussi Sergiu Ivanov, pour les discussions scientifiques passionnantes que j'ai pu avoir et qui m'ont maintenues stimulés, mais aussi pour leur soutien.

Je voudrais particulièrement souligner le soutien très important que m'a apporté Laurent Trilling. Son enthousiasme et sa clairvoyance dans la recherche m'a été salutaire ! C'est lui qui m'a maintenu actif, me convainquant de prendre des étudiants chaque année, allant les chercher parfois lui-même. Laurent a aussi passé beaucoup de temps à relire ce document, à me faire de retours, à échanger sur ses contenus et, de façon générale, sur nos recherches. Si cette HDR devait avoir eu un tuteur, ce serait assurément lui. Merci Laurent !

Je termine par exprimer toute ma reconnaissance et mon affection pour Nisma Ballet, qui m'a toujours ouvert sa porte à Brest, m'a toujours soutenu, pas que pour cette HDR, et avec qui j'ai eu des discussions philosophiques passionnantes, hors laboratoire, relative aux humanités, au rapport de l'homme avec la Nature, la technique, sa perception subjective ou objective du monde, ses projections, ses besoins et ses aspirations, réflexions profondes qui m'ont considérablement guidées dans ma démarche de recherche. Que notre amitié soit longue !

3. INTRODUCTION - LE VIVANT, LE RARE, LE FAIBLE, L'AMORPHE

Qu'un esprit pût trouver ses joies les plus hautes dans des échappées originales, aux antipodes de la routine de tous les jours, et dans les combinaisons aussi frappantes que neuves de ces images que l'habitude et la lassitude, à force de les faire repasser dans le sillon ébréché et usé de la normale, ont dépouillé de tout élément vivant, voilà qui était impensable pour cet esprit clair, pratique et éminemment logique. Pour lui, toute chose, tout sentiment avait des dimensions, des propriétés, des causes et des effets bien déterminés ; et quoiqu'il eût vaguement conscience du fait que l'esprit parfois nourrit des visions et des sensations d'une nature beaucoup moins géométrique, classifiable et utilisable, il se croyait justifié à tracer une frontière arbitraire, et à tenir pour quantité négligeable tout ce qui ne peut être vécu et pleinement compris par l'homme de la rue. De plus, il était pratiquement certain que rien ne pouvait être vraiment *indicible*¹. Cela ne lui paraissait pas sérieux.

H. P. Lovecraft - L'indicible

3.1. Introduction : Le Vivant ... une machine ?

Le Vivant ... la matière vivante, cet amas dense, partiellement organisé de molécules parfois très complexes, pourvu d'une animation permanente et autonome, sans qu'aucune étincelle initiale ne fut nécessaire. Cette vie qu'on n'a pas besoin d'insuffler comme on démarre une machine, comme on initialise un programme informatique. C'est cet élément et qualificatif insaisissable qui frustre les penseurs depuis toujours. Sa définition est impossible tant elle est diverse, souvent vue comme complexe, mais parfois perçue à la limite de l'inanimé. Sa nature nous échappe donc et, pour l'appréhender, nous fondons notre compréhension sur ce que nous savons faire de mieux : les machines². Les machines automatiques à vapeur sont nées il y a plus de 2 siècles. Leur comportement s'est depuis complexifié, leur source d'énergie diversifiée mais, depuis le début, elles inspirent l'homme qui projette dedans ses désirs, ses pulsions, sa soif de contrôle. L'homme développe un affect pour ses machines comme s'il s'agissait d'êtres sensibles. S'il s'est beaucoup inspiré de la perception du vivant pour créer des machines, l'inverse est vrai aussi. Il est en effet plus facile de comprendre les choses vivantes comme des machines, ces compagnes utiles que l'on connaît bien maintenant.

Soyons affirmatif : l'être humain est ingénieur ! Il décortique toute chose de la nature en éléments, systèmes, processus et en tire des enseignements sur la composition et l'organisation de l'entité vivante. En complément pour sa compréhension, il recompose par l'esprit ou par le modèle ces éléments, cherche l'assemblage harmonieux, celui qu'il pense avoir observé avec ses instruments (ces instruments qui mesurent et perçoivent ce que l'on pense qu'il faut observer et qui ne perçoivent

1 Larousse : Qu'on ne saurait exprimer, qui dépasse toute expression.

2. Remarque : Nous associerons le terme 'machine' à toute réalisation issue d'une quelconque ingénierie, pas forcément humaine, consommatrice d'énergie, dotée d'un niveau d'organisation important et dotée d'une fonction, à savoir les machines mécaniques, électroniques, comme les programmes informatiques.

que ce qui peut être perçu) et qu'il a complété par inférence sur ce qu'il pense, sur ce que les autres pensent (sur ce que l'humanité a fini par doucement admettre) être du vivant et ... fabrique ainsi, fantasme devrais-je dire, une machine vivante, un modèle qui obéit à des lois (d'organisation, d'interaction, de synthèse ou dégradation ...) auxquelles des machines pourraient se conformer.

Ne nous méprenons pas sur la direction que je voudrais donner à la suite de ce texte introductif. Je ne souhaite pas remettre en question l'histoire des sciences et les choix paradigmatiques qui ont permis l'accélération considérable des connaissances sur le vivant (des plus petits éléments, subcellulaires, aux écosystèmes), en particulier celui dont je discute ici, *le Vivant-Machine*, et leur utilisation dans des applications qui ont impacté la qualité de vie des êtres humains (agronomie, médecine, technologies bio-inspirées, *etc*) et l'équilibre de la planète (mesures environnementales, sauvegardes des écosystèmes, *etc*). Il me semble pourtant nécessaire d'avoir en tête qu'un biais est à l'œuvre : celui de la subjectivité. C'est un débat en soit que celui de la subjectivité en sciences (biologiques en particulier). La question sous-jacente, c'est le choix biaisé de l'instrument de mesure pour accéder à des éléments factuels sur la composition, l'ordonnement, les rythmes du Vivant. Le microscope est un outil de choix car il permet d'observer l'ordre tissulaire, cellulaire, subcellulaire, ... mais il ne peut justement nous permettre d'observer que ce qui est ordonné. Il n'existe pas de microscopes du désordre. Les mesures d'affinité entre molécules biologiques font ressortir les interactions suffisamment fortes, celles qu'on juge suffisantes pour intervenir dans un processus ... mais *quid* de la foule autrement plus importante des interactions faibles. La question de la subjectivité, c'est aussi celle de l'*a priori* sur l'organisation et les processus qui font que le vivant vit. C'est le choix biaisé des formalismes sur lesquels nous projetons ces connaissances partielles, ces formalismes qui agissent comme autant de cadres rigides pour contenir cette matière molle, pour lisser ses comportements souvent sujets au hasard, pour que le rare devienne saillant. C'est aussi l'idée, de plus en plus présente actuellement, car alléchante, que le Vivant, par sa nature de machine complexe, pourrait servir pour réaliser des tâches que nos machines peinent à résoudre, pourrait réaliser des calculs.

Ces questions ont été maintes fois réfléchies et explorées par des nombreux théoriciens de la biologie et philosophes des sciences (ex : [Rosen, 2001 ; Woese, 2004 ; Cornish-Bowden, 2007 ; Wagner, 2008 ; Nicholson, 2013 & 2019; Boldt, 2018 ; Jeancolas, 2020]).

Je vais néanmoins développer maintenant ce point à ma manière en focalisant sur le début des années 1950 avec deux développements scientifiques majeurs : la biologie moderne et l'informatique. Ceci m'amènera ensuite logiquement à signifier ce que j'entends par *Le rare, le faible et l'amorphe* et à montrer comment, en pratique, dans les recherches que je mène, j'envisage de prendre cette question en compte.

Je m'autoriserai ensuite, vers la fin de cette partie introductive, quelques digressions d'ordre philosophiques concernant la nature de la science et la motivation associée. En effet, c'est vers le terrain des limites de la science qu'on se dirige assurément lorsqu'on commence à s'interroger sur ce

qui est observable sans biais et sur comment continuer à poursuivre des recherches scientifiques dans ces conditions. Ces digressions permettront au lecteur, s'il le souhaite, de saisir mon état d'esprit ces dernières années, jusqu'à l'écriture de ces lignes et de façon générale sur la manière dont j'appréhende les sciences.

3.1.1. Naissances conjointes de la biologie moderne, de la cybernétique et de l'informatique

Dans ce paragraphe, je vais rappeler les grandes étapes de la naissance de ce qu'on peut considérer comme la biologie moderne (la biologie cellulaire et surtout moléculaire) en insistant sur les dates. Je souhaite ici montrer qu'il s'agit d'une histoire récente. Cela me permet aussi de la mettre en parallèle d'une autre histoire, dont je donnerai aussi les grandes dates, qui est celle de l'informatique et de la robotique, de l'information et du signal. Enfin, c'est aussi au même moment que naissent les 3 théories majeures visant à expliquer les origines du code génétique.

Ce paragraphe est donc pour moi assez central. C'est un nœud où se regroupent toutes mes questions. Il est important pour comprendre la question fondamentale qui m'anime et que je défends dans ce manuscrit : la nature du vivant et sa relation avec nos machines. Il est ainsi important pour comprendre que la façon dont nous percevons *a priori* le vivant a une influence jusque dans la réflexion sur ses origines. Que cherchons nous comme émergence du vivant ? Et si nous cherchons la vie ailleurs que sur Terre, que devons nous trouver³ ? Cherchons nous la machine efficace que nous décrivons actuellement, autrement dit la cellule actuelle, telle LUCA⁴ ? Le vivant de l'ère pré-LUCA est-il déjà doté d'une *machinerie cellulaire* (système de traduction notamment) comme semblent converger les propositions en vogue [Koonin, 2009 & 2017 a] ? Est-on passé brutalement de la soupe biochimique primordiale à un système capable de réplication, de traduction, *etc*, ou est ce qu'une pente douce d'améliorations progressives joint les ères primordiales à LUCA ? Quelle fut la place des lois physico-chimiques élémentaires (affinités chimiques, diffusion anisotrope, ...) face aux systèmes biologiques que nous connaissons (molécules adaptatrices ou catalyseurs complexes, canaux protéiques, ...) ? Les processus d'efficacité apparente faible, davantage fondés sur des lois physico-chimiques élémentaires plutôt que sur une *machinerie*, furent-ils suffisants pour qu'on puisse désigner de tels systèmes comme 'vivants' ?

Déterminer où se trouve ce curseur, c'est potentiellement accepter d'élargir les formes vivantes à des systèmes qui ne ressemblent pas forcément à LUCA et ses descendants et de se référer à une définition canonique (ex : système semi-ouvert capable d'exploiter matière et énergie pour se

3 C'est cette question qu'un manager de la NASA a posée à Chris Adami [Adami 2011] : « *Chris, do you please tell us, how do we look for Life outside Earth ?* » , ajoutant « *Biosignatures ! we need to look for a biosignature !* »

4 LUCA : Last Universal Common Ancestor. Ce sigle désigne un point dans le temps que nous situons à environ 3.5 milliards d'années (entre 3.5 et 3.3 milliards d'années [Doolittle, 2000 ; Glansdorff, 2008]). LUCA est déjà une cellule évoluée qui dispose d'un système de réplication, transcription et traduction proche de celui de la cellule moderne. Il s'agit plus vraisemblablement d'un ensemble de types cellulaires qui, ensemble, ont donné le monde cellulaire tel que nous le connaissons, plutôt que d'un unique ancêtre. Vu ainsi, c'est un goulot d'étranglement provoqué par la sélection naturelle des pré-LUCA (des protosystèmes non cellulaires jusqu'aux cellules vivantes pré-LUCA en passant par les protocellules) et homogénéisé par les transferts horizontaux entre cellules.

maintenir, se reproduire et évoluer) plutôt qu'à une forme cellulaire connue (LUCA). Cela a des conséquences sur la recherche de formes de vie ailleurs dans l'univers. Cela a aussi des conséquences sur les processus biologiques actifs actuels : quelle place pour les processus physico-chimiques dans la cellule actuelle ? Mais également, sur la façon dont tout un chacun manipulera le vivant, l'analysera, le modélisera, tâchera de le comprendre, d'interagir avec lui, de l'exploiter.

Quand on s'intéresse aux origines du vivant, il faut s'intéresser de près à l'histoire de la biologie et à la manière dont les concepts du vivant ont été formulés et les connaissances observationnelles associées. Au début du XX^{ème} siècle, il y a au moins deux problèmes qui se posent en biologie : comment fabriquer les briques du vivant (acides aminés, sucres, bases azotées, acides nucléiques ...) d'une part ; d'autre part, comment la cellule fonctionne-t-elle ? Quels sont les mécanismes qui lui permettent de créer de la matière biochimique à partir des éléments de son environnement, et surtout comment sont-ils orchestrés ? Si dans la première moitié du XX^{ème} siècle beaucoup de ces éléments deviennent connus (on connaît les enzymes ; on sait qu'elles sont faites d'acides aminés ; on sait mesurer leur activité ...), ce n'est réellement qu'au milieu du XX^{ème} siècle que la biologie moderne naît avec pour bon point de départ la résolution de la structure tridimensionnelle de l'ADN par Watson, Crick (aidés de R. Franklin et R.G. Gosling par leurs travaux de cristallographie [Franklin, 1952]) en 1953 [Watson, 1953]. Notons que c'est cette même année que Stanley Miller et Harold Urey réalisent une expérience désormais célèbre [Miller 1953, Miller & Urey 1959], première expérience de chimie prébiotique sur laquelle je reviendrai plus tard.

C'est à partir de ce moment (la découverte de la structure de l'ADN) et jusque dans les années 70 que la biologie moléculaire naissante met à jour la *machinerie cellulaire*. J'y reviendrai. La suite, nous la connaissons avec l'avènement de la biologie moléculaire et son incroyable potentiel. Peu après la découverte de la structure de l'ADN, il y a eu une certaine effervescence pour déterminer comment l'information que l'on savait contenue dans l'ADN cellulaire, molécule support de l'hérédité (propriété suspectée depuis 1944 [Avery, 1944] et prouvée en 1952 [Hershey, 1952], soit un an seulement avant l'identification de sa structure), était traduite sous la forme de protéines, molécules effectrices de la cellule (enzymes et molécules de structure). Autrement dit, il s'agissait de déterminer un code génétique.

En 1954, George Gamow a proposé que les acides aminés (au nombre de 20 différents dans la plupart des cellules) s'intercalent spécifiquement entre 2 doublets successifs de nucléotides dans la double hélice (Fig. 3.1 A). C'était une première tentative d'explication qui pouvait paraître cohérente car il se trouve qu'il existe exactement 20 combinaisons de paires de doublets de nucléotides dans chaque sens de lecture de l'ADN (voir Fig. 3.1 B), soit autant que de types d'acides aminés, et ce modèle permettait d'expliquer le rapprochement stérique nécessaire entre 2 acides aminés pour former des liaisons peptidiques et obtenir ainsi des protéines. On sait évidemment que ce modèle est faux mais il avait le mérite d'être simple et il a donné naissance à la théorie de l'origine stéréochimique du code génétique que nous présenterons en détail dans la sous-section suivante.

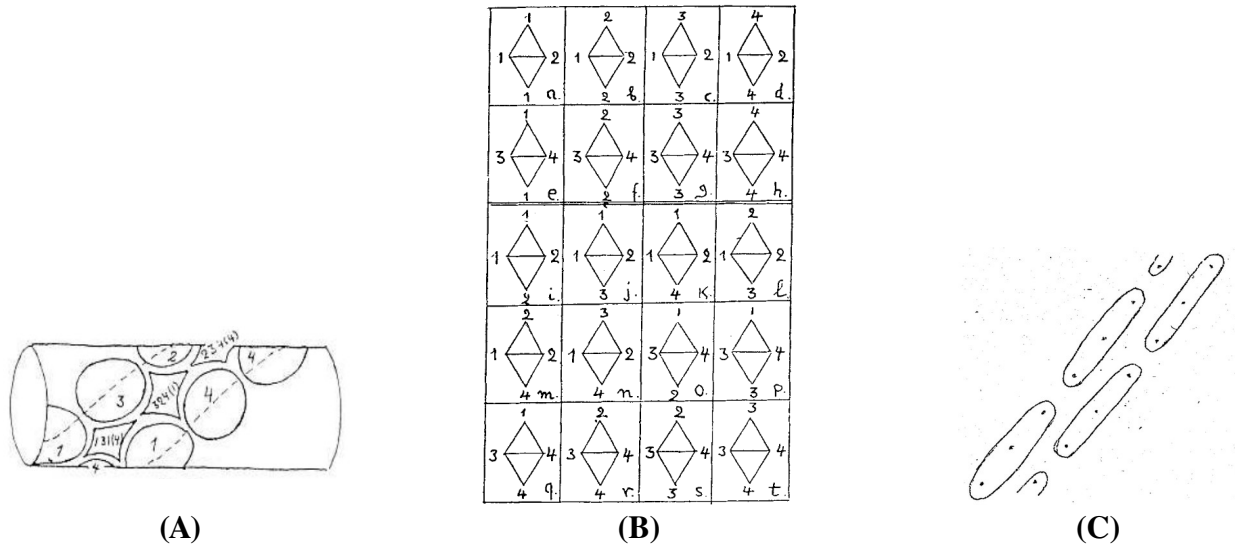


Figure 3.1. (A) Modèle du *code diamant* de G. Gamow [Gamow, 1954] et (B) ses 20 combinaisons dans un sens de lecture. (C) Modèle de triplets de F. Crick [Crick, 1955]

Peu de temps après, Francis Crick, un des découvreurs de la structure de l'ADN, répondant à Gamow, proposa dans une note manuscrite de 1955 l'idée d'un code à 3 lettres correspondant à 3 nucléotides successifs sur un brin (voir Fig. 3.1 C) et que certains ARN puissent jouer le rôle d'adaptateurs (il sous-titre notamment son article « *A note on the RNA Tie Club* » ce qui, si je ne me trompe pas dans mon interprétation signifie « Une note sur le bâton capable de se fixer et fait d'ARN », autrement dit l'adaptateur) dans la formation des protéines à partir de l'ADN [Crick, 1955]. Dans ce même article, Crick introduit la notion de triplets de nucléotides. Cette théorie de l'adaptateur devait attendre 1958 pour une confirmation expérimentale [Hoagland, 1958]. Elle donna aussi un support de réflexion pour une théorie, la « *escaped triplet theory* », une variante de la théorie de l'origine stéréochimique du code génétique. Le code génétique (mise en correspondance de codons et d'acides aminés) fut finalement découvert entre 1961 et 1966 (code génétique complet en 1966) par Nirenberg et ses collaborateurs [Nirenberg, 1961 ; Matthaei, 1962 a et b ; Leder, 1964], annoncé toutefois par F. Crick en 1961 [Crick, 1961] qui montra que ce code était constitué de triplets de bases nucléiques successifs nommés codons. La suite, c'est la découverte progressive de tout le reste de ce que l'on appelle en français comme en anglais la *machinerie de réplication et de traduction cellulaire*, composée d'une foule de composants protéiques et nucléiques (ARN messenger, ARN de transfert, Ribosome, et tout un cortège d'enzymes libres ou associées à des acides nucléiques) *a priori* finement agencés en temps et en espace, qui permet de répliquer l'ADN et de le traduire (après transcription en ARN) en protéines.

Nous avons donc vu que la biologie moderne commence avec une histoire de *support d'information* et de *code*, et se poursuit avec une *machinerie*, pour l'exécuter.

Sortons un instant de la biologie. Je trouve intéressant que cette époque coïncide avec une autre série de naissances, quelques années auparavant de la théorie de l'information [Shannon, 1949], de la cybernétique [Wiener, 1948] issue elle-même de la théorie de contrôle et des machines [Triclot,

2008], de l'informatique théorique et pratique [Turing, 1937; Lewis, 1997; Isaacson, 2015] puis de la robotique.

Ce point peut largement être discuté mais j'ai le sentiment que la biologie moderne naissante (biologie cellulaire et moléculaire) a subi l'influence immédiate de ce monde des machines, de programmes, de codes, de signaux. J'ai aussi le sentiment que c'est cette analogie qui a fortement guidé l'analyse biologique, de la simple biomolécule (ex : protéine) aux larges systèmes biologiques. Remarquons que cela ne serait pas surprenant et que ça n'aurait rien de nouveau puisque cette pensée mécaniste de l'animal-machine commence avec René Descartes au XVII^{ème} siècle et se radicalise chez certains, pour l'homme notamment avec de La Mettrie et sa vision de l'Homme Machine (publié en 1747). Les nerfs, parlons en ! des câbles électriques ! Ce c'est que montre Luigi Galvani en 1791. L'animal, l'homme, tout cela, ce ne sont que des amoncellements ordonnés de tuyaux, de câbles, d'effecteurs, ... une sorte de machine en quelle que sorte ; il doit en être de même pour son cerveau : si le cerveau de l'homme est une machine, on doit alors pouvoir le décomposer jusqu'à le comprendre voire le reproduire. C'est en tout cas ce qui sera souvent avancé jusqu'à la théorisation de l'informatique où la question du cerveau vu comme un ordinateur est posée. Bien entendu, cette pensée mécaniste a évolué avec les raffinements de la biologie et de la physiologie, mais il me semble qu'en pratique, probablement par nature, l'homme pense toujours le vivant comme une machine [Nicholson, 2013 ; Boldt, 2018 ; Keller, 2002]. Notons qu'au delà de la question culturelle et historique, l'association Vivant-Machine est peut-être à rechercher plus profondément que cela, dans la façon même dont fonctionne notre cerveau pour analyser le monde : nous autres, humains, trouvons notamment des motifs organisés dans la nature probablement parce que notre cerveau lui-même cherche à représenter les choses de la manière la plus simple possible [Lynn, 2018], il reconstruit une information de structure qui lui semble manquante comme on reconstruit mentalement un triangle lorsqu'on ne voit pourtant que ses 3 coins. Ainsi, par exemple, nous voyons des réseaux là où il n'y en a pas physiquement (réseaux de régulation, réseaux métaboliques, écosystèmes ...) ; seuls les réseaux de neurones sont physiquement câblés. Mais on peut aller plus loin en regardant les profils d'expression génétique : nous filtrons volontairement ces matrices d'expression génétique que nous révèlent les micro-arrays pour n'en faire ressortir que les interactions clairement marquées, simplifiant dès lors un sac de nœuds mêlant interactions fortes et faibles en quelques relations bien marquées.

Penser le vivant ainsi a permis de le théoriser comme une machine, d'en extraire une compréhension et éventuellement d'en utiliser des éléments pour la technologie.

Par exemple, John von Neumann s'attaque au problème de la réplication en élaborant un automate auto-réplicateur à 2 dimensions et 29 états [von Neumann, 1966], faisant apparaître la nécessité de dissocier un plan et l'exécuteur de ce plan (ce que développera Langton, avec ses boucles auto-reproductrices [Langton, 1984]), puis s'interroge sur ce qu'il nomme les *machines molles* dans son ouvrage *L'ordinateur et le cerveau* [Pignon, 1992] ; tout est dans le titre ! L'ensemble de ce champ de recherche extrêmement large et développé a donné lieu à ce que l'on nomme la Vie Artificielle [Rennard, 2002] mais également à un autre champ de recherche plus récent qui est le calcul naturel

[Conrad, 1985 ; Pfaffmann, 2000 ; Demarse, 2001 ; Grzybowski, 2005 ; Adamatzky, 2004 ; Courbet, 2015], sous-genre du calcul non-conventionnel. Cette activité de recherche, différente du calcul moléculaire [Rothmund, 2004, 2006 ; Qian, 2014 ; Courbet, 2018], vise à s'inspirer très directement de la Nature en en produisant un modèle et en l'utilisant ou à la détourner (la pirater dirais-je) en en conservant la majeure partie pour faire des calculs ou construire des dispositifs techno-biologiques (et non nécessairement biotechnologiques). Or, pour faire du calcul avec des systèmes vivants ou issus du vivant, il est souvent nécessaire d'en identifier les éléments et processus et établir un dictionnaire (mapping) des processus élémentaires du système d'intérêt vers un calcul (c'est le cas des langages de programmation amorphe ou chimiques [Asai, 2005 ; Adamatzky, 2008 ; Beal, 2011 ; Gorecki, 2014 ; Baig, 2020]), l'autre possibilité d'intégration des dispositifs de contrôle résidant dans l'usage d'un apprentissage du système biologique/chimique afin qu'il atteigne une configuration voulue, réalisant le calcul souhaité [Demarse, 2001 ; Pfaffmann, 2000 ; Budd, 2006]. Il s'agit donc de conserver au moins partiellement le contrôle. Dans ce cadre encore, le vivant est donc vu comme une machine. Pourtant, comme le souligne R. Abbott, si un arbre produit une ombre, est-il en train de me donner l'heure ? (*If a tree casts a shadow, is it telling the time ?*) [Abbott, 2006]. En d'autres termes, c'est nous, humains, qui projetons la machine sur le vivant pour gagner en contrôle.

Cela a aussi permis de modéliser le vivant avec des mathématiques et des programmes informatiques et d'avancer dans sa compréhension : nous modélisons le vivant avec des programmes qui ne sont rien d'autre que des virtualisations de machines. C'est aussi cela qui a permis d'avancer en médecine et pharmaceutique : la prothèse ou l'implant, éléments de machines issus de la technologie, s'interfacent directement avec le corps considéré aussi comme une machine.

On l'aura compris, le paradigme du vivant-machine permet aussi de considérer le modèle de machine comme *modèle générateur* des modèles du vivant. Penser la machine vivante, c'est projeter ses mécanismes sur l'objet vivant observé et les utiliser pour le modéliser.

3.1.2. Rendre plus vivant ce qui ne l'est pas !

Il est effectivement amusant de constater que ce biais subjectif (Vivant→Machine) a une réciproque (Machine→Vivant) : si pour comprendre le vivant on cherche à en faire une machine, pour le reconstruire, on cherche inversement à s'éloigner de tout ce qui peut se rapprocher d'une machine. Les chercheurs en biologie synthétique, en vie artificielle et ceux qui s'intéressent à l'émergence de la vie depuis ses débuts prébiotiques connaissent ce biais. Cyrille Jeancolas, doctorant lors de l'écriture de ce manuscrit, réalise une étude ethnographique⁵ d'un laboratoire de biologie synthétique travaillant sur les origines du vivant [Jeancolas, 2020]. Dans leur expérience de

5 c'est-à-dire une étude anthropologique mais faite de l'intérieur, C. Jeancolas étant immergé au moment de l'étude dans le laboratoire qu'il étudie (et faisant même partie des sujets étudiés vu que sa thèse de sciences portait sur l'étude de métabolismes primordiaux).

microfluidique visant à produire un simulacre d'évolution Darwinienne sur ce qu'on peine à appeler des 'protocellules', les chercheurs utilisent une batterie de dispositifs technologiques et de processus algorithmiques pour forcer une collection de micro-goutelettes d'eau et de réactifs dans l'huile à se former, à échanger des composés, à se diviser, à 'évoluer'. Là, au lieu d'utiliser le langage des machines comme le font les biologistes qui cherchent à décomposer et analyser les êtres vivants comme des machines, ces chercheurs qui visent l'abiogenèse (formation du vivant à partir de ce qui ne l'est pas, de l'inerte) utilisent au contraire le langage de la vie, celui de la biologie : 'cellule', 'vie', 'mort', 'métabolisme', 'génome' ... alors que rien de tout cela n'est en réalité présent. Ces abus de langage m'apparaissent comme le pendant symétrique de ce que font les biologistes : on cherche là à s'éloigner de la machine et à se diriger vers la vie ; on agit comme si le fait d'utiliser ces termes du vivant permettait d'insuffler cette indispensable étincelle qui manque à l'inerte pour s'animer.

D'une manière assez similaire, en vie artificielle, on rend vivants les objets simulés en leur associant des attributs du vivant. J.P. Rennard parle même de *binologistes* (biologistes du monde binaire), ajoutant une extension de logiciens du vivant aux informaticiens qui programment ces univers virtuels, comme seuls capables de mettre en évidence des « organismes » distribués dans un environnement virtuel [Rennard, 2002]. Une autre bon exemple sur lequel je reviendrai est le modèle de génétique artificielle, nommé typogénétique, proposé par D. Hofstadter puis implémenté par L. Varetto [Hofstadter, 1979 ; Varetto, 1993 & 1998]. De séquences de lettres auxquelles on attribue des règles, on cherche à tirer des enseignements sur l'émergence de boucles génétiques et métaboliques auto-entretenu. Dans ces travaux, les termes employés vont aussi vers une *biologisation* de ces objets informatiques. La vie artificielle [Rennard, 2002], c'est un monde dans lequel on parle de *biomorphes*, dans lequel les auteurs titrent leurs articles *How I created Life in a virtual universe* [Ray, 1993], dans lequel on 'apprend' sur la Vie en simulant des systèmes très artificiels. C'est également ce rapprochement entre processus physico-chimiques auto-organisés et la présence de motifs réguliers dans le vivant qui fascinait déjà Thompson d'Arcy, Nicolai Rashevsky puis Alan Turing et de nombreux autres, les poussant à proposer des causes physico-chimiques à la génération des formes du vivant et à les modéliser comme tels [D'arcy, 1917 ; Rashevsky, 1940 & 1948 ; Turing, 1952 ; Hodge, 1992 ; Murray, 1993] ; c'est aussi ce qui continua de fasciner⁶ des générations d'étudiants et de chercheurs qui, s'inspirant des travaux de ces pionniers de la biologie théorique, continuent de manipuler processus de réaction-diffusion, agents virtuels, transformation géométriques (morphing), *etc.* Obtenir par simulation une telle auto-organisation, c'est déjà obtenir *un bout de vie* ! Je ne peux dès lors m'empêcher de rappeler cette anecdote de Louis Bec qui racontait avoir perdu une de ses 'créatures' (artificielles) qu'il conservait 'en vie' dans un de ses 'aquariums' virtuels [Comm.SFBT, 2014].

Tout cela, j'en ai bien sûr aussi fait le constat moi-même. En 2017 avec un étudiant en stage de Master, Clément Hege [Hege, 2017a], nous nous sommes notamment prêtés à une petite expérience. Nous avons testé la nature perçue, vivante ou non-vivante, d'objets réels d'abord, puis de simulations numériques, par plusieurs personnes (35 participants adultes, hommes et femmes en

⁶ les induisant parfois en erreur ; ce fut mon cas [Glade, 2012]

proportion équilibrée, de domaines de compétence différents : informatique, technologie, biologie, géologie ...), sous la forme de sondages. Dans un premier sondage *Google Form* [Hege, 2017b] sans grande prétention et très largement perfectible, nous avons présenté une série de photos de coraux d’une part, de structures de nanotubes de carbones auto-organisées à une échelle macroscopique d’autre part. Ces photos montraient donc des objets biologiques et des objets issus de processus physiques, vus à différentes échelles, sans annotation ni quelconque indication pouvant donner des indices sur leur nature. Quelques exemples de photos et réponses associées sont données figure 3.2.

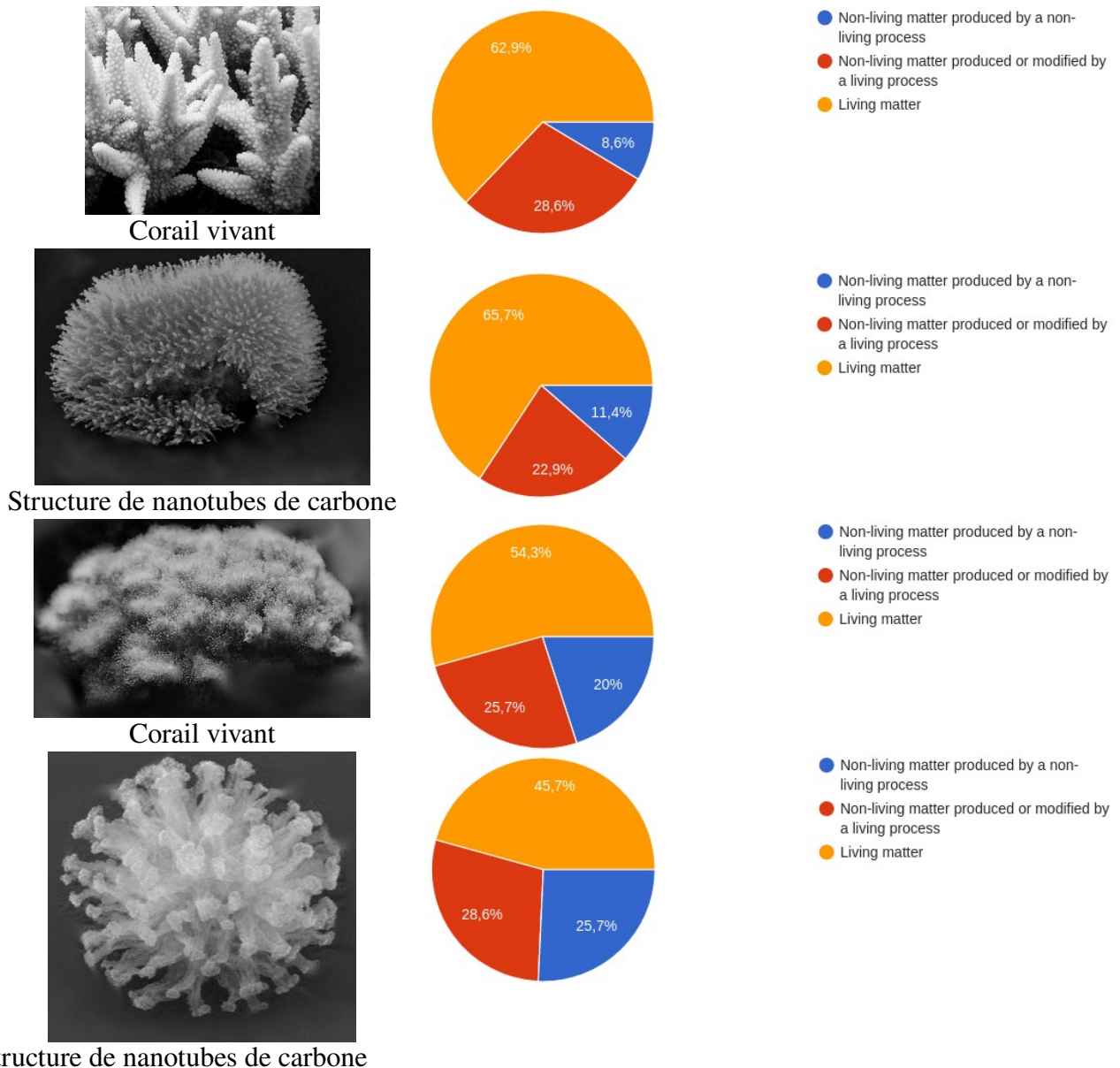


Figure 3.2. Résultats d’un sondage sur la nature perçue d’objets vivants ou inertes. A gauche, des photos montrées dans le questionnaire ; à droite les réponses des 35 participants. On voit que l’indécision domine !

Dans tous les cas, les sondés étaient mal à l’aise devant ces photos car la plupart du temps indécis. En plus d’évaluer la nature vivante ou non des objets en photo, nous demandions d’évaluer leur échelle entre le nanomètre et le mètre, ainsi que leur complexité (sur une échelle de 1 à 10). Là non

plus, pas de consensus parmi les participants et d'une photo à l'autre. Nous avons aussi demandé leur sentiment général vis-à-vis de ce test, sous la forme d'une note comprise entre 1 et 10. Le sentiment général était qu'il n'avaient que très moyennement apprécié ce test qui, d'après leurs commentaires, les plaçaient dans l'inconfort.

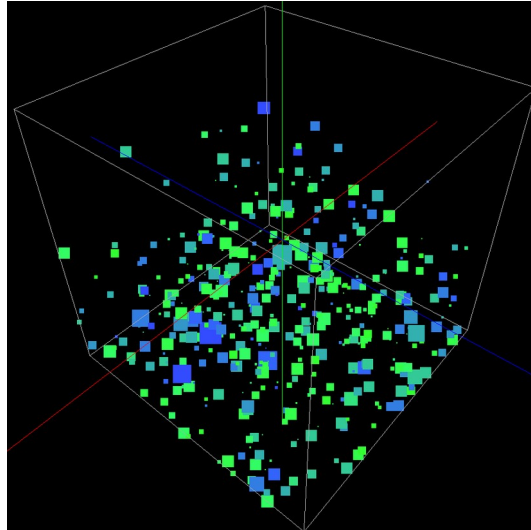


Figure 3.3. Image d'une simulation comprenant de multiples créatures se mouvant par simple agitation thermique ou sédimentation, mais aussi parfois à cause de 'nageoires'. Les couleurs sont liées à leur phénotype. [Hege, 2017a]

Un autre test a été proposé, montrant des vidéos de trajectoires de nombreux objets (figure 3.3). Il était demandé aux observateurs de qualifier le caractère vivant ou non des objets ayant produit ces trajectoires, les observateurs n'ayant aucune idée de leur origine virtuelle (il pouvait tout aussi bien s'agir de trajectoires de poissons dans un banc ou d'oiseaux ou d'insectes dans une nuée). Il s'agissait effectivement de 'créatures' de vie artificielle dotées d'un 'génomme' minimal codant pour des 'bouches', des 'senseurs' et des 'nageoires', et capables 'd'évolution' (donc de multiplication et mutations), que nous avons en réalité toutes simulées [Hege, 2017a]. Certains 'écosystèmes' se sont avérés être riches en 'créatures' aux comportements complexes, affichant des comportements de 'prédation', des déplacements complexes, des zones privilégiées de 'vie' et de 'chasse'... Ces environnements étaient riches en 'biodiversité' car ils contenaient aussi d'autres types de 'créatures' ayant des comportements très différents, misant davantage sur des 'stratégies reproductives' pour 'assurer leur survie'. D'autres au contraire, dans des 'conditions environnementales' différentes ne contenaient que des 'créatures' aux comportements inintéressants, par exemple purement sédimentaires. Là encore, les personnes interrogées, des chercheurs pourtant habitués aux simulations numériques, se sont tous trompés : pour eux, certains systèmes étaient forcément issus de mesures de systèmes biologiques tant leur comportement paraissait, non pas complexe, ni auto-organisé, ni même intelligent, mais « vivant ».

Bien que vite et pas très bien conçus, ces tests montrent que lorsque nous n'avons pas de repère particulier, d'indices faciles sur la présence possible d'organismes vivants, nous ne savons pas ce qui

est vivant et ce qui ne l'est pas avec la moindre certitude. Là, on se trouve dans une sorte de vallée dérangeante⁷ entre les mondes de l'inerte, de la machine et du vivant : on ne sait pas s'il faut rendre plus vivant un objet, en réalité inerte, en le caractérisant par exemple de structure corallienne, ou si au contraire il faut décomposer cet objet, en réalité vivant, en système doté de mécanismes et de processus.

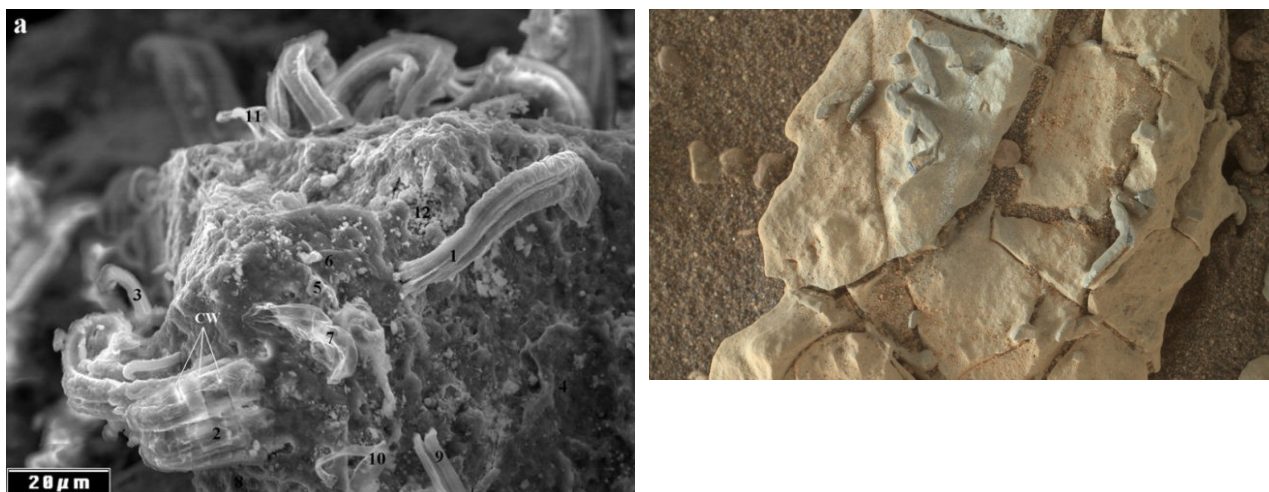


Figure 3.4. Structures remarquables dans des roches extra-terrestres. (Gauche) Vues en microscopie électronique sur des météorites carbonées de structures supposées d'origine biotique par Hooker [Hooker, 2011]. L'image est extraite d'une figure dont la légende est *Hitachi FESEM Secondary Electron Detector image at 1000 X of multiple filaments and sheaths embedded in Orgueil meteorite matrix*. (Droite) Photo prise sur Mars par le rover Curiosity de structures minérales évoquant des fossiles. Chaque structure a une longueur de l'ordre de 5mm.

C'est d'ailleurs bien le problème avec la recherche de la vie ailleurs dans le système solaire, illustré en particulier par l'embarras qu'ont provoqué certaines structures martiennes (voir figure 3.4), en particulier celles de la fameuse météorite « martienne » ALH84001 qui présentait des structures filiformes pouvant être assimilées à des fossiles cyanobactériens, leur origine biotique ou abiotique n'étant toujours pas tranchée [McKay, 2009], ou celles relevées par Hoover [Hoover, 2011] sur des séries de météorites carbonées CI1. Ce fut aussi le cas des structures macroscopiques (de l'ordre de 5mm) photographiées *in situ* par le rover martien *Curiosity* qui ont provoqué un émoi dans la presse grand public suite à un re-tweet avec un zoom sur ces structures fait à partir d'une publication du journal du rover par la NASA [Edwards, 2018]. Notons qu'ici, je ne défends absolument pas la véracité ou non de ces publications que d'ailleurs, comme beaucoup, je mets fortement en doute,

⁷ Wikipedia : « La vallée dérangeante ou vallée de l'étrange (de l'anglais *uncanny valley*) est une théorie du roboticien japonais Mori Masahiro, publiée pour la première fois en 1970, selon laquelle plus un robot androïde est similaire à un être humain, plus ses imperfections nous paraissent monstrueuses. Ainsi, beaucoup d'observateurs seront plus à l'aise en face d'un robot clairement artificiel que devant un robot doté d'une peau, de vêtements et d'un visage visant à le faire passer pour humain. Ce n'est qu'au-delà d'un certain degré de réalisme dans l'imitation, selon cette théorie, que les robots humanoïdes seront mieux acceptés. C'est pour cela qu'est utilisé le terme de vallée : il s'agit d'une zone à franchir dans laquelle chaque progrès fait vers l'imitation humaine amènera au départ plus de rejet, mais passé un certain seuil de réalisme une acceptation plus grande. » https://fr.m.wikipedia.org/wiki/Vall%C3%A9e_d%C3%A9rangeante

d'une part à cause de failles possibles dans les protocoles d'analyse (risques de contamination par des organismes terrestres) mais aussi par la piètre qualité des "revues" dans lesquelles ces hypothèses ont été défendues ; ici, j'illustre une difficulté réelle à qualifier ce qui est vivant et ce qui ne l'est pas et je me sers de ces cas typiques et extrêmes pour montrer le risque subjectif inhérent de sur-qualifier de vivant ce qui ne l'est pas forcément, cette *projection de vie sur l'inerte* qui, nous le verrons dans la suite de cette partie introductive et plusieurs fois dans le reste du manuscrit, occupe beaucoup mon esprit. Les chercheurs prudents ne s'y trompent d'ailleurs pas : dans une interview pour la revue en ligne Space.com, Seth Shostak (chercheur au Search for Extraterrestrial Intelligence (SETI) Institute in Mountain View, Calif.) est clair : « *If you look at the microscope photos, they are certainly suggestive – looking like photos made of various terrestrial bacteria, but then again, while intriguing, that's hardly proof. If similarity in appearance were all it took to prove similarity in kind, then it would be pretty easy for me to demonstrate that there are big animals living in the sky, because I see clouds that look like them.* » [Space.com, 2011].

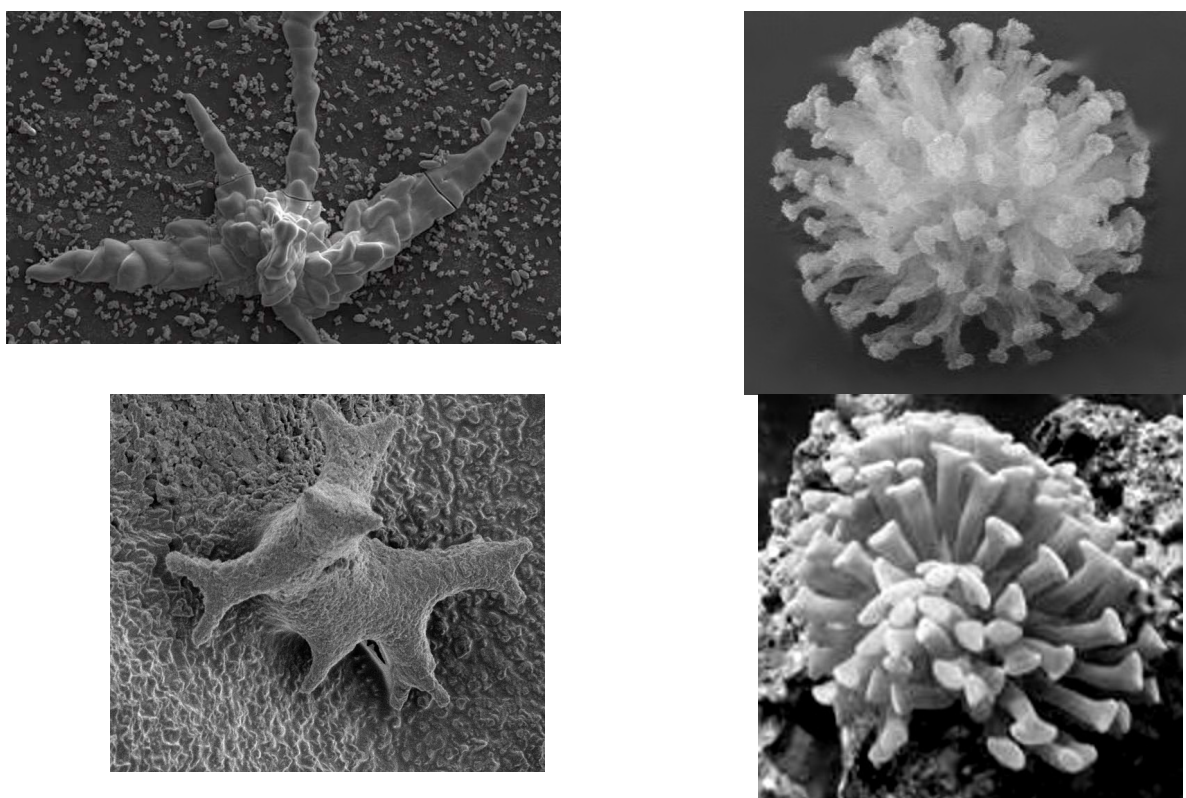


Figure 3.5. Structures abiotiques complexes vs structures biologiques. (Haut gauche) Structure de carbonates silicatés formée dans l'eau aux cascades Ney [Garcia-Ruiz, 2017 ; Phys.org, 2017]. (Haut droite) Structure auto-organisée de nanotubes de carbone [Bonnot, 1999]. Certains processus physiques peuvent être source de structures particulièrement complexes dont l'aspect est très comparable à celui d'êtres vivants comme (Bas gauche) un gros plan sur une formation corallienne ou (Bas droite) une anémone de mer.

Néanmoins, il est indéniable que ces structures ont posé un réel problème car leur forme complexe et du moins leur présence saillante et inattendue dans cet environnement minéral pouvait laisser penser à une origine biotique, d'autant qu'elles ressemblaient à des structures terrestres vivantes ou

associées à la présence de vie. De plus, on ne connaît pas toujours de processus géologique capable de générer de telles structures. L’embarras vient du fait qu’on ne sait pas s’il y a eu de la vie sur Mars. C’est pourtant quelque chose qui est absolument envisageable, Mars ayant présenté à une époque lointaine de nombreuses caractéristiques (eau liquide, géologie et volcanisme, chimie atmosphérique ...) qui ont pu permettre l’émergence de la vie pendant une période de près d’un milliard d’années, la vie sur Terre étant apparue dans un temps beaucoup plus court après la phase de refroidissement de la croûte terrestre.

C’est bien entendu le même problème lorsque l’on recherche des traces de vie fossile sur Terre. Le problème qui s’ajoute est que l’on sait que la vie est apparue sur Terre ; la tentation d’associer le minéral au vivant est donc encore plus grande que pour ce qui provient de l’espace extra-terrestre : ce pourrait être une trace de vie fossile ! Il faut pousser l’analyse, morphologique certes, mais chimique surtout, pour en déterminer l’origine avec plus de certitude. En complément, les géologues, géochimistes, physiciens, *etc.* font un travail de spécification des morphologies minérales, du moins inorganiques, et d’identification des processus physiques ou géologiques qui ont pu les générer, levant ainsi certains doutes sur certaines structures d’aspect complexe [Garcia-Ruiz, 2017 ; Phys.org, 2017] (figure 3.5). Ces travaux ainsi que le dialogue ouvert et interdisciplinaire entre communautés scientifiques permettent ainsi de suggérer des pistes de recherche de fossiles ou de traces de vie sur Terre comme ailleurs dans le système solaire [McMahon, 2018 ; Projet COOL, 2017].

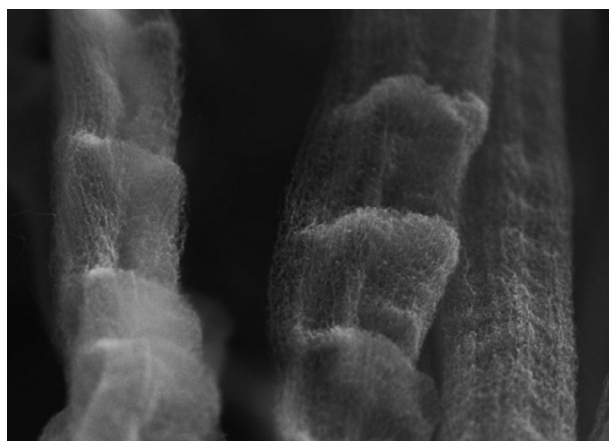
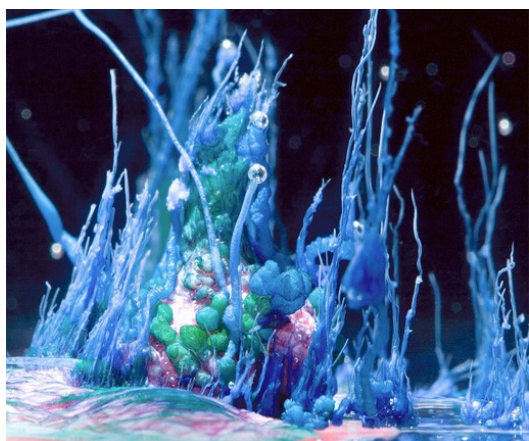


Figure 3.6. D’autres structures abiotiques. (Gauche) Jardins chimiques de LeDuc [Leduc, 1910]. (Droite) “Tricotins” faits de l’autoassemblage et l’auto-organisation de nanotubes de carbone [Bonnot, 1999].

En revanche, lorsqu’on sait à l’avance la nature de l’objet que l’on observe (des cellules sous microscope, un coquillage, une nuée d’oiseaux, ou au contraire une simulation de biomorphes, un jardin chimique, la synthèse d’une structure auto-organisée de nanotubes de carbones ; voir figures 3.6 et 3.7), on n’hésite pas à sortir de cette vallée dérangeante pour projeter ces objets vers une version plus artificielle ou au contraire plus vivante, une projection qui nous arrange pour le besoin qu’on en a, à savoir donner un sens utile à ce qui est manipulé ou observé. Cette utilité, cette rationalisation, c’est celle de ce à quoi se destine en partie la Science : créer des connaissances

utilisables, point dont je discute dans la section suivante. Je pense que la question se pose aussi pour des formes non vivantes mais issues de modélisations du vivant, en particulier des réseaux : un (modèle de) réseau de régulation biologique par exemple, possède-t-il dans sa structure (que des biologistes et modélisateurs ont établie) des caractéristiques propres au vivant (qui ne correspondent pas à des réseaux générés au hasard ou au contraire générés par un processus artificiel) ? Ou bien risqué-t-on là encore une vision paréidolique⁸ ?

Je ferais maintenant une autre remarque sur ce sujet de la projection Machine→Vivant : son lien avec la complexité des objets étudiés. La seule solution apparente pour qualifier un objet (naturel) semble être l'analyse physico-chimique approfondie ; la morphologie, elle, est trompeuse. Pourtant, il arrive un moment où celle-ci est sans ambiguïté. Si un rover trouvait un coquillage sur Mars, comme celui montré figure 3.7 (ce qui est hautement improbable, soyons clairs!), on saurait sans le moindre doute possible que la vie est apparue sur Mars. La figure 3.5 montre pourtant des structures très proches d'animaux marins. Peut-être leur complexité n'est-elle pas assez élevée pour être qualifiés sans le moindre doute possible d'animaux marins ? Même question pour les nuées d'oiseaux en comparaison à des boïds, bancs de créatures artificielles en déplacement collectif auto-organisé. Où se trouve cette limite, ce curseur de complexité apparente qui permet de sortir de cette zone grise pour une identification nette de ce qui est vivant ou non ? Dès lors que ce jugement est subjectif, il faut forcément regarder la manière dont notre cerveau traite l'information (visuelle notamment) et se sert de notre expérience de la nature pour classer ce qui la compose ... sur Terre. C'est une thèse entière qu'il faudrait développer à ce sujet, mais déjà l'article *Testing for emergence in Artificial Life* [Ronald, 1999] donne une piste intéressante, introduisant une définition de l'émergence (et potentiellement de la complexité sous-jacente) fondée sur l'expérience humaine vis-à-vis d'un système étudié. Pour ces auteurs, des formes, des motifs, des comportements complexes, émergents, ne le sont pour l'observateur que tant que celui-ci n'est pas habitué à ces formes ou comportements.

Lorsque c'est le cas, l'observateur intègre celles-ci comme primitives dans sa compréhension du système étudié : l'objet ou le comportement n'est plus émergent ; il devient un élément fondamental d'un système potentiellement plus complexe. Un banc de poissons ou une nuée d'oiseau nous paraît incroyable et magnifique les premières fois qu'on les observe ; avec le temps, on y prête moins attention car on comprend (on infère, devrais-je écrire) le fonctionnement sous-jacent, l'émergence de la nuée ou du banc n'est plus un mystère et le système "banc" ou "nuée" peut être pris comme un tout pouvant intégrer un système plus large et potentiellement plus complexe, par exemple l'ajout de prédateurs sociaux comme certains requins, des dauphins ou espadons, et de "règles" de prédation, générant un système plus complexe « banc de poisson + groupe de prédateurs auto-organisés » dans lequel on verra apparaître de nouvelles formes auto-organisées et de nouveaux comportements émergents. Par extension, il est envisageable de penser que dans la nature, des processus physico-chimiques (auto-assemblage, réaction-diffusion, collections d'agents "intelligents" auto-organisés, ...) se mêlent à quelque chose d'additionnel. Ce composant

8 Une impression basée sur des apparences. Par exemple, voir des formes animales dans les nuages.

additionnel peut être perçu par l'observateur mais non nécessairement identifié ni compris ; c'est pourtant ce petit plus qui les englobe et les rend complexes, sinon vivants.

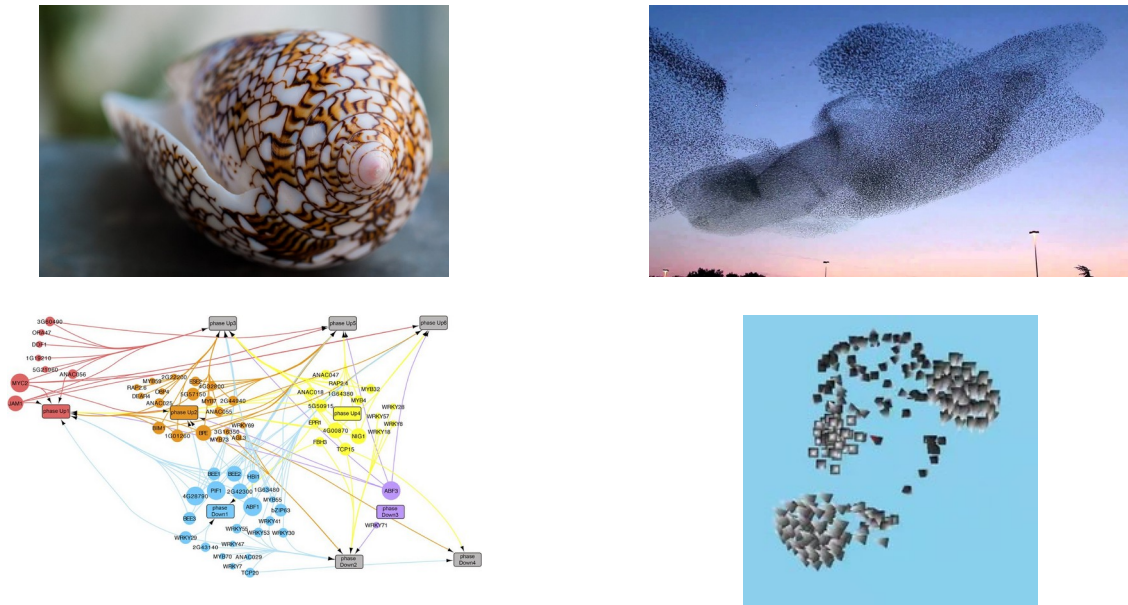


Figure 3.7. Structures du vivant et modèles du vivant (Haut gauche) Un coquillage (Creative Commons CC0). (Haut droite) une nuée d'étourneaux (Creative Commons CC0). (Bas gauche) Réseau génétique établi pour la régulation du gène de l'acide jasmonique (de Hickman et al., *Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network*, Plant Cell, 2017). (Bas droite) simulation d'un boid (image Wikipédia).

Enfin, faut-il posséder une forme pour prétendre appartenir au vivant ? Quel est le degré minimal d'organisation requis pour cela et à quelles échelles ? L'analyse de l'organisation interne et de la complexité des réseaux biologiques en comparaison d'autres réseaux (ex : aléatoires) montrent que ceux-ci sont différenciables [Costa, 2007]. Une mesure d'ordre peut aider à distinguer ce qui est issu du vivant de ce qui ne l'est pas, même dans des objets abstraits comme les réseaux d'interaction (génétiques, sociaux, écosystémiques ...). Mais est-ce toujours vrai ? Un concept comme les *Functioning-Dependent Structures* (structures dépendantes de la fonction), proposées par Ripoll, Thellier & Norris [Thellier, 2004 & 2006], des structures auto-formées dans les cellules lorsque la demande par d'autres éléments cellulaires se fait, va dans le sens d'une absence *a priori* de formes tant que la fonction n'est pas nécessaire. D'autres idées, soutenues par l'observation [Pervushin, 2007 ; Kulkarni, 2020], comme les protéines déstructurées et la structuration dynamique des protéines, mettant à mal notamment le modèle clef-serrure du fonctionnement des enzymes, vont aussi dans ce sens. Des composants (moléculaires là) du vivant, non structurés ni organisés, se structurent en apparence selon une *nécessité* de la cellule, nécessité matérialisée par des changements de concentration ou de structuration d'autres composants cellulaires, revisitant ainsi le modèle clef-serrure des enzymes ; face à un modèle statique d'interaction enzyme-ligand dans lequel la serrure et la clef sont déjà structurées et prêtes à s'emboîter, il est envisageable de considérer un modèle dynamique où ni la clef, ni la serrure ne sont structurées, sauf quand elles se rapprochent suffisamment, forçant la structuration mutuelle et permettant l'interaction. J.J. Kupiec

et d'autres, défendent ardemment l'idée que le hasard domine dans la cellule [Kupiec, 2016].

Cette idée, j'ai envie de la rejoindre : voir le hasard comme fonctionnel et ce qui est ordonné comme un biais sélectionné par l'évolution qui évite à ce hasard de partir dans toutes les directions. Cette idée est assez simple à illustrer avec une analogie sur les sociétés humaines et les lois qui les régissent. Il n'y a pas, pour dicter chaque faits et gestes de notre vie en tant qu'individu pris isolément ou dans le contexte sociétal, de règles imposées par nos lois. Dans la plupart des sociétés, la vie d'un individu est principalement dictée par son libre-arbitre et influencée par des processus d'auto-organisation spatiaux et temporels liés aux interactions entre individus ou avec l'environnement [Moussaïd, 2019]. Les lois et règles morales, vues comme un ordre global s'imposant aux sociétés peuvent alors être vue comme un biais, empêchant les groupes d'individus de fonctionner en totale autonomie (auto-organisation des groupes et foules et libre arbitre des individus), les guidant dans des directions particulières (dictées par des politiques nationales, européennes, *etc*). Une image plus détachée de l'humain est celle de la formation des fourmilières et termitières: Grassé [Grassé, 1959] décrit cela comme le couplage de la stigmergie (processus auto-organisé) et du gabarit (terme de forçage). Ces insectes sociaux communiquent entre eux par contacts inter-individuels et usent aussi de phéromones à la durée de vie temporaire pour échanger des informations via l'environnement. Ce faisant, elles n'ont pas le plan de leur future construction en tête, fusse-t-elle d'une extrême complexité [Khuong, 2016 ; Singh, 2019]. Lors de la construction d'un nid, aux communications inter-individuelles s'ajoute une stigmergie : lorsqu'un insecte dépose un élément architectural (de la terre, ...) il modifie les déplacements possibles et les flux d'air, donc la diffusion des phéromones ; cela affecte en retour l'ensemble des individus au voisinage de ces changements de l'environnement. Tous ces processus sont eux-même affectés par un biais global imposé par « la reine », vue là comme un centre organisateur : celle-ci émet une phéromone qui diffuse autour d'elle. Le profil de diffusion, décroissant avec la distance, constitue un gabarit spécifiant la taille globale de la structure à venir et sa forme. Ce gabarit, ainsi que des facteurs environnementaux pouvant l'affecter comme le vent, l'ensoleillement, ..., constituent un forçage global qui s'applique à l'ensemble de la colonie.

Je ne sais pas du tout si cette idée est transposable telle-quelle à l'échelle subcellulaire mais je pense que, déjà, l'abandon du tout-génétique [Atlan, 1999] en tant que contrôle des actions de la cellule, est un pas vers une dé-mécanisation de celle-ci. Là, la machine comme *modèle générateur* en prend un coup ! En effet, si formes et structures sont majoritairement absentes du fonctionnement du vivant, si le hasard, le libre-arbitre, le comportement individuel dominant, comment imaginer un modèle de machine sous-jacent ? Certes, l'émergence et l'auto-organisation viennent à grands renforts, mais on sent bien que dans ce cadre conceptuel particulier, le paradigme du vivant-machine bat de l'aile.

Mais alors, comment se libérer de cette subjectivité ? Ne faudrait-il pas accepter de rester dans la vallée de l'étrange ou tout reste indéfini, non-vivant et non-machine ? ! Et s'il fallait lâcher le *Vivant-Machine* pour comprendre le *Vivant* ? Cela implique de risquer de s'éloigner du paradigme

existant sur lequel toute la biologie et la biologie théorique, ses outils, ses modèles, ses théories reposent. Du point de vue pratique de la recherche (et non de la philosophie), cela pose un véritable problème. Et dès lors qu'on tente une sortie de route importante par rapport à un cadre scientifique établi et fonctionnel, on touche aussi là une bordure de la science et le risque de se fourvoyer vers une approche a-scientifique.

Avant de clore ce chapitre introductif en venant enfin au problème posé par mes thématiques de recherche, je vais m'autoriser une digression sur ma perception personnelle de ce qu'est la Science et de la façon dont je conçois sa pratique.

3.1.3. Qu'est ce que la science, pour moi ?

Il y a 5 ans de cela (en 2017), j'ai été sollicité comme plusieurs autres chercheurs, par M. Silberstein, éditeur des Editions Matériologiques⁹, pour écrire un petit chapitre de 2 pages sur le thème « Qu'est ce que la Science, pour vous ? ». Depuis, 2 tomes sont sortis. En ce qui me concerne, autant le dire tout de suite : j'ai échoué à cet exercice ! Mais il ne fut pas vain car il m'a emmené assez loin dans mes lectures de philosophie des sciences et dans mes réflexions sur mes thématiques de recherche, en particulier sur leur place dans ce document (l'HDR) et le rôle, pour moi, de ce document. Quand j'ai tenté de répondre à cette question que je trouve extrêmement difficile, car très peu claire pour moi, je me suis surtout posé la question de (1) comment et (2) pourquoi je faisais de la science ? et, plus généralement, comment et pourquoi les hommes font-ils de la science ?

Si l'on demandait à Achille son avis sur son talon, il serait bien embarrassé de reconnaître sa faiblesse. Je pense que demander à un chercheur son avis sur pourquoi et aussi comment il fait de la science l'embarrasse tout autant, dès lors qu'il cherche rentrer dans l'intimité du sujet !

Cette question d'apparence banale est en réalité extrêmement complexe et multiple et doit provoquer malaise et introspection chez le chercheur pour peu qu'il prenne le temps d'y penser. Dans la routine du métier (les observations, les mesures, la lecture, les modèles, la rédaction d'articles et de rapports ... mais aussi tout l'à côté moins réjouissant et qui ancre le chercheur dans une réalité sociale) cette question n'est, fort heureusement, pas particulièrement présente ni complètement indispensable. Fort heureusement aussi, le chercheur part, c'est à espérer, avec un bagage scientifique et une formation à l'analyse, à la critique, accumulée sur de longues années et, fort de ce bagage, le chercheur est confiant dans sa clairvoyance et son jugement utiles à son métier de tous les jours. Néanmoins, avec cette question pensée ainsi, avec une réponse dans un format aussi court (2 pages), la nuance nous est interdite ; nous sommes contraints à la prise de position, au choix épistémologique difficile. Marc Silberstein, dont je trouve l'initiative excellente, n'aurait pas pu s'y

9 ... avec qui nous avons travaillé, Angélique Stéphanou et moi, pour l'édition de 2 recueils d'articles de conférence [Materiologique 2013, 2015] dans le cadre des écoles de printemps de la Société Francophone de Biologie Théorique.

prendre mieux pour stimuler les routards de la mesure et du modèle, et afficher au grand jour la psyché des chercheurs.

Pour aider dans ce choix (ou provoquer un vertige, c'est selon), des générations de philosophes des sciences et de scientifiques qui s'y sont essayés (Bachelard, Barrau, Cariou, Chalmer, Descartes, Kuhn, Feyerabend, Heisenberg, Morin, Popper, Buffon, Rousseau, Voltaire, pour ne citer qu'eux, étant donné que ce sont ceux que j'ai lus) se sont escrimés à proposer des définitions de la science, des méthodologies, pour appréhender cet objet insaisissable qu'est la science. Encore heureux ! Comment penser sans avoir appris à penser ? Si leurs textes sont finalement assez peu lus par le monde scientifique qui vit plus que jamais dans l'immédiateté du résultat et la confrontation factuelle, leurs idées finissent tôt ou tard par diffuser en chacun, à travers une phrase isolée dans un article, lors d'une conférence, ... Les chercheurs sont aussi des penseurs qui tôt ou tard relèvent la tête de leurs instruments et s'interrogent sur leur démarche.

Peu d'entre eux (les chercheurs, les penseurs, les philosophes) par contre ont développé la question de « Pourquoi fait-on de la science ? » Voici ce qui, au fond, me motive ; voici ce que j'ai tenté de répondre à cette question provocante « Qu'est ce que la science pour vous ? » et qui m'a amené tout doucement à me rassurer sur mes choix thématiques, sur mon positionnement en tant que scientifique :

Le « Pourquoi ? » Mon opinion personnelle est que la Science est un moyen comme un autre qu'a développé l'humanité pour répondre à une angoisse ! Mieux, la Science est une jouissance nécessaire à l'humanité.

Certains objecteront, disant qu'il s'agit surtout pour l'humanité, depuis toujours, de la recherche de solutions pour améliorer son quotidien, sa survie. Je dis que cela revient au même. Le lecteur me pardonnera ce passage pessimiste :

L'Univers et la vie ont une origine qui effraie : elle est incertaine, débattue, lointaine, tumultueuse. S'y identifier est déjà difficile. Mais la vie dans l'univers, la vie telle que nous la connaissons, ainsi que l'Univers, ont une fin. D'une manière ou d'une autre, tout ce que l'homme a produit (ses savoirs, ses écrits, ses mythes, ses arts ...) disparaîtra. C'est un fait que la physique elle-même ne saurait récuser. Dès lors qu'on est un tant soit peu conscient de cela, comment peut-on alors acceptablement trouver l'envie et la force de continuer à produire Sciences, Arts et Littérature ? L'altruisme immédiat, celui qu'on réserve à notre descendance et à nos proches, le gène social par opposition au gène égoïste de Dawkins [Dawkins, 1976], est une explication : c'est la force de l'empathie, l'espoir de la survie de ce que l'on a fait ou dit pour autrui. J'ai bien peur que cet espoir soit mal fondé et résiste peu à une réflexion sur le long terme. Je ne vois donc qu'une seule autre explication : la jouissance ! Chalmer, lui, parle d'« *entreprise [humaine] excitante* » [Chalmer, 1987]. Nous, l'humanité, cherchons la jouissance, sous toutes ses formes, pour cacher, pour enterrer nos angoisses sous suffisamment de bons moments et de bonnes choses. La recherche du savoir devient donc *in fine* un assouvissement d'un désir de jouissance face à deux angoisses : d'une part l'angoisse de l'inconnu et du vide, une angoisse relative à celle de solitude ; l'angoisse de la mort,

du néant final d'autre part. Bien que je n'imagine pas que cela puisse en réalité être le cas, l'homme est jusqu'à preuve du contraire le seul être pensant dans l'univers, la vie est jusqu'à preuve du contraire unique. De plus, la nature du monde physique ou biologique nous échappe ...

Remplir ce vide de connaissances sur ce cosmos mystérieux qui nous englobe, chercher autrui (autres vies, autres consciences), vouloir créer d'autres formes de vie ou de conscience, tout cela permet aux hommes de se rassurer grâce aux sciences, un réconfort immédiat, temporaire, ... une jouissance. Je ne développerai pas la suite qui pourrait poser que l'angoisse est un instinct (lié à la survie) sélectionné par l'évolution, au même titre que les réponses à cet instinct par le développement de l'intelligence animale (voyons plus large que l'homme) comme une nécessité poussée par la mise en place de circuits du plaisir et de la récompense (cf. K. Lorenz ; voir [Leboucher, 2012]). C'est intéressant, mais je risquerais de m'éloigner beaucoup trop du fil de ce paragraphe, déjà singulier, qui vise à faire comprendre quelles motivations m'animent.

Pour clore ce paragraphe et passer en toute logique au suivant, citons Le Quellec : « *L'homme est un animal curieux et il a horreur des vides cognitifs. Il préfère toujours croire à quelque chose plutôt qu'à rien. Il ne cesse donc de sélectionner dans le monde qui l'entoure des éléments qu'il interprète et il en tire des récits d'ordre mythologique.* » J.-L. Le Quellec¹⁰

Le « Comment ? » Derrière la question du comment, il faut commencer par s'interroger à « comment l'humanité gère-t-elle ses interrogations, ses angoisses ? » Ensuite seulement, on est en mesure de réfléchir à comment il faut procéder pour pratiquer une *bonne science* ? Cela amène inmanquablement à revoir ce que sont les canons de la méthodologie scientifique. Là, on se demande « quels sont les canons de la méthodologie auxquels il est recommandé de se conformer finalement ? » et enfin, où sont les limites de la science, comment les identifier et surtout comment les traiter scientifiquement ? Je me suis inévitablement intéressé à ces deux derniers points, particulièrement quand j'ai passé en revue mon activité scientifique antérieure et lorsque j'ai abordé mes nouveaux sujets d'études, d'intérêt du moins.

Si je devais résumer en une page ce que je pense de *la Science*, ce que je crois en comprendre, une définition que je souhaiterais retenir, je dirais qu'il s'agit à la fois d'une culture et d'une méthode dotée de son éthique propre, visant à expliquer le monde (émettre des théories scientifiques) à partir de faits observables ou issus de théories antérieures. Cette récolte des données, de conception d'instruments et de protocoles dédiés, ces recoupements multiples et échanges d'idées, de faits, de théories, c'est une forme d'artisanat dans laquelle chaque individu contribue à générer une vue instantanée des choses de la Nature, une photographie qui a une durée de vie limitée et qui trouve un temps son usage dans la société des hommes. Cette photographie résulte de fait d'une construction de l'esprit et requiert souvent, pour permettre son expression et servir à la réflexion un système d'externalisation des idées [Abbott, 2006] : papier et crayon, ordinateur, présentation orale, etc.

Dite ainsi, cette définition pourrait finalement tout-à-fait correspondre à celle de la littérature, des

¹⁰ Il n'y a pas de référence. J'avais, il y a quelques années, noté cette phrase de lui mais je suis bien incapable de retrouver dans quoi je l'avais lue...

arts ou des religions (ce que je désignerai, en incluant les Sciences par *les humanités*) qui sont d'autres moyens valables pour l'homme d'exprimer ce qu'il comprend des choses de la Nature. Un peintre capture d'une certaine façon un paysage ou une expression humaine mieux que ne pourraient les décrire des biologistes ou des médecins. Les religions proposent des réflexions sur la place de l'homme et de ses actions dans le monde et servent de guide spirituel (de l'esprit) à de nombreuses personnes. Leur perception de la psyché humaine diffère de celle que donneraient des psychologues, mais elles ont un rôle pour beaucoup, entre autre de donner des repères aux hommes et de lever leur angoisse du néant ou de la mort. Les descriptions de la nature de l'homme, que ce soit par un artiste peintre ou par une personne de foi, ont un rôle dans nos sociétés ; ce qu'elles apportent est sensé apaiser ou stimuler ceux qu'elles ciblent ; elles sont donc fonctionnelles. Je pense par ailleurs que la science, comme toutes les humanités, n'échappe pas non plus à une certaine forme de sacralisation du monde. Je ne parle pas là de la sacralisation au sens émergeant des religions, mais au sens du sacré universel, celui qui crée l'émoi et un respect infini devant la beauté du monde. Le peintre la saisit, la religion en fait symbole, le scientifique, lui, la perçoit aussi dans l'harmonie, l'agencement incroyable des processus de la nature. Il pourrait rejeter cela, mais se contente de le mettre de côté un moment et de s'en servir pour relancer régulièrement sa motivation à comprendre ce qui fait que le monde physique et le monde biologique lui paraissent ainsi. La science et ses représentants humains, sujets malgré eux à la notion du sacré, essaient juste de s'en affranchir le plus possible.

En quoi la science diffère-t-elle des autres humanités, notamment pour résoudre cette double angoisse du néant et de la mort ?

Selon moi, elle diffère des autres humanités sur au moins deux points. D'abord, elle porte en elle un vœu de se débarrasser du subjectif de façon acceptable. On arrive là évidemment à une question qui est tout aussi complexe, relative au choix de la méthode scientifique pour éliminer tant que possible ce biais. Je trouve à cet égard qu'A. Barrau, dans son essai « La vérité dans les Sciences » [Barrau, 2016] explique bien cette difficulté et les choix humains (nécessairement subjectifs, tant dans les choix d'observation que dans les choix des modèles sur lesquels projeter ces observations) que l'on doit accepter. Kuhn, dans « La structure des révolutions scientifiques » [Kuhn, 2008] détaille particulièrement l'ensemble de relations complexes qui unissent les idées reçues, les connaissances « établies », les croyances et les hommes, de science en particulier. Il formule cette inquiétude légitime de la dépendance des sciences aux croyances des hommes et idées reçues acquises par l'enseignement ou la transmission directe par des pairs. Il répond à ce propos : « *L'observation et l'expérience peuvent et doivent réduire impitoyablement l'éventail des croyances scientifiques admissibles, autrement il n'y aurait pas de science. Mais à elles seules, elles ne peuvent pas déterminer un ensemble particulier de ces croyances* [Note : elle ne peuvent pas déterminer celles qui à la fin, seront retenues comme bonnes, acceptablement valides, utiles à la science]. *Un élément apparemment arbitraire, résultant de hasards personnels et historiques, est toujours l'un des éléments formatifs des croyances adoptées par un groupe scientifique à un moment donné. Cet élément arbitraire n'indique pas cependant qu'une communauté scientifique puisse fonctionner sans un ensemble d'idées reçues. Il ne diminue pas non plus les conséquences de l'ensemble particulier*

d'idées auxquelles la communauté souscrit à un moment donné. La recherche réelle ne commence guère avant qu'un groupe de scientifiques estime qu'il est en possession de réponses solides à des questions [...]»

S'ajoutent à cela deux écueils supplémentaires. D'abord le non-observable (trop petit, trop rapide, inaccessible à la physique comme l'intérieur des trous noirs). La théorie peut venir au secours de l'observateur qui a perdu la main sur l'objet de son étude. On posera dans ce document la question de la pertinence de faire des recherches dans ce sens. Ensuite, il y a aussi le fait que, de nos jours, alors que la recherche vit sous pression de publication et d'obtention de crédits durement arrachés aux rares sources de financement, la reproductibilité des travaux de recherche n'est plus favorisée. Qui, de nos jours, se lancerait dans des travaux visant à reproduire ceux d'un autre ? Cela se pratique bien sûr dans certains domaines comme les études cliniques ou de lourdes expériences de physique, mais il faut bien admettre que beaucoup d'études de reproductibilité n'apporteront pas la publication salvatrice à leurs auteurs !

Le deuxième point qui selon moi introduit une distinction majeure avec les autres humanités concerne son temps d'actualisation propre. Ces humanités, les arts, les sciences, la philosophie, les religions, *etc*, sont confrontées au constat de l'évolution des idées avec l'avancée des siècles ou simplement des années. Ceux qui en bénéficient ont des attentes différentes vis-à-vis d'elles ; répondre à leurs questionnements fait partie des attentes fondamentales dans tous les cas. Comme nous l'avons dit, il n'y a pas vraiment de « meilleure » ou de « moins bonne » réponse, d'une part parce qu'il n'y n'est pas possible d'atteindre une quelconque meilleure vérité sur le monde, il est juste possible de s'y intéresser avec une méthode donnée, d'autre part les modèles du monde proposés par ces humanités sont conformes aux attentes de ceux qui les espèrent ; ils leurs donnent la compréhension dont ils ont besoin. Chacune de ces humanités diffère par la méthode bien entendu, mais ce que je veux signifier ici, c'est que si l'attente évolue, la vision des choses de la Nature devra aussi évoluer quelle que soit cette méthode. Ce sont les temps de mise à jour des idées qui diffèrent en fin de compte, ceux-ci dépendant directement des attentes de ceux qui en bénéficient. Les idées, dans ces humanités sont stables ! Elles doivent l'être le plus possible ! On ne peut construire une vision du monde, des paradigmes, si tout ceci évolue sans cesse de façon instable. Les arts ont leur durée propre de stabilité, les styles graphiques [Ellis, 2018] par exemple durent un certain temps puis subissent un changement. Les religions, la philosophie connaissent aussi leur propre évolution liée à ce besoin de stabilité. Les sciences également ! C'est l'utilité du paradigme, ou noyau dur, selon comment le nomment les épistémologistes.

Pourquoi parler de tout cela ? A ce stade, celui qui lit ces lignes se doute déjà que ce sont mes propres interrogations sur la nature des choses du monde et surtout la façon dont je les conçois, les modèles sur lesquels je projette ces conceptions qui me posent problème. Mes choix thématiques touchent des objets, des systèmes, qui se situent souvent dans les limites de l'observable ; les méthodes de travail que j'emploie (la modélisation principalement, mais aussi des mesures expérimentales de très petites interactions) contiennent une part importante de subjectivité tant dans

les choix que dans l'interprétation. Elles génèrent en moi un malaise, mais il s'agit aussi d'une inquiétude salvatrice puisqu'elle m'incite à la prudence. Je pense aussi, pour en avoir discuté avec quelques chercheurs et étudiants en thèse sujets à ces questions du fait de leurs thématiques, souvent en biologie théorique, que le texte que j'ai développé ici peut les aider dans leurs réflexions sur la nature des objets de leur recherches et sur la façon dont ils les étudient.

Venons en maintenant à la dernière partie de ce chapitre dans laquelle j'introduirai le programme de recherche que je présente dans ce document, un programme qui parle de très faibles ou très rares interactions, de systèmes non-structurés, de non-unicité des modèles.

3.1.4. Le rare, le faible, l'amorphe

Un des buts de la Science c'est donc, entre autres, de créer de la connaissance (j'utilise là une vue constructiviste de la science, la *Connaissance* en tant que *Vérité Scientifique* sur la nature véritable du monde n'étant définitivement pas accessible [Chalmer, 1987 ; Barrau, 2016]) faisant suffisamment consensus et permettant de faire le pas qui suivra dans notre compréhension du monde et éventuellement de servir l'homme, d'améliorer son quotidien, sa survie ...

A cet égard, bien qu'emprunts des biais subjectifs précités, le dialogue entre le monde de la biologie, des mathématiques, de l'informatique, de la technologie, ainsi que les analogies entre les objets de ces mondes sont profitables à la Science ! Faut-il pour autant ne pas tenter de s'éloigner d'un paradigme dont le biais est identifié ? Comment procéder et quel risque y a-t-il à faire un peu dissidence ? Rejeter un paradigme existant, brutalement, sans transition douce, bien que cela se présente au cours de l'histoire des sciences, est rarement une bonne chose car on perd un large socle théorique et, de là, toute possibilité de projection des observations pour rendre celles-ci intelligibles dans un cadre conceptuel donné. Même les développements rapides de la physique quantique d'une part, de la théorie de la relativité générale d'autre part, qui nous apparaissent souvent aujourd'hui comme un écart sauvage par rapport aux théories ondulatoire et de l'éther, ont émergé dans un contexte en réalité plutôt préparé à ce changement [Univers, 1989 ; Thorne, 1997]. La catastrophe ultraviolette était déjà bien identifiée comme une faille majeure des théories en vigueur. Le remplacement théorique qui s'ensuivit était donc attendu. La relativité, elle, née d'Einstein, était également attendue, en particulier pour résoudre l'énigme du milieu de propagation de la lumière, Michelson et Morley montrant qu'il ne pouvait s'agir de cet éther pour lequel tout un champ théorique avait été développé [Michelson, 1887]. Cette théorie subit malheureusement des attaques très lourdes liées notamment à l'anti-judaïsme de l'époque et aux anti-germaniques, en particulier en France [Moatti, 2007] ; la science vit dans un système humain et subit aussi ses défauts. Il me semble que le seul exemple de cassure brutale de paradigme que je connaisse est le coup de pied, très mal accueilli, mais chanceux, que Fritz Zwicky infligea au monde de l'astronomie et de l'astrophysique naissante en 1933 : F. Zwicky observant la distribution et le mouvement de la matière dans l'univers, notamment grâce au catalogue galactique imposant qu'il contribue à enrichir,

suggère contre vents et marées la présence de la matière noire, une matière invisible nécessaire à l'évolution de la matière dans l'univers. Zwicky, fort de son sale caractère, d'un fort entêtement et d'une certitude de détenir des vérités que les autres n'avaient pas, produisait beaucoup d'idées parfois farfelues, certaines géniales, pas toujours testables, en tout cas à l'époque. Dans son esprit, ce que défendaient les autres, des *spherical bastards*, comme il les désignait, à savoir les paradigmes d'une époque, était faux et devait être remplacé sans discussion. Bien qu'il eut raison plusieurs fois (mais pas toujours !), son esprit brillant aidant, sa démarche était a-scientifique. L'histoire ne s'y trompe pas : ses idées sur la matière noire par exemple ne furent considérées qu'à partir des années 1970 lorsque des éléments observationnels tangibles et des théories bien construites (ce que ne faisait pas Zwicky) les remirent au goût du jour sur le devant de la scène. Freeman Dyson dira de son approche de la science « *Unfortunately, the approach does not seem to work so well if your name is not Fritz Zwicky* »¹¹ [Dyson, 2005].

Changer de paradigme en vigueur (la théorie cellulaire, le vivant-machine) n'est pas le but. Je souhaite davantage considérer un point de vue différent sur la Nature car je crois, et c'est ma formation et mon expérience transdisciplinaire qui veulent ça, que changer de prisme soit théorique, soit méthodologique, soit disciplinaire, est toujours un excellent moyen pour étudier et réfléchir sur un système biologique.

Pour commencer, qu'est ce qui ne relève pas de ce *paradigme du vivant-machine* ? Eh bien, c'est ce que je désigne par *rare, faible et amorphe*, c'est-à-dire l'opposition directe de tout ce qu'on se fait comme idée d'un processus biologique pouvant être étudié. Là se pose inévitablement la question de la réalité d'une interaction ou d'un mécanisme : s'il est rare, *trop rare*, faible, *trop faible*, ou amorphe, *trop indéfini*, comment un tel système (que dès lors on peine à appeler *système*) peut-il faire l'objet d'une étude ? De plus, à partir de quelle fréquence et quelle intensité peut-on *acceptablement* retenir une interaction, un mécanisme comme fonctionnel ? Conrad propose une version computationnelle de ce problème dans un de ses articles notoires, *Scaling of efficiency in programmable and non-programmable systems* [Conrad, 1995]. Dans cet article, Michael Conrad, père du calcul moléculaire, interroge sur la proportion d'interactions disponibles pour réaliser une tâche effective dans les cellules ou des dispositifs technologiques, en comparaison de la foule d'interactions non-fonctionnelles. Notons que Conrad développait cette idée en vue d'exploiter les processus microscopiques qui régissent la vie des cellules, pour faire des calculs. Mais il me semble que c'est à l'évolution surtout qu'il faut le demander : si un processus ou un système, si rare, si faible, si amorphe soit-il, persiste au cours de l'évolution, c'est en toute vraisemblance qu'il n'a pas été éliminé comme élément non parcimonieux¹² du vivant par la sélection naturelle.

Comment procéder ?

Cette opération est complexe, d'abord parce qu'elle impose de limiter notre utilisation de concepts

11 Malheureusement, cette approche ne semble pas fonctionner si bien si vous ne vous appelez pas Fritz Zwicky.

12 La question de la parcimonie du vivant restant un autre problème ouvert.

classiquement utilisés pour représenter, mimer, simuler les processus du vivant. Là où l'on a établi un réseau de régulation génétique, un beau câblage bien net, que faut-il mettre à la place ? Une voie de différenciation, finalement, est-ce aussi simple qu'une cascade d'activations ou d'inhibitions qui ressemble à la mise en route d'un cockpit d'avion, bouton après bouton ? Les voies de différenciation sont-elles seulement celles qu'on apprend sur les bancs de l'université, un système d'adressage fonctionnant tel le courrier à la poste. Existe-t'il d'autres voies de communication intracellulaire, moins spécifiques, exploitant certaines structures internes (ex : microtubules) pouvant briser l'isotropie de la diffusion ? La compartimentation cellulaire, par exemple des systèmes membranaires, est-elle vraiment cette description d'anatomiste différenciant la membrane nucléaire, le réticulum endoplasmique et l'appareil de Golgi ou est-ce un continuum comme on le montre maintenant. Comment représenter de tels systèmes dans un ordinateur en essayant de ne pas encoder une machine *stricto sensu* ?

Elle l'est aussi car, si comme je l'ai dit on cherche à mettre en évidence l'existence de phénomènes rares, trop rares, faibles, trop faibles, et trop peu structurés, il va falloir souvent procéder à l'inverse du traditionnel observation → modèle → déduction. Ce n'est pas de l'observation qu'on déduira la théorie par construction progressive. C'est à partir d'hypothèses théoriques et, dans ce que j'avance, d'hypothèses non-fondées sur le paradigme du vivant-machine, couplées à de l'expérimentation numérique qu'on cherchera à inférer des modèles plausibles pour le vivant, à simuler leur comportement, en particulier à montrer l'importance de la présence de tels systèmes rares, faibles ou amorphes dans le fonctionnement naturel des systèmes vivants et, *a posteriori*, on cherchera à identifier les traces de ces systèmes en comparant ce que suggèrent les simulations numériques avec les observations du vivant.

3.1.5. Structure du document – recommandations sur sa lecture

Dans ce document, par l'intermédiaire de diverses thématiques, origines du vivant, réseaux, diffusion anisotrope, j'illustrerai ces questions que j'ai abordées dans l'introduction et tâcherai de montrer que l'on peut mener des recherches et proposer à des étudiants des sujets de thèse viables dans ce cadre.

Dans ce document, je couche enfin et de manière personnelle, ce que j'ai fait ces 20 dernières années et ce que j'envisage pour l'avenir. Il commence par mon dossier scientifique et académique (partie II, incluant les sections 4, 5 et 6) puis par mes travaux antérieurs (partie III, section 7) qui, d'une part sont requis dans un manuscrit d'HDR et d'autre part permettront au lecteur de se faire une idée de la trajectoire qui m'a emmenée ici ; la lecture de cette partie du document n'est, j'en ai bien peur, pas des plus passionnantes, aussi peut-être puis-je recommander au lecteur de le survoler et de passer rapidement à la suite, le programme de recherche (partie III, sections 8, 9 et 10), et venir se référer aux travaux antérieurs et éventuellement au dossier scientifique lorsque je le recommande. Ensuite seulement, je présente mes travaux de recherches actuels (section 8), à commencer par une longue partie sur les origines du vivant, un thème de recherche très riche que j'affectionne. J'aborde

ensuite les interactions dans les systèmes actuels en partant des réseaux d'interaction, exposant au passage la façon que je vois pour se libérer un peu de leur pouvoir structurant, et en allant vers les *trail systems*¹³ et la diffusion guidée, d'autres façons de concevoir l'auto-organisation et la communication dans les systèmes vivants.

Je termine ce document par deux chapitres (sections 9 et 10) qui sont davantage une ouverture qu'une conclusion. Je me sers de ces parties pour combiner les différentes thématiques auxquelles je réfléchis et ouvrir sur de nouvelles questions que m'ont inspirées les quelques centaines d'heures de rédaction que m'a demandé ce document, et finalement ... conclure.

J'ajoute à cette version révisée un chapitre (le chapitre 11) qui fait suite à la soutenance. Ce chapitre répond aux questions très pertinentes (et que j'ai trouvées très intéressantes) des rapporteurs et examinateurs, des questions et remarques qui m'ont été adressées dans les rapports, oralement lors de la soutenance, ou au cours d'échanges informels. J'y développe notamment un certain nombre de points que je n'avais pas pris le temps de bien traiter avant, en particulier l'affine mon propos sur mon positionnement vis-à-vis des approches mécanistes ou holistes, du rôle et du statut des modèles et approches théoriques, et surtout je discute de la question de la finalité dans le vivant. Je pense que ça lecture *a posteriori*, après la conclusion donc, présente un intérêt pour le lecteur dont j'aurais déjà retenu l'attention jusque là.

Bonne lecture !

13 Les *trail systems* sont des systèmes naturels complexes dont les éléments doués d'une activité propre laissent des traces de leur passage dans l'environnement ; en retour ces traces sont exploitées causant leur auto-organisation.

PARTIE II.
DOSSIER SCIENTIFIQUE

4. CURICULUM VITAE

Dernière chance d'échapper à cette partie bien peu enrichissante : je suggère de passer directement à la section 8, éventuellement à la section 7.

ETAT CIVIL ET COORDONNEES

Nom : GLADE

Prénoms : Nicolas, Gabriel

Date de naissance : 20 février 1976

Lieu : Ollioules, Var

Tel : 04 56 52 00 42

E-mail : nicolas.glade@univ-grenoble-alpes.fr

URL: <https://sites.google.com/site/nicolasglade>
https://www.researchgate.net/profile/Nicolas_Glade2

Adresse : Laboratoire TIMC, CNRS UMR5525, Domaine de la Merci, 38710 La Tronche.

POSITION ACTUELLE

Je suis Maître de Conférences à l'Université Grenoble Alpes (anciennement Joseph Fourier), rattaché à l'UFR Polytech Grenoble, en section 27, recruté sur un poste 26/27 avec un besoin fort en mathématiques, traitement des données, interdisciplinarité math-info-biomédecine. J'enseigne l'informatique, les mathématiques, les statistiques et la modélisation statistique, le traitement de l'information, la physique et suis responsable d'année 3 de la filière Technologies de l'Information pour la Santé à Polytech depuis 2011 (biseau en 2010). Mes recherches se situent en mathématiques et informatique et en application des mathématiques et de l'informatique à la biologie. Je suis rattaché au laboratoire TIMC, CNRS UMR5525 et mène mes recherches au sein de l'équipe BCM (Biologie Computationnelle et Modélisation).

INTÉRÊT SCIENTIFIQUE

- Origines du vivant : origine stéréochimique du code génétique ; reliques ARN et fonctions ancestrales des miRs ; auto-organisation spatiale ; influence des champs externes vectoriels sur l'émergence et la biodiversité dans des conditions prébiotiques.
- Méthodes formelles pour la détermination de réseaux d'automates booléens à seuil. Application à la programmation modulaire (composition) et taxonomie des programmes. Application à la théorie ataviste du cancer.

- Ordre et désordre intra-cellulaire ; diffusion anisotrope. Voies de signalisation et de transport, chenaux amorphes et puits entropiques dans les cellules vivantes.
- Computation naturelle fondée sur la dynamique collective d'agents fibrillaires microscopiques de type *trail systems*. Modélisation multi-échelles de la dynamique et de l'organisation des trail systems et plus particulièrement d'éléments du cytosquelette.
- Fibration isochrone, fluctuations vs robustesse dans les systèmes dynamiques.
- Substitution sensorielle : perception de formes et évolution dans des espaces à 3 dimensions comprenant obstacles et objets mobiles – Applications : substitution visuelle et suppléance perceptive.

PARCOURS PROFESSIONNEL

- **Ingénieur de recherche** (programme spatial en biologie) – CEA/CNES-ESA, France / Suède 2002-2003
- **Post-Doctorat** (microsystèmes plastiques) – laboratoire LETI, CEA, Grenoble 2003-2004
- **Post-Doctorat** (modélisation biologique) – laboratoire CMM, Santiago du Chili 2005-2006
- **Chercheur contractuel** (modélisation biologique) – laboratoire TIMC-IMAG, Grenoble 2006-2007
incluant 2 périodes sous convention de chercheur bénévole (09/04 à 04/05) et (01/07 à 07/07) et 2 contrats d'agent contractuel (05/06 à 08/06) et (07/07 à 10/07)
- **A.T.E.R.** (50%) à la faculté de pharmacie et au laboratoire TIMC-IMAG, Grenoble 2007-2008
- **A.T.E.R.** (100%) à la faculté de pharmacie et au laboratoire TIMC-IMAG, Grenoble 2008-2009
- **MCF** (section 27) à Polytech, actuellement rattaché au laboratoire TIMC, CNRS UMR5525 2009- ...

CURSUS UNIVERSITAIRE

- Équivalence de DEUG 1^{ère} année par **Médecine 1^{ère} année (PCEM 1)** 1993-1995
Faculté de Médecine de la Timone, Aix-Marseille II.
- **DEUG B (SNV) dominante Chimie Organique** 1996
Université de Provence, Aix-Marseille I.
- **Licence et Maîtrise de Biochimie dominante Immunologie** 1996-1998
Université de la Méditerranée, Aix-Marseille II.
- **DEA Modèles et Instruments en Médecine et Biologie** 1999
Université Joseph Fourier, Grenoble I.
- **Thèse de Doctorat en Biophysique** 1999-2002
ICH-DRDC, INSERM U548, CEA Grenoble.
TIMC-IMAG, CNRS UMR5525, Université J. Fourier (UJF), Grenoble I.
Effet des champs externes faibles sur l'auto-organisation des microtubules. Comparaison entre expériences et simulations numériques.
Directeurs de Thèse : Jacques Demongeot (TIMC-IMAG) et James Tabony (DRDC, CEA)
Thèse soutenue le 10/12/2002

QUALIFICATIONS UNIVERSITAIRES

- Qualification en section 85 – session 2008 : *Personnels enseignants-chercheurs de pharmacie en*

sciences physico-chimiques et ingénierie appliquée à la santé (anciennement 39 Sciences physico-chimiques et technologies pharmaceutiques) : Biostatistiques et bioinformatique, Biophysique, Bioanalyse, métrologie des médicaments et autres produits de santé, Instrumentation analytique, Chimie physique appliquée à la technologie pharmaceutique et au médicament, Pharmacie galénique, pharmacotechnie, biopharmacie et biomatériaux

- Qualification en section 65 – Session 2006, *Biologie Cellulaire*

COMPÉTENCES POUR LA RECHERCHE

▪ Informatique

Confirmé en programmation objet C++ moderne, JAVA et OpenCL (C99 sur GPU), et des bibliothèques wxWidgets, GLFW, SFML, SDL, OpenGL moderne.

Bonne maîtrise des algorithmes pour l'infographie.

Confirmé aussi en langages divers comme R, et C et forte utilisation par le passé d'ASM x86, Turbo Pascal.

Utilisation régulière de quelques autres langages (notamment de programmation par contraintes comme PROLOG et ASP (Answer Set Programming)).

Très bonne culture des technologies et algorithmes bio-inspirés, des développements en vie artificielle, algorithmes pour la dynamique des systèmes complexes et collectifs, modélisation de dynamiques stochastiques, des principes théoriques associés à l'informatique (machine de Turing, lambda calcul ...), de calculabilité ..., de programmabilité et de structurabilité des processeurs (en particulier programmation structurale et auto-organisation). Je travaille notamment dans le domaine de la composition de réseaux Booléens.

Développement utilisé principalement pour l'implémentation de modèles mathématiques en vue de simulations numériques, et pour la création d'outils d'analyse et de traitement des données (signaux, images, analyse de systèmes multi-agents ...). La programmation objet est adaptée à la représentation des objets et systèmes que je simule. De plus, ma thématique de recherche principale en computation naturelle m'a permis de considérablement approfondir ma maîtrise des fondements théoriques de l'informatique et de la programmation structurale. Je pratique aussi beaucoup de programmation graphique, en tant qu'utilisateur d'une part pour la visualisation dans mes simulations, en tant que concepteur d'autre part pour la création d'un pipeline graphique (et à des applications graphiques) dédié à l'affichage d'informations tridimensionnelles sur un périphérique interfacé avec la langue. J'utilise R et Julia pour le traitement statistique de mes données.

▪ Biophysique, Biochimie, Biologie Cellulaire et Moléculaire

Techniques courantes de biologie moléculaire et de biochimie. Culture cellulaire (MCF7, HeLa, CHO ...), Élevage de drosophiles. Isolement, purification et marquage d'éléments cellulaires (nucléiques, protéiques).

Expériences biophysiques sur le cytosquelette. Effet des champs et facteurs externes sur le vivant.

Excellentes connaissances de la biophysique de l'adhésion cellulaire.

Compétences en biologie cellulaire développées principalement pour l'étude des microtubules et des fonctions associées in vivo et in vitro. Connaissances de biologie moléculaire acquises lors de l'élaboration d'un système de plasmides permettant de faire du 'double hybride' en cellules Eucaryotes.

▪ Techniques Optiques et Microscopie

Parfaite maîtrise de tous types de microscopie optique, notamment microscopie Polarissante : *Biréfringence, Polarisation circulaire, Differential Interference Contrast (DIC)*. Microscopie en fluorescence. Microscopie fond noir. Microscopie en contraste de phase. Micromapping de surfaces par microscopie à contraste interférentiel.

Traitements et amplification des images par des techniques numériques ou optiques.

Techniques de spectrophotométrie (dosages colorimétriques et turbimétriques, marquages fluorescents).

Techniques utilisées notamment pour l'observation de structures fines comme des microtubules ou des particules subcellulaires (billes de latex, chromosomes, ...), et pour l'observation cinétique et l'analyse de structures auto-organisées et du mouvements de particules. Également, analyse de nano et micro topologies de surfaces.

- **Microsystèmes**

Design et fabrication de surfaces microstructurées par photolithographie et casting. Technique de '*double bounding*' sur film sec pour la microfluidique.

Traitements des surfaces par dépôts, gravures plasma et traitements UV.

Analyse des énergies de surfaces de substrats et de leurs composantes. Analyses de rugosité.

- **Langues**

Espagnol et anglais couramment parlés et écrits (couramment pratiqués).

PRIX

- **Prix Pierre Delattre.** Meilleure présentation au '*XXIème Séminaire de la Société Francophone de Biologie Théorique (SFBT)*', Paris, mai 2001
- **Prix de l'Innovation Technologique du Defi H.** Je suis initiateur de ce projet et encadrant scientifique de l'équipe ARMIND de Polytech Grenoble qui a gagné le prix de ce défi pour le handicap organisé par Sogeti, Paris, mai 2013

5. ENSEIGNEMENT, ENCADREMENT ET ANIMATION SCIENTIFIQUE

L'enseignement est une activité indissociable de l'enseignant-chercheur. D'abord, de nombreux doctorants font de l'enseignement et les enseignants-chercheurs encadrant des thèses tiennent compte de la charge liée à la préparation des enseignements par leurs doctorants et aux cours qu'ils donnent. L'encadrant EC peut guider le doctorant dans ses premiers pas d'enseignement-recherche. Ensuite, parce que l'enseignement, c'est du temps pris sur la recherche ! En 2019-2020 par exemple, j'ai cumulé 349 h ETD d'enseignement, une aberration ! La place qu'occupent de plus en plus les enseignements et les charges administratives liées dans les UFR en sous-effectifs devrait davantage être prise en compte dans les réflexions des laboratoires (notamment dans celles précédant les évaluations HCERES). Il est indispensable qu'une complémentarité s'établisse entre l'EC et le doctorant dans le maintien d'une recherche : l'étudiant, c'est celui qui garde pied dans l'activité de recherche en cours et qui permet à l'EC de s'y reconnecter. Enfin, l'enseignement rentre au moins partiellement dans un processus de valorisation de la recherche : ma recherche, la façon dont je la pratique, transparait dans mes cours. Pour ces raisons, il me paraît inconcevable de ne pas rendre visible mon dossier d'enseignement à coté de celui de recherche.

ACTIVITES ADMINISTRATIVES ET ANIMATION DE LA RECHERCHE

- Responsable universitaire d'année 3 (1ère année du cycle ingénieur) à Polytech TIS sur les périodes 2011-2015, 2017-2020.
- Activité d'évaluation pédagogique :
 - Participation, chaque année depuis ma nomination au poste MCU à la sélection des candidats pour Polytech (toutes filières + sélections spécifiques TIS) : analyse et sélection des dossiers, auditions.
 - Expertise dans les audits des PFE et des projets d'année 5 de Polytech.
 - Participation chaque année (présidence des jury) aux soutenances des stages d'année 4 et 5 (environ 6 par an)
 - Président de jury du Baccalaureat en 2011.
- Activité d'évaluation scientifique :
 - **Expertise d'articles :**
 - Chapitre dans le livre *Approches symboliques de la modélisation et de l'analyse des systèmes biologiques*, ISTE, 2020
 - *Acta Biotheoretica* (revue à CL) : **8 articles de mathématiques et informatique pour la biologie** (2010,2015,2017,2019)
 - Actes du *Colloque ARIMA* (congrès à CL) : **1 article de mathématiques** (2014)
 - *Journal of Royal Society Interface* (revue à CL) : **1 article de biologie théorique** (2013)
 - *Theory in Biosciences* (revue à CL) : **1 article de biologie théorique** (2013)
 - *Cytoskeleton* (revue à CL) : **1 article de biologie** (2015)
 - *IEEE Transactions on Systems, Man, and Cybernetics--Part A: Systems and Humans* (congrès à CL) : **1 article de psychomotricité** (2011)
 - *IEEE ICC 2017 SAC Symposium Molecular, Biological, and Multi-Scale Communications Track* (congrès à CL) : **1 article d'informatique** (2017)
 - *Adv. Artif. Neural Networks* (revue à CL) : **1 article de neurosciences** (2012)
 - **Evaluations de M2, Thèses et HDR:**
 - Examinateur invité, thèse d'Arthur Douillet, UBO Brest, 2020
 - Diverses évaluations de thèse à mi-parcours et de M2.

- 2012, examinateur de l'**HDR** de P. Ballet, UBO Brest.
 - **Expertise de projets :**
 - 2010, rapporteur d'un projet d'**ANR SysComm**
 - 2014, rapporteur d'un projet d'électrochimie et de biophysique pour la Fondation des Sciences Tchèque (GACR)
 - **Comités de sélection :**
 - (2014) sur le poste CNU 27 N°0194 pour le LabSTIC / UFR Sciences de l'UBO, à Brest. Rapporteur de 4 dossiers.
 - (2014) sur le poste CNU 27 N°0050 pour le LIRIS / INSA Lyon. Rapporteur d'1 dossier.
 - (2016) sur le poste CNU 26 N°147 pour le TIMC-IMAG Grenoble. Rapporteur de 9 dossiers.
- Édition et direction d'ouvrages scientifiques :
- 2012-2013, éditeur invité de la revue *Acta Biotheoretica* (gestion de 18 articles soumis)
 - 2012-2013, direction d'ouvrage, livre « *Le Vivant Discret et Continu* » (avec A. Stephanou) aux Editions Matériologiques.
 - 2014-2015, éditeur invité de la revue *Acta Biotheoretica* (gestion de 10 articles soumis)
 - 2014-2015, direction d'ouvrage, livre « *Le Vivant Critique et Chaotique* » (avec A. Chauvière, A. Stephanou, E. Fanchon)
 - 2017-2018, éditeur invité de la revue *Acta Biotheoretica* (gestion de 12 articles soumis)
 - 2018-2019, éditeur invité de la revue *Acta Biotheoretica* (gestion de 11 articles soumis)
- Animation scientifique et Organisation d'événements:
- Animateur à la fête de la Science : « substitution visuelle et reconnaissance de formes » (avec F. Robineau), 2007.
 - Organisateur du séminaire Dautreppe de la Société Française de Physique (SFP) avec P. Tracqui, du 6 au 10 décembre 2010. Thème : "*Imagerie Médicale et Modélisation Multi-Echelle - Interface Physique Biologie*" - 16 conférenciers.
 - Organisateur de l'école de printemps et du séminaire de la SFBT avec A. Stéphanou et O. Bastien du 4 au 13 juin 2012.
Thèmes de l'école : "*Expérimentation Numérique et Systèmes Hybrides comme alternatives efficaces dans la compréhension du vivant*" et "*Quels enseignements en biologie théorique au XXIème siècle*" - 18 conférenciers.
Thème incitatif du séminaire : "*Robustesse, résistance et résilience dans les systèmes naturels*" - 30 conférenciers.
Total de 63 participants, Budget de 23 k€ pour l'école et 8k€ pour le séminaire. Soutiens financiers du CNRS, de l'IXXI, de l'EDISCE, soutien et participation des écoles doctorales EDISCE et MST2I de l'UJF, Grenoble, et SICMA de l'UBO, Brest.
URL : <http://sfbt-2012.imag.fr/> et <http://sfbt-2012.imag.fr/seminaire.html>
 - Organisateur de l'école de printemps et du séminaire de la SFBT avec A. Stéphanou, E. Fanchon et A. Chauvière du 21 au 28 mai 2014.
Thème de l'école : "*Le Vivant Critique et Chaotique*" - 13 conférenciers
Thème incitatif du séminaire : "*La Vie Synthétique et le point de vue de la Biologie Théorique*" - 11 conférenciers
35 participants, Budget de 13 k€. Soutiens financiers des GDR METICE et PHENIX.
URL : <http://sfbt-2014.imag.fr/> et <http://sfbt-2014.imag.fr/seminaire.html>
 - Organisateur du séminaire de la SFBT avec S. Ivanov du 11 au 13 juin 2018.
28 conférenciers.
34 participants, Budget de 7,6 k€.
URL : <http://sfbt-2018.imag.fr/>
 - Organisateur du workshop ATAC (Atelier sur la Théorie Ataviste du Cancer) avec S. Ivanov et A. Stéphanou du 7 au 9 juin 2022.

Budget : 7000 €

- Sociétés savantes et GDR
 - Membre de la SFBT depuis 2001. Membre du bureau de la **SFBT** sur les périodes 2010-2015 et 2017-...
 - Coordinateur du groupe **Biocomputing** du programme d'**Epigénomique**, Génomole d'Evry de 2007 à 2011. L'activité de l'atelier de ce groupe s'est arrêtée suite à la fin des activités (ateliers) du programme épigénomique.

ACTIVITES D'ENSEIGNEMENT ET PEDAGOGIQUES

- **Responsable d'année** en filière *Technologies d'Information pour la Santé*, 3ème année, à Polytech Grenoble, UJF / UGA, sur les périodes 2011 (biseau en 2010)-2015 et 2017-2020

Cette activité m'a pris un temps considérable (temps inévitablement déductible de mon temps de recherche), l'année universitaire TIS3 étant particulièrement délicate à gérer, d'une part parce que les étudiants, encore scolaires, rentrent dans le cycle ingénieur, qui plus est dans une filière interdisciplinaire dans laquelle ils peinent à retrouver leurs marques (ce que je les aide à faire), d'autre part par la nature de la filière (interdisciplinaire) et la composition en enseignements et enseignants associés (des MCU/Pr de nombreux CNU, mais aussi des médecins et PU, des industriels ...), tout un système hétéroclite dont il faut maintenir la cohérence et le bon fonctionnement. De plus, la quantité de tâches administratives et liées aux personnels (dont gestion du personnel enseignant et des volumes horaires associés, des EDT, des problèmes de salles...) a explosé ces 3 dernières années.

- **Direction de Master 2 MITI**, UFR de Médecine. Depuis début 2021, j'assume la codirection du Master Modélisation, Innovation Technologique et Imagerie en binôme avec Joris Giai, médecin.

Là aussi, cette activité (non rémunérée) me demande un temps considérable, notamment parce que j'ai entrepris de réformer intégralement ce master 2. Le master avait décliné ces dernières années faute d'une gestion efficace. Devant le besoin des laboratoires interdisciplinaires, TIMC en premier, mais également de la demande croissante dans les entreprises innovantes dans le domaine biomédicale de personnes aux compétences variées dans ces domaines à l'interface maths-info / physique / biomédecine, il a fallu repenser ce M2. Aidé de mes collègues chercheurs, enseignants chercheurs, médecins, de plusieurs laboratoires (TIMC, mais aussi Liphy), j'ai ainsi créé une dizaine d'UE allant de la modélisation en biologie computationnelle en passant par la biomécanique et la biologie systémique, jusqu'au traitement des données d'imagerie médicale ou au traitement statistique avancée et au design des études cliniques, l'ensemble devenant un tout cohérent dans lequel les étudiants apprennent la démarche scientifique dans le contexte interdisciplinaire à l'interface du monde biomédical. Le montage d'abord et la synchronisation de l'ensemble demande un travail colossal. Je suis pour cela allé chercher et convaincre les enseignants susceptibles de donner de tels enseignements. De plus, je passe beaucoup de temps dans le relationnel étudiant et la mise en réseau avec les entreprises et équipes de recherche pour les stages de M2 et pour les projets applicatifs tutorés que les étudiants ont à réaliser durant le premier semestre.

- **Enseignements universitaires**

Le volume horaire de mes cours et activités pédagogiques est compris entre 200 et 350 heures ETD par an, incluant :

- *Cours à Polytech (2009 - ...) niveaux L3 et M1.*
 - Mathématiques de tronc commun en 3ème année (de l'ordre de 100 h ETD par an) dont un

module de Mathématiques générales, un module d'Analyse de Fourier, un module de Calcul matriciel, un module de Probabilités.

- Traitement de l'information en filière *TIS (Technologies de l'Information pour la Santé)*, 3^{ème} et 4^{ème} année
 - Modélisation statistique en santé (responsable du module) : 69 h ETD par an (réparties sur 2 années).
 - Traitement des données multivariées : 6 h ETD par an (34 h en 2010, 36h en 2014).
- Traitement de l'information en filière *PRI (Prévention des Risques Industriels)*, 3^{ème} année
 - Statistique (responsable du module) : 26 h ETD.
- *Cours en L3 MIAGE (2016 – ...), UFR IM2AG*
Langages pour le Web : 40h ETD par an.
Algorithmique et Programmation Objet : 40h ETD par an.
Projets intégrateurs : 30h ETD par an.
- *Cours en Master 2 ISM, Ingénierie pour la Santé et le Médicament, UGA (2009 – 2020), UFR Médecine*
Statistiques : 8h ETD par an
Physique : 3h ETD par an
Initiation à la modélisation : 5h ETD par an
- *Cours en Master 2 Modélisation, Innovation Technologique et Imagerie, UGA (2021 – ...), UFR Médecine*
Biomécanique et biomatériaux : 5h ETD par an
Initiation à la modélisation : 12h ETD par an
Biologie computationnelle : 8h ETD par an
Bioinformatique et Biologie Systémique : 4 h ETD par an
- *Cours en Licence 3 Biothechnologies (2009 - 2014), UFR Pharmacie*
Statistiques (*Cours en Master 1 Biotech, Faculté de Pharmacie de Grenoble*) : 18 h ETD par an.
- *Cours en Pharmacie (faculté de pharmacie de Grenoble) niveaux L1 à L3.*
 - 1^{ère} année (2007-2009) :
 - Mathématiques (158 h ETD) dont : Analyse et mathématiques générales, Probabilités, Statistiques
 - Informatique bureautique : 52h ETD
 - Biophysique (spectroscopie, radioactivité, focométrie, optique et microscopie, polarimétrie, photoelectricité, viscosimétrie et réfractométrie) : 85h ETP
 - 2^{ème} et 3^{ème} année (2011 - 2014) : Statistiques (initiation et statistiques avancées) : 53 h ETD par an.
- *Autres :*
 - Formation continue de médecins en informatique et modélisation pour la biologie et la médecine (2h) (2002)
 - Expérimentation en informatique appliquée à la médecine (4h) en 3^{ème} cycle, Master 2 MIIMB (2008)

▪ Tutorats.

Dans le cadre de Polytech TIS, je suis chaque année tuteur d'au moins 2 étudiants en année 4 et 2 autres en année 5. Le manque de tuteurs disponibles fait que j'en prends souvent plus.

ACTIVITES D'ENCADREMENT SCIENTIFIQUE

Pendant ma thèse et mes deux premiers post-doctorats, j'ai participé à la formation technique (expérimentale) et théorique (mathématiques, fondements théoriques) de 2 étudiants en DEA, et d'étudiants en physique (formation technologique). Ces étudiants n'étaient pas sous ma responsabilité. Voici, ci-dessous la liste des étudiants que j'ai officiellement encadré, classés par niveau d'étude.

- **BTS.**
 - 2012 – 3 BTS Services Informatiques aux Organisations, « Création du site WEB de la SFBT ». Coencadrement avec A. Stephanou (TIMC-IMAG)
 - 2013 – 3 BTS SIO, « Création du site WEB de la SFBT » - Fin du projet. Coencadrement avec A. Stephanou (TIMC-IMAG)
 - 2015 – 3 BTS SIO, « Conception d'un simulateur multi-agent - calcul massivement parallèle » - Fin du projet.
 - 2016 – 6 BTS SIO, « *Development of a simulation system for biological modelling (first releases of MESS the simulator, and OVES its modeler) in C++* » et « *Prototyping of a graphical modeler for OVES.* »
 - 2017 – 2 BTS SIO, « *MESS implementation of a group-based hybrid model in epidemiology.* »
 - 2018 – 2 BTS SIO, « *Amélioration de l'ergonomie d'un logiciel de modélisation, simulation et analyse des réseaux biologiques.* »
 - 2021 – 2 BTS SIO, « Développement du système de gestion de projets et de la configuration de Network Designer »
- **Licence 3.**
 - 2012 - Roselyne Chauvin, un travail exceptionnel ouvrant sur un M2 et une thèse sur un ensemble de questions ouvertes : « Programmation par contrainte pour la détermination de réseaux d'automates booléens à seuil – Taxonomie et héritage ». Coencadrement avec Eric Fanchon (TIMC-IMAG, CNRS UMR5525) et Hedi Ben Amor (AGIM, CNRS FRE3405). L'étudiante a suivi ensuite le M2 Cognition de l'ENS Paris et est actuellement en thèse en Hollande.
 - 2013 – Zhuofan Xie. Stage de 3ème année de Polytech 3I (informatique industrielle) en double avec Bing Yi Liu sur le sujet « Contrôle et validation d'un système de substitution visuelle haute résolution ».
 - 2016 - Anita Ravat – Stage L3 Physique - Tuteurs : Eric Beaugnon et Nicolas Glade, « Effets des champs magnétiques intenses sur le potentiel hydrogène (pH) de solutions aqueuses »
 - 2018 - Corentin Dervieux – Stage L3 MIAGE - Tuteurs : Nicolas Glade et Laurent Trilling, « Refonte d'un logiciel de modélisation, simulation et analyse des réseaux biologiques. »
 - 2019 - Nasser-Eddine Monir – Stage L2 - Tuteurs : Nicolas Glade et Laurent Trilling, « Composition de fonctions dans les réseaux Booléens à seuil par programmation par contraintes. »
 - 2019 - Emre Aydin – Stage L3MIAGE - Tuteur : Nicolas Glade, « Design et développement d'une IHM pour le logiciel NetworkDesigner en JavaFX et Java. »
 - 2021 – Nicolas Marino – Stage L3MIAGE – Tuteur : Nicolas Glade, « Programmation de l'interface de conception des réseaux dans Network Designer en Java et JavaFX »
 - 2021 – Remy Defossez – Stage L3 ENS Paris Ulm – Tuteurs : Sergiu Ivanov et Nicolas

Glade, « Le mille-feuille évolutionnaire, une nouvelle formalisation de l'évolution »

- 2022 – Vincent Ferrari – Stage L3 ENS Paris Ulm – Tuteurs : Sergiu Ivanov et Nicolas Glade, « Contrôlabilité des systèmes complexes. Systèmes à membranes et réseaux booléens »
- **Master 1 et 2 et Stages d'ingénieurs.**
 - 2007-2008 - Hedi ben Amor. Projet de fin d'études de l'Ecole Nationale des Sciences de l'Informatique de Tunis, sur le sujet « *Etude de la robustesse dans les réseaux de neurones et les réseaux de régulation biologique* ». Ce travail a donné lieu à 2 publications et a permis d'obtenir une équivalence de Master 2 à l'UJF, Grenoble, puis à débouché sur une thèse. Coencadrement avec J. Demongeot (TIMC-IMAG, CNRS UMR5525).
 - 2010 - Abdessalem Chekhchoukh. Stage de Master 2 du master MOSIG de l'UJF, Grenoble, sur le sujet « *Électrostimulation linguale et perception visuelle 3D* ». Ce travail a donné lieu à 2 publications et a débouché sur une thèse. Coencadrement avec N. Vuillerme (TIMC-IMAG, CNRS UMR5525).
 - 2011 - Malik Goumidi. Stage de Master 2 du master MOSIG de l'UJF, Grenoble, sur le sujet « *Suivi de trajectoire tri-dimensionnelle par électro-stimulation linguale* ». Ce travail n'a pas été poursuivi sous la forme d'une thèse. Coencadrement avec Y. Payan (TIMC-IMAG, CNRS UMR5525).
 - 2013 – Bing Yi Liu. Stage de 4ème année d'école d'ingénieur, Polytech Grenoble, Filière 3I (informatique industrielle) sur le sujet « Contrôle et validation d'un système de substitution visuelle haute résolution ». Coencadrement avec O. Chenu (ingénieur, société Taxisense).
 - 2017 – Quoc-Trung Vuong (MOSIG Master, UGA, Grenoble) - Tuteurs : Laurent Trilling et Nicolas Glade, « Taxonomy and inheritance in thresholded boolean automaton networks inferred from logical constraints »
 - 2017 – Clément Hege (Ensimag, Grenoble, UGA, Stage IRL) - Tuteurs : Nicolas Glade et Sergiu Ivanov, « Étude par expérimentation numérique de la diversité et de la survie de formes de vie artificielles dans un environnement stratifié par sédimentation différentielle »
 - 2017 – Marion Pellicer (Polytech Grenoble, UGA, RICM4 - Ingénierie) - Tuteurs : Nicolas Glade et Angélique Stéphanou, « Développement d'un logiciel de simulation de tumeur pour la thérapie guidée par ordinateur - Ré-usinage du code vers du C++11/14 et intégration à un framework unifié »
 - 2018 – Urvan Christen (Ecole Polytechnique, Paris) - Tuteurs : Nicolas Glade et Laurent Trilling, « Evaluation of distance between Threshold Boolean Automaton Networks ». Grand prix du stage de recherche à l'école polytechnique.
 - 2018 – Rémi Segretain (MOSIG Master, UGA, Grenoble) - Tuteurs : Nicolas Glade et Laurent Trilling, « Inference, simulation et analyse des réseaux d'automates Booléens à seuil. »
 - 2019 – Hector Bouvier (MOSIG Master, UGA, Grenoble) - Tuteurs : Nicolas Glade et Laurent Trilling, « Modélisation des phénomènes spatiaux émergents aux origines du vivant. »
 - 2020 – Justine Garcia (M2 MITI, UGA, Grenoble) – Tuteur : Nicolas Glade, « Mesure des affinités spécifiques entre acides aminés et trinuécléotides par titrage microcalorimétrique et spectrofluorométrie ».
 - 2021 – Florian Birkner (M1 Phelma, INP, Grenoble) – Tuteurs : Rémi Segretain et Nicolas Glade, « Conception d'un système de calcul générique et parallélisé pour la simulation et l'analyse des réseaux formels dans Network Designer »

- 2022 – Vincent Ferrari-Dominguez (ENS Paris) – Tuteurs : Sergiu Ivanov et Nicolas Glade, « Contrôlabilité des systèmes dynamiques finis : réseaux booléens et P systèmes booléens »
- 2022 – Ai-Ling Nguyen Bonnet (M1 Centrale Supélec, Paris) – Tuteurs : Sergiu Ivanov et Nicolas Glade, « Recherche de chemins viables dans les mille-feuilles évolutionnaires »
- **Thèses**
 - 2008-2012 (thèse soutenue en juillet 2012) - Hedi Ben Amor. « *Méthodes numériques et formelles pour l'ingénierie des réseaux biologiques - Traitement de l'information par des populations d'oscillateurs, approches par contraintes et taxonomie des réseaux biologiques* ». Mention *excellent* au manuscrit et *exceptionnel* à la soutenance. Ce travail a donné lieu à 8 publications, 1 logiciel, 1 projet de recherche financé (IXXI), et de nombreuses communications. Coencadrement avec J. Demongeot (AGIM CNRS FRE3405). H. Ben Amor Travaille actuellement comme cadre à haut niveau de responsabilité chez Accenture Technology, dans l'IA, après avoir démarré sa carrière jusqu'en 2018 dans l'entreprise CoSMo à Lyon en modélisation en urbanisme.
 - 2011 – 2015 (thèse soutenue en octobre 2015) - Abdessalem Chekhchoukh. Thèse à mi-temps, l'étudiant étant salarié d'une entreprise (DeFacto, Moirans). « *Substitution visuelle par électro-stimulation linguale et perception en 3 dimensions* ». Coencadrement avec N. Vuillerme (AGIM, CNRS FRE3405) et Y. Payan (TIMC-IMAG, CNRS UMR5525). La soutenance est prévue pour fin novembre, début décembre 2014 (date non fixée). A. Chekhchoukh travaille actuellement comme cadre à forte responsabilité chez DeFacto, Moirans, société de technologies biomédicales.
 - 2012 - 2016 (thèse soutenue en juin 2016) – Hessam Hessami. « *Phénomènes critiques en épidémiologie: résistance, spatialité, dynamiques humaines - Des modèles aux applications* ». Encadrement par J. Demongeot (AGIM) et M. Rachdi (AGIM) de Nov. 2010 à Nov. 2012 ; Coencadrement par J. Demongeot (AGIM) et N. Glade (AGIM) de Nov. 2012 à Nov. 2013 ; Encadrement par N. Glade (TIMC-IMAG) à partir de Nov. 2013. La thèse a connu un conflit entre directeurs de thèse en 2012 et le sujet a changé. La thèse a été réalisée à mi-temps, le doctorant travaillant en parallèle comme salarié de plusieurs entreprises successivement au cours de la thèse.. Le changement de directeur de thèse est une recommandation de l'école doctorale EDISCE qui a autorisé que j'encadre la thèse seul étant donné les circonstances. Le sujet actuel (que j'ai donné et qui a été soutenu) date de novembre 2012. H. Hessami travaille actuellement comme Data Scientist chez Digital Factories, Servier Labs, à Paris.
 - 2019 – 2022 – Rémi Segretain. « *Complexité et évolution : étude des réseaux Booléens à signe* ». Encadrement par N. Glade (TIMC-IMAG).
- **Post-docs**
 - 2016 – Sergiu Ivanov - PI. Nicolas Glade - *Modélisation du cytosquelette de microtubules in vivo*.

APPARTENANCES SCIENTIFIQUES

- Membre des groupes **Biocomputing** et **SMABio** du programme d'**Epigénomique**, Génopole d'Evry 2007-2016
- Membre de l'**Institut Rhone-Alpin des systèmes complexes (IXXI)** depuis 2007

- **Complex Systems Society (CSS, now called IXXI)**. Membre depuis 2006
- **Société Francophone de Biologie Théorique (SFBT)**. Membre depuis 2001, Membre du bureau de 2010 à 2014 et 2017-...

COLLABORATIONS SCIENTIFIQUES ACTUELLES

- **P. Ballet**, Lab-STICC, Univ. Bretagne Occidentale, Brest
- **O. Bastien**, LPCV, CEA Grenoble.
- **E. Beaugnon**, LNCMI, CNRS Grenoble
- **H. Guillou**, Institut Néel, CNRS - UGA, Grenoble.
- **S. Ivanov**, Univ. Evry val d'Essonne
- **A. Stephanou**, TIMC-IMAG, CNRS – UGA, Grenoble
- **L. Trilling**, TIMC-IMAG, CNRS - UGA, Grenoble.

FINANCEMENTS DE RECHERCHE OBTENUS

Depuis 2004, j'ai écrit ou co-rédigé (je n'ai jamais participé sans écrire au moins une bonne partie de la demande) bon nombre de demandes de financement (RNSC, IXXI, Clusters et Région, Appels d'offre CNRS dont PEPS, ANR, NoE, ERC Starting Grants ...). Je n'indique ici que les financements que j'ai obtenu ces dernières années (relatifs à mes activités de recherche actuelles). Au delà de ce qui est obtenu, j'ai fait de très nombreuses demandes, les dernières en date étant (i) une ANR 2013 (ANR SysVac avec F. Villier sur le thème de l'empoisonnement des végétaux par les métaux lourds), (ii) deux PEPS (EcoSanté 2014 et 2017) et une demande à l'INS2I en 2014 sur le thème de l'épidémiologie, (iii) une demande de bourse de thèse d'excellence UJF (2014), 2 projets AGIR 2013 (dont un dont je suis porteur), une demande à l'IXXI en 2014, et une demande AGIR 2014-2015 faites avec l'IAB, Grenoble, sur la modélisation de l'organisation du cytosquelette dans les plaquettes sanguines, puis (iv) 2 demandes, une à l'ANR 2016, une PEPS 2016 sur l'effet des champs magnétiques intenses sur le vivant, (v) un PEPS blanc (2018) et un CNRS Defi (2018) sur la modélisation dans le cadre de la théorie ataviste du cancer.

J'ai connu de nombreux échecs sur ces demandes et l'analyse que l'on peut en faire est assez simple : je suis encore trop souvent isolé sur les thématiques soumises, malgré la recherche et l'agrégation de partenaires extérieurs à ces projets, mais surtout ces thématiques sont clairement transdisciplinaires là où les financements sont souvent attribués à des projets inter-disciplinaires.

Les financements obtenus sont :

- **IXXI** sur la thématique « **Biocomputing** » (2007). Je suis porteur de ce projet mené en collaboration avec O. Bastien (équipe Chémogénomique comparative, LPCV, DRDC, CEA Grenoble). Son intitulé est « Formalisation et quantification de l'auto-organisation des *trail systems* fibrillaires ». Nous avons obtenu 5000 €.
- **IXXI** sur la thématique « **MicroARNs** » (2007). J'ai imaginé et écrit ce projet mais n'en suis pas le porteur (déjà porteur du préprojet Biocomputing). Son intitulé est « Rôle des microARNs en tant que clefs potentielles de la coévolution dans les écosystèmes viraux ». Nous avons obtenu 5000 €. Ce projet à l'interface entre les mathématiques et la biologie a été développé pour renforcer une collaboration entre le TIMC (avec J. Demongeot), le laboratoire de virologie moléculaire et structurale (UVHCI), EMBL Grenoble (avec E. Drouet) et l'université de Valparaiso au Chili (avec A. Moreira).
- **Réseau d'Excellence Européen – Virtual Physiological Human (NoE-VPH)** (2008). J'ai une participation active dans ce réseau d'excellence qui permet notamment le financement de l'étudiant en thèse, H. Ben Amor, que j'ai co-encadré. Notre participation au réseau concerne le développement

d'outils et de méthodes informatiques pour la modélisation, la simulation et l'analyse de réseaux de régulation de Hopfield (automates Booléens à seuil), potentiellement stochastiques.

- **RNSC** sur la thématique « **Origines du vivant** » (2009). Je suis porteur de ce projet mené en collaboration avec P. Ballet (Laboratoire LISyC, UBO Brest). Intitulé : « *préBioDyn – Expérimentation numérique et phénomènes émergents en chimie prébiotique* ». Nous avons obtenu 4000 €.
- **CSVSB** (financement du pôle Chimie, Sciences du Vivant) sur la thématique « **Substitution visuelle** » (2010-2011). Je suis porteur de ce projet mené en collaboration avec N. Vuillerme (AGIM, CNRS FRE3405, UJF) et Y. Payan (TIMC-IMAG, CNRS UMR5525). Son intitulé est « Substitution Visuelle Linguale en Trois Dimensions ». Nous avons obtenu 25 k€.
- **IXXI** sur la thématique « **Réseaux de régulation** » (2010). Je suis porteur de ce projet qui a été imaginé et monté par H. Ben Amor, alors étudiant en thèse sous ma responsabilité. Son intitulé est « Approche déclarative pour la modélisation de réseaux de régulation biologiques ».
- **IXXI** sur la thématique « Origines du vivant » (2010). Je suis porteur de ce projet mené en collaboration avec P. Ballet (Laboratoire LISyC, UBO Brest), O. Bourgeois et H. Guillou (MCBT, CNRS, Institut Néel, Grenoble). Son intitulé est « *preBioDyn - Origine du code génétique et phénomènes émergents en chimie prébiotique* » (il a le même titre que le projet RNSC de 2007. Il s'agit de la suite). Nous avons obtenu 5000 €.
- **FRM** – Projet **Platon** porté par K. Sadoul (IAB Grenoble) – TIMC-IMAG a obtenu un financement de postdoc sur un an (modélisation de l'organisation du cytosquelette dans les plaquettes sanguines) et 30k€ de matériel de microscopie.
- **IDEX** – j'ai activement participé au montage du **CDP COOL** (CDP – Origins Of Life) sur les origines du vivant et des planètes en 2017, dans le cadre de l'IDEX UGA. Le CDP COOL porté par X. Delfosse a été obtenu.
- **IDEX CDP COOL** – Dans le cadre du Cross Disciplinary Program COOL (CDP Origins of Life) de l'IDEX, Université Grenoble Alpes, un appel à projets a été proposé. J'ai obtenu 2 financements en tant que porteur de 8k€ chacun sur la période 2018-2022 :
 - Projet expérimental *Confirmation expérimentale de la possible origine stéréochimique du code génétique*, N. Glade (TIMC-IMAG) et E. Beaugnon (LNCMI Grenoble), 8 k€
 - Projet de modélisation *Modélisation des phénomènes spatiaux émergents aux origines du vivant*, N. Glade (TIMC-IMAG) et O. Bastien (LPCV, CEA Grenoble)
 - j'ai restitué 5000€ au CDP pour le soutien des prolongations de thèses retardées par Covid.
- **IDEX IRS** – sur la thématique Réseaux Booléens et régulation génétique (*Inférence de réseaux Booléens à seuil par programmation logique par contrainte pour modéliser l'homéostasie cellulaire et tester la théorie atavistique du cancer*), j'ai obtenu, en tant que porteur, une allocation doctorale de recherche en 2019 (thèse de Rémi Segretain) avec un support de 5k€.
- **IXXI** – *Test de la théorie ataviste du cancer par l'expérimentation numérique*. Ce projet financé fin 2019 à hauteur de 2500 euros est dédié à l'organisation d'une conférence sur le cancer et ses approches de modélisation.
- **INS2I – Projet ATAVISME** - *Modélisation de la bascule physio-pathologique dans le cadre de la théorie ataviste du cancer par des systèmes dynamiques et des réseaux Booléens*. Ce projet obtenu en 2021 est financé à hauteur de 10000 euros.
- **IXXI** – *Exploration de l'évolutivité et de la robustesse des réseaux biologiques – Caractérisation des composantes connexes structurelles et dynamiques des réseaux Booléens à signes et contraintes de viabilité*. Ce projet obtenu en 2021 est financé à hauteur de 5000 euros.
- **UGA - Soutien financier à l'organisation d'une manifestation scientifique** – Soutien pour l'organisation de l'Atelier sur la Théorie Ataviste du Cancer organisé en juin 2022. Ce projet obtenu en

2022 est financé à hauteur de 1700 euros.

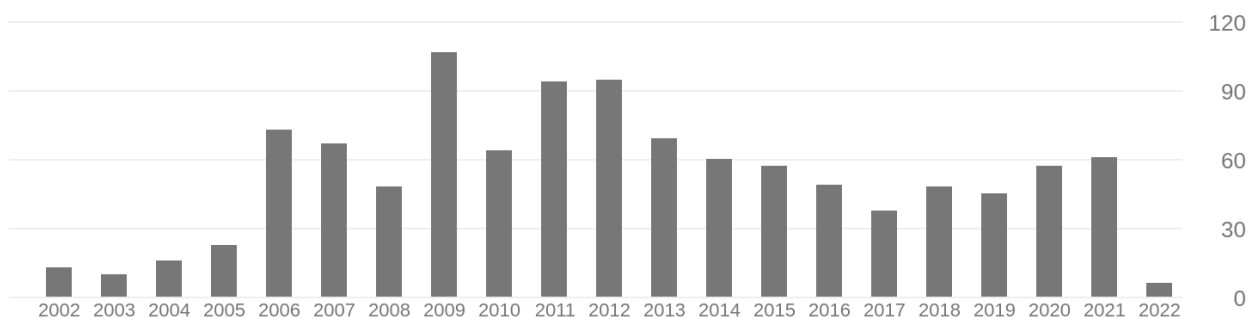
6. PRODUCTION SCIENTIFIQUE

METRIQUE

Je suis auteur et co-auteur dans 60 publications écrites dont 34 dans des revues avec comité de lecture, 13 actes de conférence avec comité de lecture, 4 actes de conférence sans comité de lecture, 3 chapitres de livres, 5 éditoriaux pour des revues et des livres et 1 article de vulgarisation. Elles sont réparties équitablement entre d'une part des journaux de mathématiques, biologie théorique, bioinformatique et sciences computationnelles et, d'autre part, des revues de biologie, biochimie, et physique ; j'ai également 4 articles en biotechnologie. Je suis 1^{er} ou dernier auteur dans 29 d'entre elles. J'ai une contribution importante dans mes articles (souvent plus de 50 %), la plupart du temps comme principal protagoniste et initiateur (principalement depuis 2006). Je suis aussi inventeur d'un brevet en technologie et co-auteur de 3 logiciels sous licence.

La plupart de mes publications attestent de mon inter-disciplinarité. J'ai des activités de recherche en STIC (math-info), avec des publications dans les Comptes Rendus Mathématiques, mais aussi en biologie théorique (Acta Biotheoretica notamment) et en *computation* non conventionnelle. Mes travaux de recherche actuels visent à renforcer ma participation à la recherche en informatique et en technologies pour la santé. L'autre moitié de mes travaux est davantage orientée biologie, biophysique et biotechnologie. Actuellement, je continue de travailler et de publier dans toutes mes thématiques de recherche. Une bonne partie de mes publications est le fruit de collaborations que j'ai initiées.

La figure ci-dessous est un rapport de citations généré par Google Scholar en mars 2022. Mon h-index avec le *Publish or Perish* (utilisant *Google Scholar*) de Harzing est de 20 (18 sans auto-citations). Ces travaux sont régulièrement cités à hauteur d'environ 40-50 citations par an sans auto-citations.



PRODUCTION SCIENTIFIQUE ECRITE

Publications dans des revues avec comité de lecture (35)

- Glade N., Tabony J., *The effect on microtubule self-organisation of reduced gravity levels produced in ground based experiments; comparison with space flight experiments*, J. Phys. IV, France (2001) **6**, 255-260.
- Tabony J., Glade N., Papaseit C., Demongeot J., *The effect of gravity on microtubule self-organisation*, J. Phys. IV, France (2001) **11**, 239-246.
- Tabony J., Glade N., Papaseit C., Demongeot J., *Microtubule self-organisation and its gravity dependence*, Chapter in: 'Advances in Space Biology and Medicine: Cell biology and biotechnology in

- space' (2002), Elsevier Science, A. Cogoli (Ed.) 8, pp 19-58. ISBN: 0444507353
- Tabony J., Glade N., Demongeot J., *Microtubule self-organization: a biological exemple of 'emergent' phenomena in a 'complex' system*, Recent. Res. Devel. Biophys. Chem. (2002) **3**, 11-53
 - Glade N., Demongeot J., Tabony J., *Comparison of reaction-diffusion simulations with experiment in self-organised microtubule solutions*, C. R. Biologies (2002) **325**, 283-294.
 - Tabony J., Glade N., Demongeot J., Papaseit C., *Biological self-organization by way of microtubule reaction-diffusion processes*, Langmuir (2002) **18:19**, 7196-7207.
 - Glade N., Demongeot J., Tabony J., *Numerical simulations of microtubule self-organisation by reaction-diffusion*, Acta Biotheoretica (2002), **50:4**, 239-268.
 - Tabony J., Glade N., Papaseit C., Demongeot J., *Gravity dependence of microtubule preparations*, J. Gravitational Physiology (2002) **9**, pp 10-14.
 - Demongeot J., Bézy-Wendling J., Mattes J., Haignon P., Glade N., Coatrieux J.-L., *Multiscale modeling and imaging: The challenges of biocomplexity*, Proc. IEEE (2003) **91**, 1723-1737.
 - Tabony J., Glade N., Papaseit C., Demongeot J., *Microtubule self-organisation as an example of the development of order in living systems*, J. Biol. Phys. Chem. (2004) **4**, 50-63
 - Glade N., Demongeot J., Tabony J., *Microtubule self-organisation by reaction-diffusion processes causes collective transport and organisation of cellular particles*, BMC Cell Biol. (2004) **5:23**.
 - Vulto P., Glade N., Altomare L., Bablet J., Medoro G., Chartier I., Manaresi N., Tartagni M., Guerrieri R., *Microfluidic channel fabrication in dry film resist for production and prototyping of hybrid chips*, Lab on a Chip (2005) **5**, 158-162.
 - Glade N., Tabony J., *Brief exposure to magnetic fields determine microtubule self-organisation by reaction-diffusion processes*, Biophys. Chem. (2005) **115**, 29-35.
 - Cortes S., Glade N., Chartier I., Tabony J., *Microtubule self-organisation by reaction-diffusion processes in miniature cell-sized containers and phospholipid vesicles*, Biophys. Chem. (2005) **120**, 168-177.
 - Glade N., Beaugnon E., Tabony J., *Ground-based methods reproduce space-flight experiments and show that weak vibrations trigger microtubule self-organisation*, Biophys. Chem. (2006) **121**, 1-6.
 - Demongeot J., Glade N., Forest L., *Lienard systems and potential-hamiltonian decomposition. I Algorithm*, C. R. Mathématique (2007) **344**, 121-126.
 - Demongeot J., Glade N., Forest L., *Lienard systems and potential-hamiltonian decomposition. II Methodology*, C. R. Mathématique (2007) **344**, 191-194.
 - Glade N., Forest L., Demongeot J., *Lienard systems and potential-hamiltonian decomposition. III Applications*, C. R. Mathématique (2007) **344**, 253-258 (IF: 0.462).
 - Forest L., Glade N., Demongeot J., *Lienard systems and potential-hamiltonian decomposition. Applications in biology*, C. R. Biologies (2007) **330**, 97-106.
 - Sordel T., Kermarec-Marcel F., Garnier-Raveaud S., Glade N., Sauter-Starace F., Pudda C., Borella M., Chatelain F., Bruckert F., Picollet-D'hahan N., *Influence of glass and polymer coatings on CHO cell morphology and adhesion*, Biomaterials (2007) **27**, 1572-1584
 - Tabony J., Rigotti N., Glade N., Cortes S., *Effect of weightlessness on colloidal particle transport and segregation in self-organising microtubule preparations*, Biophys. Chem. (2007) **127**, 172-180
 - Demongeot J., Glade N., Hansen O., Moreira A., *An open problem: the inner mitochondrial membrane (IMM) as a free boundary problem*, Biochimie (2007) **89**, 1049-1057
 - Glade N., *Computing with the cytoskeleton: A problem of Scale*. Int. J. Unconventional Computing (2008) **4:1**, 33-44.
 - Demongeot J., Glade N., Moreira A., *Evolution and RNA relics. A system biology view*, Acta Biotheor. (2008) **56:1-2**, 5-25.

- Demongeot J., Glade N., Moreira A., Vial L., *RNA Relics and the Origin of Life*. Int. J. Mol. Sci. (2009) **10**, 3420-3441.
- Abbas L., Demongeot J., Glade N., *Synchrony in Reaction-diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves*, Phil. Trans. Roy. Soc. A (2009) **367**, 4829-4862.
- Ben Amor H., Glade N., Lobos C., Demongeot J., *The isochronal fibrillation: characterization and implication in biology*, Acta Biotheoretica (2010) **58**, 121-142.
- Norris V., Zemirline A., Amar P., Audinot J. N., Ballet P., Ben Jacob E., Bernot G., Beslon G., Cabin A., Fanchon E., Giavitto J. -L., Glade N., Greussay P., Grondin Y., Foster J. A., Hutzler G., Kepes F., Michel O., Molina F., Signorini J., Stano P., Thierry A. R., *Computing with bacterial constituents, cells and populations: from bioputing to bactoputing*, Theory in Biosciences (2011) **130**, 211-228.
- Glade N., Ballet P., Bastien O., *A stochastic process approach of the drake equation parameters*, Int. J. Astrobiology (2012) **11**, 103-108.
- Chekhchoukh A., Glade N., *Influence of sparkle and saccades on tongue electro-stimulation-based vision substitution of 2D vectors*, Acta Biotheor. (2012) **60**, 41-53.
- Glade N., *On the nature and shape of tubulin trails : Implications on microtubule self-organization*, Acta Biotheor. (2012) **60**, 55-82.
- Ben Amor H., Corblin F., Fanchon E., Elena A., Trilling L., Demongeot J., Glade N., *Formal Methods for Hopfield-like networks*, Acta Biotheor. (2013) **61**, 21-39.
- Glade N., Bastien O., Ballet P., *Diversity and survival of artificial lifeforms under sedimentation and random motion*, Theor. Biosci. (2017) doi: 10.1007/s12064-017-0254-1
- Christen U., Ivanov S., Segretain R., Trilling L., Glade N., *On computing structural and behavioural complexities of threshold Boolean automata networks - Application to biological networks*, Acta Biotheor. (2019) **68**, 119-138.
- Alhazov A., Ferrari V., Freund R., Glade N., Ivanov S., *A P Systems Variant for Reasoning about Sequential Controllability of Boolean Networks ?*, Theor. Comput. Sci. (2023) in press [BioArXiv : <https://arxiv.org/abs/2303.00110>]

Publications dans des actes de congrès avec comité de lecture (13)

- Vulto P., Glade N., Altomare L., Bablet J., Medoro G., Leonardi A., Romani A., Chartier I., Manaresi N., Tartagni N., Guerrieri R., *Dry film resist for fast fluidic prototyping*, Micro Total Analysis Systems, Proceedings of the μ TAS 8th conference, Special Publication – Royal Society of Chemistry (2004), **297**, 43-45
- Glade N., *Existence and persistence of microtubule chemical trails – A step toward microtubule collision-based computing*. In A. Adamatzky & C. Teuscher (eds): From Utopian to Genuine Unconventional Computers (Splendeurs et misères du calcul peu usuel), UC'06 Workshop proceedings. Luniver Press, UK (2006), 37-65. ISBN: 0955117097. (ranking confs info : dans les premiers 30%)
- Elena A., Ben-Amor H., Glade N., Demongeot J.: *Motifs in Regulatory Networks and their Structural Robustness*. IEEE Transactions on Information Technology in Biomedicine, 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. BIBE 2008., Athens : Grèce (2008). (ranking confs info : dans les premiers 50%)
- Ben-Amor H., Glade N., Demongeot J., *Mnesic Evocation: An isochron-based analysis*, IEEE AINA' 10 & BLSMC' 10, Perth, Australia, IEEE Proceedings, Piscataway, 745-750 (2010)
- Chekhchoukh A., Vuillerme N., Glade N., *Vision substitution and moving objects tracking in 2 and 3 dimensions via vectorial electro-stimulation of the tongue*, Proceedings of the ASSISTH'2011 : 2^{ème} Conférence Internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situations de

- handicaps, "De l'usage des STIC à une plus grande autonomie : des recherches interdisciplinaires", Vigouroux N., Gorce P., Abraham M. and Piette F. (eds), Paris, France, 283-297 (2011)
- Glade N., Elena A., Corblin F., Fanchon E., Demongeot J., Ben-Amor H., *Determination, optimization and taxonomy of regulatory networks - The example of Arabidopsis thaliana flower morphogenesis*, IEEE AINA' 11 & BLSMC' 11, Singapore, IEEE Proceedings, Piscataway (2011)
 - Chekhchoukh A., Vuillerme N., Payan Y., Glade N., *Effect of saccades in tongue electrotactile stimulation for vision substitution applications*, Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13)
 - Chekhchoukh A., Goumidi M., Vuillerme N., Payan Y., Glade N., *Electrotactile vision substitution for 3D trajectory following*, Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13)
 - Vuong Q.-T., Chauvin R., Ivanov S., Glade N., Trilling L., *A logical constraint-based approach to infer and explore diversity and composition in thresholded Boolean automaton networks*, [Studies in Computational Intelligence Series](#), Proceedings of the Complex Networks 2017 conference (2018) DOI: 10.1007/978-3-319-72150-7_46
 - Manificier I., Chauvière A., Verdier C., Chagnon G., Cheddadi I., Glade N., Stéphanou A., A minimal biomechanical model for random cell migration, *Comp. Methods Biomech. Biomed. Eng. - Issue sup1: 45th Congress of the Société de Biomécanique*, (2020) **23**, pp S183-S185, DOI : <https://doi.org/10.1080/10255842.2020.1813420>
 - Segretain R., Ivanov S., Trilling L., Glade N., Implementation of a Computing Pipeline for Evaluating the Extensibility of Boolean Networks' Structure and Function, *bioRxiv* (2020) DOI: 10.1101/2020.10.02.323949 [Complément de l'article *A Methodology for Evaluating the Extensibility of Boolean Networks' Structure and Function*, *Proceedings of the Complex Networks 2020 conference* ayant été reviewé avec l'article]
 - Segretain R., Ivanov S., Trilling L., Glade N., A Methodology for Evaluating the Extensibility of Boolean Networks' Structure and Function, *Complex Networks & Their Applications IX*, [Studies in Computational Intelligence Series](#), Proceedings of the Complex Networks 2020 conference (2020)
 - Segretain R., Trilling L., Glade N., Ivanov S., Who Plays Complex Music? On the Correlations Between Structural and Behavioral Complexity Measures in Sign Boolean Networks, Proceedings of the BIBE 21, IEEE International Conference on BioInformatics and BioEngineering, Kragujevac, Serbia (2021)

Publications dans des actes de congrès sans comité de lecture (4)

- Tabony J., Glade N., Papaseit C., Demongeot J., *Microtubule self-organisation as an example of the development of order in living systems*, In 'Modelling Complex Biological Systems in the Context of Genomics', Proceedings of the Evry Spring School (2004), P. Amar, J.-P. Cornet, F. Kepes, V. norris (Eds), pp 93-112
- V Norris, A Zemirline, P Amar, P Ballet, E Ben Jacob, G Bernot, G Beslon, E Fanchon, J-L Giavitto, N Glade, P Greussay, Y Grondin, J A Foster, G Hutzler, F Kepes, O Michel, G Misevic, F Molina, J Signorini, P Stano, A Thierry, *Bactoputing: computing with bacteria*, In 'Modelling Complex Biological Systems in the Context of Genomics', Proceedings of the Lille Spring School (2008), P. Amar, F. Kepes, V. Norris, B. Vandebunder (Eds), ISBN: 978-2-7598-0075-9
- Norris V., Amar P., Aimar M., Ballet P., Batto A.-F., Barlovatz G., Bernot G., Beslon G., Cabin A., Chevalier S., Delaune A., Delosme J.-M., Fanchon E., Gao H., Glade N., Grondin Y., Hernandez-Verdun D., Janniere L., Képès F., Lange C., Legent G., Loutelier-Bourhis C., Molina F., Orange N., Raine D., Ripoll C., Thellier M., Thierry A., Tracqui P., Zemirline A.: *Hyperstructures 2008-2009*, In 'Modelling

Complex Biological Systems in the Context of Genomics', Proceedings of the La Colle sur Loup Spring School (2009) P. Amar, F. Kepes, V. Norris (Eds) pp 71-84

- Glade N., Ben Amor H., Bastien O., *Trail Systems as fault tolerant wires and their use in bio-processors*, In 'Modelling Complex Biological Systems in the Context of Genomics', Proceedings of the La Colle sur Loup Spring School (2009) P. Amar, F. Kepes, V. Norris (Eds) pp 85-119

Vulgarisation scientifique (1)

- Tabony J. & Glade N., *L'union fait la forme*, La recherche (Nov. 2002) N°9 "Ordre et désordre"

Chapitres de livres (3)

- Tabony J. & Glade N., *Auto-organisation biologique et structures hors-équilibre: l'exemple des microtubules*, Book : Eléments de biologie à l'usage d'autres disciplines. De la structure aux fonctions (2003) EDP Sciences, pp 144-147. ISBN: 2868835872
- N. Glade, O. Bastien, H. Hessami, *Phénomènes critiques en épidémiologie: résistance, spatialité, dynamiques humaines - Des modèles aux applications*, chapitre dans *Le Vivant Critique et Chaotique*, A. Stephanou, A. Chauvière, E. Fanchon & N. Glade (Eds), Collection *Sciences et Philosophie*, Editions Matériologiques (2015) 64 pages
- Ribeiro T., Folschette M., Trilling L., Glade N., Inoue K., Magnin M., Roux O., *Les enjeux de l'inférence de modèles dynamiques des systèmes biologiques à partir de séries temporelles*, chapitre dans *Approches symboliques de la modélisation et de l'analyse des systèmes biologiques*, C. Lhoussaine, E. Rémy, <https://hal.archives-ouvertes.fr/hal-02634235> (2020) pp 95-137

Editoriaux (5)

- H. Berry, S. R. Thomas, J. Demongeot, S. Bernard, V. Volpert, N. Bessonov, N. Eymard, A. Tosenberger, A. Chauvière, H. Hatzikirou, M. Tektonidis, A. Deutsch, E. Promayon, J. Berro, S. Portet, P. Amar, P. Baconnier, E. Fanchon, S. Lespinats, P. Ballet, A. Jeannin-Girardon, A. Pothet, G. Misevic, A. Fronville, V. Rodin, A. Stéphanou, N. Glade, *Le vivant Discret et Continu*, A. Stephanou & N. Glade (Eds), Collection *Sciences et Philosophie*, Editions Matériologiques (2013) - éditorial
- Glade N. & Stephanou A., Editorial de *Proceedings of the XXXIInd Seminar of the French-Speaking Society for Theoretical Biology; Saint-Flour (Cantal), France, 10–13 June, 2012*, dans *Acta Biotheoretica* (2013) **61**, 289-290
- F. Varenne, M. Montévil, J. Demongeot, J.-P. Françoise, M. Lahutte-Auboin, B.R. Caré, A. LoVan, H. Berry, H.A. Soula, H. Hessami, N. Glade, S. Jacquir, B. Xu, S. Binczak, J.-M. Bilbault, V. Thomas-Vaslin, *Le Vivant Critique et Chaotique*, A. Stephanou, A. Chauvière, E. Fanchon, N. Glade (Eds), Collection *Sciences et Philosophie*, Editions Matériologiques (2015) - éditorial.
- Glade N., A. Chauvière, E. Fanchon & Stephanou A., Editorial de *Proceedings of the XXXIVth Seminar of the French-Speaking Society for Theoretical Biology; Saint-Flour (Cantal), France, 26–28 May, 2014*, dans *Acta Biotheoretica* (2015)
- Glade N., Cheddadi I., Ivanov S., Editorial de *Proceedings of the XXXVIIIth Seminar of the French-Speaking Society for Theoretical Biology; Saint-Flour (Cantal), France, 11–13 June, 2018*, dans *Acta Biotheoretica* (2020)

VALORISATION SCIENTIFIQUE, TRANSFERS DE TECHNOLOGIES, LOGICIELS

Brevet (1)

- Plissonnier M., Chartier I., Borella M., Glade N., *Procédé de localisation d'une espèce chimique ou biologique sur un substrat, microsysteme d'analyse et biopuce*. Brevet N°0407531 (07/07/2004)

Logiciels sous licence (4)

- ***NetworkDesigner***, Ben Amor H., Glade N., Logiciel sous licence GPL2.0, développé en C++ et QT4 dans le cadre du Network of Excellence *Virtual Physiological Human*. Ce logiciel permet le design graphique et la simulation de réseaux d'automates. Plusieurs types de formalismes peuvent être pris en compte (réseaux de Thomas, réseaux de Hopfield ...). Il comprend 2 modes : i) un mode de simulation qui génère d'une part une analyse dynamique des états du système, notamment des états asymptotiques et de leurs bassins d'attraction et d'autre part analyse la robustesse des attracteurs sous conditions stochastiques (machine de Boltzmann) ; ii) un mode de recherche automatique de réseaux par programmation par contraintes, grâce à une interfaçage avec le logiciel GNBox dont nous avons développé un module pour les réseaux de Hopfield.
- ***GNBox Module for Hopfield Networks***, Ben Amor H., Corblin F., L. Trilling, Fanchon E., Glade N., GNBox est un logiciel développé par F. Corblin dans le cadre du formalisme des réseaux de Thomas. Il utilise la programmation par contrainte (langage PROLOG pour GNBox) et la résolution des problèmes de satisfiabilité associés grâce à un solveur SAT. Il permet à partir d'un ensemble de clauses et contraintes sur la structure et la dynamique des réseaux, de définir l'ensemble de tous les modèles satisfiables. GNBox est utilisé principalement dans le domaine des réseaux de régulation. Ma contribution, avec H. Ben Amor, est d'avoir étendu ses possibilités aux réseaux de Hopfield (des automates booléens à seuil, Turing complets).
- ***DSAnalyser***, Ben Amor H., Glade N., Logiciel sous licence GPL2.0 développé en C++. Ce logiciel permet l'analyse des systèmes dynamiques présentant un cycle limite. Ses fonctionnalités sont la détermination du cycle limite (et de sa période), le calcul des trajectoires du système dynamique et des isochrons du système (fibration isochrone), le calcul de la PRC (phase response curve) et du maximum phase shift, la simulation de grilles d'oscillateurs couplés.
- ***PCGME : Gaussian Mixture Estimate on a Periodic Compact - A R-class Implementation***, Glade N., Module R sous licence GPL. Ce programme est une implémentation R d'une méthode pour estimer les densités de mélanges Gaussiens sur un compact périodique (comme des distributions d'angles sur le cercle trigonométrique). Il utilise une méthode fondée sur un algorithme EM (expectation-maximization) inspirée de celle de C. Tomasi présentée dans "*Estimating Gaussian Mixture Densities with EM - A Tutorial*".

PRESENTATIONS ORALES

- *Computing the structural and behavioral complexities and distances in threshold Boolean networks*, Séminaire TIMC-IMAG, 2019
- *Inference, simulation et analyse des réseaux d'automates booléens à seuil*, 38ème séminaire de la SFBT, St Flour, France (2018)
- *Multiscale and Hybrid Modeling of Communicable Diseases - Towards New Numerical Tools for the Control of Epidemics*, Vietnamese-French Joint Conference on Applications of Mathematics to Ecology, Bio-economics, Epidemiology and Healthcare, VIASM, Ha Noi, Viet Nam (Dec. 2016)
- *Direct and solvent-mediated effects of IMF on biological systems*, Séminaire donné au LNCMI, Grenoble (2016)
- *Illusions perceptives et effet miroir dans le contexte de la substitution visuelle*, 34ème séminaire de la SFBT, St Flour, France (2014)
- *Quelques propositions pour l'épidémiologie - Effet de la spatialisation des modèles sur criticité auto-organisée en épidémiologie*, Ecole de printemps de la SFBT, St Flour, France (2014)

- *Taxonomie et Composition des réseaux d'automates booléens à seuil*, Séminaire donné au LabSTIC à Brest (2014)
- *Modélisation hybride en épidémiologie : entre espace homogène et SMA*, Séminaire donné au LabSTIC à Brest (2014)
- *Détermination de réseaux d'automates booléens à seuil par programmation par contraintes: dégénérescence, taxonomie et héritage*, Journée de l'IXXI, Grenoble, France (2012)
- *Expérimentation Numérique et Systèmes Hybrides comme alternatives efficaces dans la compréhension du vivant*, Conférence d'ouverture de l'école 2012 de la Société Francophone de Biologie Théorique, St Flour, France (2012)
- *Etude de l'Influence de Procédés Hérités de la Vision sur des Taches Psychomotrices utilisant un Dispositif d'Electrostimulation Linguale dédié à la Substitution Visuelle*, Journée INSIS, Grenoble, France (2012)
- *préBioDyn - Biodiversité et survie dans un système de vie artificielle*, Journées RNSC, Roscoff, France (2011)
- *Vision substitution and moving objects tracking in 2 and 3 dimensions via vectorial electro-stimulation of the tongue*, ASSISTH'2011 : 2ème Conférence Internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situations de handicaps, Paris, France (2011)
- *Modélisation multi-échelle et méthodes formelles pour les réseaux de régulation biologique et les réseaux métaboliques*, Séminaire Dautreppe 2010, SFP, Grenoble, France (2010)
- *De l'agitation thermique et de la gravité pour une Vie - artificielle - plus intéressante*, Séminaire Dautreppe 2010, SFP, Grenoble, France (2010)
- *Electro-stimulation linguale et perception visuelle 3D – A la recherche d'une certaine autonomie pour les non-voyants*, 3ème conférence internationale de la Société Francophone de Biologie Théorique, Tunis, Auger P. & Ben Miled S. (eds), (2010)
- *Et si la gravitation et l'agitation thermique avaient agit de concert aux origines du vivant ?*, 3ème conférence internationale de la Société Francophone de Biologie Théorique, Tunis, Auger P. & Ben Miled S. (eds), (2010)
- *Modelling the Biological Richness : the Beauty and the Best - Robustness, Synchrony, Diversity and Completeness. (Invited)* LISyC, UBO, Brest, France (2009)
- *La fibrillation isochrone : développements récents et applications en biologie*. Journée Bio-Math en mémoire de Loïc Forest, Laboratoire mathématiques, INSA Rouen
- *Trail Systems as reliable bio-wires for biocomputing applications*. Séminaire sur invitation de P. Ballet au laboratoire LISyC, UBO, Brest (2009)
- *Robustness of single trail system-based wires for bio-processing architectures*. Conférence au 'Computability in Europe' (CIE), Heidelberg, Allemagne (2009)
- *Trail Systems as fault tolerant wires and their use in bio-processors*, Séminaire Bioputing, Programme Epigénomique, Evry, France (2008)
- *From Ants to Molecular Ants : Fibrillar Trail Systems as Adaptive Information Processors*, Deuxième conférence internationale de la SFBT, Winnipeg, Canada (2007)
- *The endless story of tubulin trails ... or how to recycle utopic ideas*, séminaire sur invitation de F. Nedelec dans le laboratoire EMBL, Heidelberg, Allemagne (2007)
- *Enhanced and localized adhesion of chemicals and living cells by UV or plasma helium functionalizing of polysiloxane PECVD deposited layers*, Séminaire donné sur invitation dans le laboratoire 'Adhésion cellulaire et inflammation' dirigé par P. Bongrand, INSERM U600, à Marseille, France (2007)
- *Microtubule-based computing. From genuine to utopian ... (?)*, Workshop "Utopian to Genuine

- Unconventional Computers (Splendeurs et misères du calcul peu usuel)” part of the UC'06 Unconventional Computing conference, University of York, UK (2006)
- *A molecular model to estimate the chemical communications between biological microtubules*, Séminaire donné à l'occasion du XXVI^{ème} séminaire de la Société Francophone de Biologie Théorique 'Caractérisation et Analyse de l'Hétérogénéité dans les Systèmes Biologiques', à St Flour, France (2006)
 - *L'informatique appliquée à la médecine et à la biologie*. Séminaire de formation continue (cours de 2 h) à l'Association des Médecins du Var (AmiVar), à Théoules, France (2002)
 - Spring Meeting of the Topical Team CNES *Vibrational phenomena under micro-gravity*, ESPCI Paris (2002)

POSTERS

- Chen B.¹, Glade N.², Beaugnon E.¹ (¹LNCMI CNRS UGA, Grenoble ; ²TIMC-IMAG CNRS UMR5525 - UGA, Grenoble), *Magnetic field effect on water wetting*, International Conference for Magneto-Science, Hefei, Chine (2019).
- Christen U.¹, Ivanov S.², Segretain R.¹, Glade N.¹, Trilling L.¹ (¹TIMC-IMAG CNRS UMR5525 - UGA, Grenoble ; ²IBISC, Univ. Evry val d'Essonne), *Searching for Viability Graphs of Biological Functions by Computing Distances between Thresholded Boolean Automaton Networks*, Présentation à l'Ecole Polytechnique, (2018), DOI: 10.13140/RG.2.2.18606.97603
- Vuong Q.-T.¹, Chauvin R.², Ivanov S.³, Glade N.¹, Trilling L.¹ (¹TIMC-IMAG CNRS UMR5525 - UGA, Grenoble ; ²Radboud Univ. Netherland ; ³IBISC, Univ. Evry val d'Essonne), *A logical constraint-based approach to infer and explore diversity and composition in thresholded Boolean automaton networks*, Complex Networks 2017 conference (2017), DOI: 10.13140/RG.2.2.13037.23520
- H. Hessami, N. Glade (TIMC-IMAG CNRS UMR5525 Grenoble), *Effect of spatialization on the catastrophic dynamics in an epidemiological model*, ECMTB, Göteborg, Suède, (2014)
- H. Hessami, N. Glade (TIMC-IMAG CNRS UMR5525 Grenoble), *A Study of Epidemiological Processes by Criticality Analysis*, 34^{ème} séminaire de la SFBT, St Flour France (2014)
- H. Hessami, J. Demongeot, N. Glade (AGIM/TIMB, CNRS FRE3405, Grenoble), *Epidemiology : Yesterday, Today and Tomorrow*, Grenoble Interdisciplinary Days 2013, First edition, Grenoble, France (2013)
- H. Hessami, J. Demongeot, N. Glade (AGIM/TIMB, CNRS FRE3405, Grenoble), *Caractérisation des processus épidémiologiques par analyse de criticalité. Quel rapport entre une avalanche et une maladie contagieuse?*, 4^{ème} conférence internationale de la SFBT, IRD - Université de Dakar, Sénégal (2013)
- A. Chekhchoukh ^{1,2}, Y. Payan², N. Vuillerme¹, N. Glade¹ (¹AGIM/AFIRM, CNRS FRE3405, Grenoble ; ²TIMC-IMAG/GMCAO, CNRS UMR5525, Grenoble), *Effect of saccades on electrotactile stimulation of the tongue for vision substitution applications*, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13), Osaka, Japon
- A. Chekhchoukh ^{1,2}, Y. Payan², N. Vuillerme¹, N. Glade¹ (¹AGIM/AFIRM, CNRS FRE3405, Grenoble ; ²TIMC-IMAG/GMCAO, CNRS UMR5525, Grenoble), *Electrotactile vision substitution applied to 3D trajectory following*, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13), Osaka, Japon
- H. Ben Amor ¹, F. Corblin ², A. Elena ¹, E. Fanchon ², L. Trilling ², J. Demongeot ¹, N. Glade ¹ (¹AGIM/TIMB, CNRS FRE3405, Grenoble ; ²TIMC-IMAG/BCM, CNRS UMR5525, Grenoble),

- Approches par contraintes et classification par critère de robustesse des réseaux de régulation génétique, 32^{ème} séminaire de la SFBT, St Flour, France (2012)
- A. Chekhchoukh ^{1,2}, Y. Payan ², N. Vuillerme ¹, N. Glade ¹ (¹ AGIM/AFIRM, CNRS FRE3405, Grenoble ; ² TIMC-IMAG/GMCAO, CNRS UMR5525, Grenoble), Comment spatialiser l'information visuelle sur un dispositif d'électro-stimulation linguale ? - Nouveaux développements, 32^{ème} séminaire de la SFBT, St Flour, France (2012)
 - P. Ballet ¹, O. Bastien ², N. Glade ³ (¹ Université de Brest, LISyC EA3883, Brest ; ² PCV, CNRS UMR5168, CEA, INRA, iRTSV, Grenoble ; ³ Université Joseph Fourier, AGIM/AFIRM, CNRS FRE3405, Grenoble), Sédimentation, agitation, turn-over et biodiversité dans des écosystèmes artificiels : la vie est belle, 31^{ème} séminaire de la SFBT, Autrans, France (2011)
 - A. Chekhchoukh ^{1,2}, Y. Payan ², N. Vuillerme ¹, N. Glade ¹ (¹ AGIM/AFIRM, CNRS FRE3405, Grenoble ; ² TIMC-IMAG/GMCAO, CNRS UMR5525, Grenoble), Comment spatialiser l'information visuelle sur un dispositif d'électro-stimulation linguale ?, 32^{ème} séminaire de la SFBT, Autrans, France (2011)
 - H. Ben-Amor, F. Corblin, E. Fanchon, L. Trilling, N. Glade, J. Demongeot, (TIMC-IMAG, CNRS UMR5525, Grenoble) *Constraint-based approach for Boolean networks*, 3^{ème} conférence internationale de la SFBT, Tunis, Tunisie (2010)
 - H. Ben Amor ¹, M. Block ², E. Maréchal ², N. Glade ², O. Bastien ², (²TIMC-IMAG/TIMB, CNRS UMR5525, Grenoble ; ²Laboratoire de Physiologie Cellulaire Végétale, UMR 5168 INRA-CNRS-CEA Grenoble - Université de Grenoble I, France) Implication of both positive and negative feedbacks in plastid envelope membrane galactolipid biosynthetic pathway, Biarritz, France (2010)
 - J-P. Alcaraz ¹, L. Liguori ², B. Stidder ², N. Glade ³, P. Cinquin ¹, D. K. Martin ¹, (¹TIMC-IMAG/GMCAO, CNRS UMR5525, Grenoble ; ² Fondation RTRA Nanosciences, Grenoble ³ TIMC-IMAG/TIMB, CNRS UMR5525, Grenoble) Characterization of diffusion in a free thin polyelectrolyte membrane, Nanobio Europe, Münster, Allemagne (2010)
 - N. Glade¹, H. Ben-Amor¹, O. Bastien², (¹TIMC-IMAG, CNRS UMR5525, Grenoble, ; ²CEA-PCV, Grenoble) *Trail Systems as fault tolerant wires and their use in bio-processors*, Epigenomic Program Spring School, La Colle sur Loup, France (2009)
 - N. Glade¹ & O. Bastien², (¹TIMC-IMAG, CNRS UMR5525, Grenoble, ; ²CEA-PCV, Grenoble) *On the self-organizing properties of trail systems and their possible use as structurally programmable bio-processors*, 3rd MPG-CNRS Joint Workshop on Systems Biology, Grenoble, France (2008)
 - N. Glade, *Multi-scale model of microscopic trail systems for natural computing studies*, Workshop Systembiology, Berlin, Allemagne (2007)
 - Borella M.¹, Plissonnier M. ¹, Glade N. ², Belmonte T. ³, (¹ CEA-LTME (Laboratoire des Technologies des Microsources d'Énergie), Grenoble ; ² DOPT, CEA-LETI, Grenoble; ³ LSGS (Laboratoire de Science et Génie des Surfaces) CNRS-UMR7570, Nancy), *Fonctionnalisation de surface de films minces de type polysiloxane pour une adhésion localisée de substances biologiques et chimiques*, Colloque IEACM-2 'Innovations dans l'élaboration et les Applications des Couches Minces', Nancy (2005)
 - N. Glade¹, J. Demongeot², J. Tabony¹, (¹ INSERM 548 CEA, Grenoble, et ² TIMC-IMAG, CNRSUMR5525, Grenoble), *Microtubule self-organisation: an example of emergent phenomena in a biological complex system*, 5th EMBL PhD Student International Symposium *Design of Life – Learning from Nature*, Heidelberg, Allemagne (2004)
 - P.Vulto¹, N. Glade², L. Altomare¹, J. Bablet², G. Medoro³, A. Leonardi¹, A. Romani¹, I. Chartier², N. Manaresi³, M. Tartagni¹, R.Guerrieri¹, (¹ ARCES, University of Bologna, ² CEA-LETI, ³Silicon Biosystems S.R.L.), *Dry film resist for fast fluidic prototyping*, MicroTAS, Malmö, Suède (2004)
 - A. Viallat¹, M. Faivre¹, M. Abkarian¹, B. Pepin-Donat¹, C. Vezy¹, N. Glade², (¹ Laboratoire de spectrométrie physique, CNRS-Université Joseph Fourier, Grenoble, ² CEA-LETI), *Vesicles and cells*

under shear flow: motion, dynamics of adhesion, Gordon conference: colloidal, macromolecules and polyelectrolyte solutions, Ventura, USA (2004)

- A. M. Bonnot¹, C. Filippini¹, M. Chaumont¹, L. Marty¹, S. Chenot¹, T. Fournier², J. Tabony³, N. Glade³, (¹ LEPES-CNRS, ² CRTBT-CNRS, Grenoble et ³ INSERM U548 CEA, Grenoble), *Self organized growth mode of self-assembled carbon nanotubes synthesized by hot filament assisted CVD*, '1st Workshop of the COST Chemistry Action D27' sur le thème *Origin of Life And Early Evolution*, Ravello, Italie.(2002)
- N. Glade, J. Tabony, *Microtubule self- organisation as an example of the development of order in living systems*, '1st Workshop of the COST Chemistry Action D27' sur le thème *Origin of Life And Early Evolution*, Ravello, Italie.(2002)
- N. Glade, J. Demongeot, J. Tabony, *Comparison of reaction-diffusion simulations with experiments in self-organised microtubule solutions*, Euroconférence SAFIN Self-Assembled Fibrillar Networks, Autrans (2002)
- N. Glade, J. Demongeot, J. Tabony, *Self-organisation of in-vitro microtubule solutions by reaction-diffusion processes. Comparison and dialogue between experiment and numerical simulation*, Colloque 'AcM/MdA Model Driven Acquisition', Grenoble (2001)
- N. Glade, J. Tabony, *Auto-organisation biologique par des processus proches du chaos*, Congrès 'Exobiologie et Origines du Vivant', La Colle sur Loup (2001)
- N. Glade, J. Tabony, *Rôle de la gravité dans l'auto-organisation des microtubules : comparaison entre apesanteur simulée au sol et vol spatial*, Colloque CNES 'Sciences de la Matière et Microgravité', ESPCI, Paris(2001)
- N. Glade, J. Demongeot, J. Tabony, *Auto-organisation de populations de microtubules par des processus de réaction-diffusion : expériences et simulations numériques*, Séminaire de la Société Française de Biologie Théorique, 'Les Niveaux d'Organisation en Biologie: Enjeux et Perspectives', Paris (2001)
- N. Glade, J. Demongeot, J. Tabony, *Structuration spatiale des réseaux de microtubules : un modèle fondé sur des processus de réaction-diffusion*, Colloque Dynacell Contrôle de la migration et des déformations cellulaires : approches théoriques et expérimentales, Grenoble (2001)
- N. Glade, J. Demongeot, J. Tabony, *Structuration spatiale des réseaux de microtubules: un modèle fondé sur les processus de réaction-diffusion*, Les 48 Heures de Biologie Structurale et Fonctionnelle, Autrans (1999)

ÉVÉNEMENTS SCIENTIFIQUES ET CULTURELS

- F. Robineau, O. Chenu, N. Glade, *Voir des formes électriques avec la langue : une application de la substitution sensorielle*, Fête de la Science 2007 à Grenoble, France. Financement de 500 €.
- A. Stephanou, O. Bastien, N. Glade (Organizing comitee), Thematic school of the SFBT on biological modeling 2012 : "Numerical experiments and hybrid systems as efficient modelling alternatives for biological understanding" - "Expérimentation Numérique et Systèmes Hybrides comme alternatives efficaces dans la compréhension du vivant". Sessions : "Information and signalling in biological systems" - "Information et signalisation dans les systèmes biologiques" ; "Self-organisation and morphogenesis" - "Auto-organisation et morphogénèse" ; "Quels enseignements pour la biologie théorique au XXIème siècle". Budget de 23 k€.
- A. Stephanou, O. Bastien, N. Glade (Organizing comitee), Annual conference of the SFBT on theoretical biology 2012 : "Robustness, resistance and resiliency in natural systems" - "Robustesse, résistance et résilience dans les systèmes naturels". Budget de 8 k€.
- A. Stéphanou, E. Fanchon, A. Chauvière, N. Glade (Organizing comitee) Thematic school of the SFBT on biological modeling 2014 : "Le Vivant Critique et Chaotique". Budget de 13k€.

- A. Stéphanou, E. Fanchon, A. Chauvière, N. Glade (Organizing comitee), Annual conference of the SFBT on theoretical biology 2014 : "La Vie Synthétique et le point de vue de la Biologie Théorique"
- S. Ivanov, N. Glade (Organizing comitee), Annual conference of the SFBT on theoretical biology 2018.
- S. Ivanov, N. Glade, A. Stéphanou (Organizing comitee), Workshop on the Atavistic Theory of Cancer (ATAC : Atelier sur la Théorie Ataviste du Cancer), 2022. Budget de 7000 €

PARTIE II.
RECHERCHE

7. TRAVAUX ANTERIEURS

Dans cette section, je vais présenter les travaux de recherche que j'ai menés par le passé et que je continue de mener pour certains de façon périodique. J'indiquerai par une astérisque entre crochets dans le titre [*] les projets qui font toujours l'objet d'une recherche ou que je serais susceptible de relancer avec un étudiant. Au delà de mes projets de recherches développés dans le cadre de cette HDR, plusieurs des travaux que j'ai menés constituent donc une large réserve d'activités futures dans lesquelles d'autres recherches peuvent être menées, y compris dans le cadre de thèses.

Pour ceux qui souhaiteraient continuer directement sur mes projets de recherche (section 8), voici un bref résumé de ce chapitre.

Pendant ma thèse et un 1er post-doc, j'ai étudié l'auto-organisation des solutions de microtubules. J'ai réalisé un modèle et j'ai étudié expérimentalement l'effet des champs externes faibles comme la gravité ou les champs magnétiques sur ces solutions. Ce travail s'est achevé avec le lancement d'une expérience en fusée sonde. Grâce à des recherches sur le thème du calcul (au sens computationnel) fondé sur des systèmes microscopiques fibrillaires multi-agents, j'ai ensuite repris mes travaux théoriques sur les microtubules, remettant en question certaines idées que j'avais développé durant ma thèse, en particulier l'auto-organisation par réaction-diffusion. J'ai également repris mes travaux expérimentaux avec Caterina Tomba et Eric Beaugnon concernant l'effet des champs magnétiques intenses sur le vivant, en particulier les microtubules.

Durant mon 2ème post-doc, au LETI, j'ai travaillé sur les phénomènes physico-chimiques d'adhésion cellulaire sur des substrats synthétiques. J'ai déposé un brevet concernant la localisation d'espèces chimiques et biologiques par des traitements d'énergies de surface. J'ai aussi mis au point une méthode de fabrication par films secs de résine de dispositif de microfluidique. Pendant cette même période précédant mon recrutement de MCU, j'ai participé à d'autres travaux, comme par exemple la décomposition potentielle-hamiltonienne de systèmes de Liénard ou encore la modélisation du réseau de régulation du système GSH-fer et celui de la mitochondrie.

J'ai ensuite encore diversifié mes travaux avec notamment une incursion dans la psychomotricité (travaux sur la substitution visuelle sur la langue) et en épidémiologie.

7.1. Travaux sur les microtubules

7.1.1. Prérequis

Les microtubules sont des assemblages biologiques supramoléculaires en forme de tubes d'environ 20 nm de diamètre et pouvant atteindre plusieurs dizaines de micromètres de long. Ils sont un des trois éléments fibrillaires qui constituent le squelette intracellulaire. Ils peuvent s'assembler à partir

de tubuline, leur brique élémentaire, ou se désassembler rapidement (de l'ordre du micromètre par minute). Ils interviennent dans de nombreuses fonctions cellulaires (organisation intracellulaire, division cellulaire, transport vésiculaire, cils et flagelles ...), souvent assistés par des protéines associées. De plus, le cytosquelette est le candidat le plus sérieux pour la transduction des champs externes faibles (gravité, champs magnétiques ...) dans la cellule.

En 1990, J. Tabony et D. Job, CEA Grenoble, ont montré que des solutions de microtubules ne contenant que de la tubuline purifiée et du GTP (guanosine tri-phosphate), un nucléotide énergétique permettant la réaction microtubulaire, pouvaient dans certaines conditions réactives former progressivement des structures striées, semblant auto-similaires sur plusieurs niveaux d'échelle (micro à macroscopique). Sur la base de la ressemblance de ces morphologies très ordonnées et compliquées et de l'aspect dissipatif de ces solutions, ils ont proposé que ces motifs se formaient par des processus de réaction-diffusion en accord avec la théorie morphogénétique de A. Turing, et des travaux sur la dynamique hors-équilibre de I. Prigogine. De très nombreuses expériences biophysiques sont allées dans le sens de cette hypothèse. L'équipe de Tabony a aussi prouvé la gravisensibilité du phénomène, celui-ci n'ayant pas lieu en apesanteur. En 1999 et jusqu'à récemment (Tabony, Biol. Cell 2006), J. Tabony a avancé que l'auto-organisation des microtubules pourrait être due, au niveau microscopique, à des processus collectifs dans lesquels les microtubules 'communiqueraient' entre eux par l'intermédiaire de variations locales de concentration en tubuline, conséquences de la réactivité de leurs extrémités réactives. Plusieurs travaux remettent maintenant en question cette hypothèse et invalident le modèle théorique d'auto-organisation spatiale des microtubules par réaction-diffusion (Baulin, J. Chem. Phys. 2003; Liu et al, PNAS 2006; Glade, Int. J. Unconventional Computing 2008). Cependant, si les processus de réaction-diffusion s'effacent devant les aspect biomécaniques dans le cadre de l'auto-organisation spatiale des microtubules, ils expliquent toutefois très bien les processus d'auto-organisation spatio-temporelle (vagues de concentration en microtubules) mis en évidence par l'équipe de Mandelkow (Mandelkow & Mandelkow, Cell Motil. Cytosk. 1992 ; Marx & Mandelkow, Eur. Biophys J. 1994) notamment et pourraient jouer un rôle important dans la transmission d'informations dans la cellule.

Ce sujet m'ayant occupé longtemps, je le développe particulièrement. Il est à la source d'un certain nombre de mes questionnements sur le rôle et de l'importance de l'ordre dans le vivant.

7.1.2. Thèse (1999-2002) : Effet des champs externes faibles sur l'auto-organisation des microtubules

Dépôt: <http://tel.archives-ouvertes.fr/tel-00005407/fr/>

Titre : (français) *Effet des champs externes faibles sur l'auto-organisation des microtubules. Comparaisons entre expériences et simulations numériques.* (anglais) *Weak external fields effect on microtubule self-organisation. Comparisons between experiments and numerical simulations*

Composition du Jury de thèse :

- Michel LAURENT (Pr.) - Président du jury et rapporteur de la thèse
Imagerie et Dynamique Cellulaires, UPRESA 8080, Bât. 440, Université Paris Sud, 91 405 ORSAY Cedex
- Patrick DE KEPPER (Pr.) - Rapporteur de la thèse
Centre de Recherche Paul Pascal, CRPP, CNRS UPR 8641, Avenue Albert Schweitzer, 33600 Pessac
- Augusto COGOLI (Dr.) - Examineur
Space Biology Group, ETH Technopark Zurich, Technoparkstr. 1, CH - 8005 Zurich - Switzerland
- Patrice MARCHE (Dr.) - Examineur
Département de Réponse et Dynamique Cellulaires, Laboratoire d'Immunochimie, DRDC / ICH /CEA Grenoble, C3 P450B, 17 rue des martyrs, 38054 Grenoble Cedex 9
- Pierre VIGNAIS (Pr.) - Examineur
DRDC, Laboratoire Biochimie et Biophysique des Systèmes Intégrés, DRDC / BBSI /CEA Grenoble, C3 P423, 17 rue des martyrs, 38054 Grenoble Cedex 9
- James TABONY (Dr.) - Directeur de thèse
DRDC / ICH /CEA Grenoble, D3 P200, 17 rue des martyrs, 38054 Grenoble Cedex 9
- Jacques DEMONGEOT (Pr.) - Directeur de thèse
Institut des Mathématiques Appliquées de Grenoble, Laboratoire des Techniques de l'Imagerie, de la Modélisation et de la Cognition, TIMC-IMAG, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche cedex

Durant ma thèse j'ai étudié la façon dont les champs externes faibles tels que la gravité, les champs magnétiques ou les champs électromagnétiques affectaient l'auto-organisation spatiale des microtubules. A ce moment, nous pensions que les solutions de microtubules s'auto-organisaient principalement à cause de processus de réaction-diffusion. Il était ainsi proposé que des populations de microtubules, très réactifs dans ces solutions, puissent s'auto-organiser du simple fait de leurs réactions et du couplage de celles-ci à travers l'espace de la solution par diffusion, d'une façon semblable à celle décrite par A. Turing dans son '*On the chemical basis of morphogenesis*' publié dans les Philos. Trans. Roy Soc. London en 1952. Bien que maintenant la majorité des personnes remettent en question cette interprétation (moi compris), ma thèse s'inscrivait à l'époque dans ce contexte qui avait peu d'alternatives étant donné les résultats expérimentaux qui appuyaient fortement cette hypothèse. Mes résultats expérimentaux ne sont en revanche pas remis en question. Ils montrent que les solutions de microtubules sont sensibles aux champs externes faibles : la présence, l'absence (des conditions d'apesanteur peuvent être obtenues dans des clinostats rotatifs ou en lévitation magnétique), la substitution (un champ vibratoire peut compenser des effets de l'apesanteur), ou des modifications de la gravité (lévitation magnétique, hypergravité et changement de la direction du vecteur gravité par centrifugation) font disparaître ou affectent la formation des

extraordinaires morphologies auto-organisées de microtubules (ces morphologies se présentent sous la forme de stries de concentration et d'orientation de 1 mm d'épaisseur et présentent aussi des niveaux d'organisation aux échelles inférieures). J'ai aussi montré dans quelle mesure les champs magnétiques orientent les microtubules. Mes travaux constituent d'ailleurs une mise en garde (fort bien ignorée d'ailleurs) contre l'utilisation de champs magnétiques trop intenses dans les appareillages d'imagerie magnétique (IRM) puisque des effets sérieux s'observaient déjà à partir de 4 Teslas (plusieurs scanners IRM médicaux fonctionnent maintenant à cette intensité). En tout cas, ils fournissent une base pour la compréhension cellulaire des effets biologiques qui seront probablement observés dans les études d'IRM en champs très intenses comme cela sera le cas dans le cadre de l'institut Neurospin (en fait certaines études ont déjà observé des effets mais n'ont pas identifié les causes). J'ai aussi étudié le mouvement de particules colloïdales entraînées par le processus d'auto-organisation et ai conçu un modèle d'auto-organisation des microtubules individu centré et fondé sur des processus de réaction (assemblage et désassemblage microtubulaire, régénération de la tubuline par échange nucléotidique) couplée à la diffusion des petites molécules (tubuline, nucléotides).

7.1.3. Premier post-doctorat (2002-2003) : (contrat ESA-CNES). Programme de lancement d'une expérience en fusée sonde de l'Agence Spatiale Européenne

J'ai pu avoir un contrat post-doctoral de 7 mois au laboratoire ICH-DRDC, au CEA Grenoble et sur la base de lancement spatial de l'ESRANGE, à Kiruna, Suède, sur un projet de *'biologie en apesanteur'* concernant l'étude de *l'effet de l'apesanteur sur l'auto-organisation des microtubules et sur les fonctions qui y sont associées*. J'ai participé à une expérience embarquée dans une fusée sonde MAXUS de l'Agence Spatiale Européenne, cofinancée par le Centre National d'Études Spatiales CNES (PI: James Tabony (ICH-DRDC, CEA Grenoble)) en collaboration étroite avec l'agence spatiale suédoise (SSC). J'étais financé par le CNES.

Depuis la moitié de ma thèse (en parallèle), j'ai participé à tous les stades de la mission: l'élaboration du projet initial, la gestion de la mission (protocoles de conception, de test, expérimentaux...), le développement du banc expérimental et du module, les tests, la campagne de lancement en Suède et l'analyse des résultats.

Ce module comprenait 3 expériences d'un total de 42 échantillons en cuves de verre et de plus de 10 plaques de 50 micro-échantillons, répartis dans 2 compartiments, l'un en apesanteur l'autre, témoin (en plus des témoins au sol), accéléré à 1 g terrestre par une roue tournante. Les échantillons et le matériel d'observation et de mesure devait pouvoir subir les conditions extrêmes du décollage (15 g) et de réentrée dans l'atmosphère (50 g), et devait être maintenu à 35 °C durant plusieurs heures après atterrissage. Durant le vol parabolique (apogée à 600 km), ils étaient soumis à 13 minutes d'apesanteur. Les 3 expériences consistaient à suivre (1) par microscopie (avec 2 microscopes polarisant embarqués portant chacun un objectif de 20x) le déplacement organisé et l'auto-ordonnement de particules colloïdales de latex de 0,8 µm de diamètre du au processus d'auto-organisation (ou son absence ou ralentissement en l'absence de pesanteur), dans chacun des 38

échantillons, (2) le niveau d'assemblage des microtubules par turbidimétrie UV (spectroscopie à 350 nm), (3) l'auto-organisation des microtubules dans des micro-échantillons de dimensions cellulaires (20 x 20 x 20 μm à 200 x 200 x 70 μm), ces échantillons étant observés 2h après le lancement, après récupération de la charge utile et extraction du module de vol.

Le retour de vol s'est mal passé : une mauvaise évaluation des conditions climatiques au point de ré-entrée a provoqué le crash (malgré le parachute) de la charge utile. Cela a provoqué une ouverture du blindage de la charge utile et la coupure du système de chauffage provoquant le refroidissement progressif des échantillons (l'air étant à -30°C à l'extérieur). Ceux qui étaient les plus sensibles à la température (les micro-échantillons) ont été perdus. Les autres (qui ont pu être maintenu à une température acceptable durant 2h grâce à une très bonne isolation du module) ont constitué des résultats partiels confirmant des expériences que j'ai faites au sol dans des clinostats. Ce travail a été publié dans [Cortès, 2006]. Il montre que dans les conditions d'apesanteur, l'auto-organisation de ces solutions de microtubules ne se fait pas en apesanteur et que les billes de latex ne se déplacent plus, et que les microtubules s'assemblent correctement, même en apesanteur. A noter que si les résultats sont irréprochables, je ne cautionne plus, depuis 2004, leur interprétation par des processus de réaction diffusion, l'article (ainsi que tous ceux copublicés avec J. Tabony en 2007) l'ayant été sans mon accord.

7.1.4. Remises en question sur le mécanisme d'auto-organisation des microtubules (2005-2007)

Au cours de mes travaux sur le calcul fondé sur les microtubules (voir plus bas et programme de recherche), j'ai commencé l'élaboration d'un nouveau modèle du cytosquelette microtubulaire. D'abord, j'ai conçu un modèle de dynamique microscopique, mais j'ai souhaité disposer d'un autre modèle s'adressant à des problèmes plus macroscopiques sans avoir besoin de représenter tous les réactifs individuellement. J'ai vite vu que je pouvais utiliser une architecture commune et concevoir un modèle multi-échelles tel que décrit dans ce projet. Ce modèle n'est pas concurrentiel avec des modèles biomécaniques très performants tels que ceux développés par l'équipe de F. Nedelec (EMBL Heidelberg) car, contrairement à ces derniers, il est focalisé sur la spatialisation de l'interrelation entre les microtubules et leurs constituants libres en solution.

Les premiers résultats n'ont pas tardé. Ils remettent en question la proposition selon laquelle les microtubules, dans les solutions *in vitro* qui s'auto-organisent spatialement (travaux de J. Tabony), s'auto-organiseraient par des processus collectifs de réaction-diffusion, c'est-à-dire comme des fours moléculaires. Les simulations prouvent en effet l'inefficacité des communications chimiques entre microtubules dans ces solutions à l'échelle microscopique. Néanmoins, elles ne remettent pas en question le fait que certains systèmes microtubulaires *in vitro* puissent se comporter comme des systèmes de réaction-diffusion (solutions temporellement ou spatialement oscillantes (Pirollet et al, 1987; Carlier et al, 1987; Mandelkow et al, 1989)) à un niveau plus macroscopique. Au contraire,

elles renforcent cette interprétation pour ces systèmes. Le critère pour savoir si de l'ordre peut apparaître par réaction-diffusion dans ces solutions est le même que celui décrit dans les systèmes classiques de réaction-diffusion de type Turing (A. Turing, Phil. Trans. Roy. Soc., 1952), à savoir le ratio Périodicité=Réaction/Diffusion. Dans les systèmes microtubulaires oscillants la dynamique des microtubules est telle que des structures spatiotemporelles peuvent apparaître. Ce n'est pas le cas dans les solutions auto-organisées de J. Tabony où les microtubules sont particulièrement stables.

Ce travail a été partiellement publié dans 2 articles dans des journaux de sciences computationnelles. De plus, il a été ensuite accepté en 2012 dans la revue *Acta Biotheoretica*.

7.1.5. Travaux en computation non-conventionnelle. Projet Biocolliders (2004-2007) [*]

Sur l'impulsion de E. Winfree (Caltech Institute, Californie) qui m'a contacté à propos de mes travaux sur l'auto-organisation des microtubules et avec qui j'ai pu avoir à plusieurs reprises des échanges très précieux ces dernières années, j'ai souhaité démarrer un projet de recherche original (projet présenté au CNRS en 2005): développer une approche computationnelle non conventionnelle fondée sur la dynamique collective d'agents fibrillaires microscopiques, tels que les microtubules ou les filaments d'actine. Ce projet, difficile à mettre en place et à formuler, a considérablement évolué depuis que je l'ai initié.

Partant du principe que le vivant œuvre depuis plus de 3.5 milliards d'années à développer les plus efficaces formes de traitement de l'information et de nano et micro machinerie qui soient, ce type d'approche connaît un développement très important outre manche et outre atlantique mais reste pratiquement inexistant en France. Ces approches permettraient de pallier aisément certaines limitations des systèmes de calcul classique dans certaines applications. En effet, elles s'appuient sur des systèmes capables de réaliser des tâches de façon massivement parallèle et analogique, et souvent capables d'auto-organisation. De plus, si de tels travaux paraissent risqués (il n'est pas évident *a priori* de pouvoir manipuler les microtubules et de les diriger par des biais chimiques ou physiques vers un calcul particulier), ils enrichissent en réalité considérablement les thématiques concernées (biologie, sciences computationnelles, biophysique) en nouveaux concepts, outils et résultats.

Fort de ma connaissance des systèmes collectifs auto-organisés, en particuliers des solutions de microtubules, et de mon intérêt pour les technologies bio-inspirées, j'ai commencé ce travail avec pour objectif de montrer que l'on peut utiliser les capacités des fibres dynamiques, comme les microtubules, l'actine, les nanotubes de carbone ou d'ADN, à s'auto-assembler et à s'auto-organiser pour réaliser des systèmes de calcul (processeurs chimiques ou biochimiques) ou pour concevoir des machines à l'échelle du nanomètre et du micromètre. En particulier je souhaitais fonder mes travaux sur l'hypothèse, défendue par J. Tabony et moi-même pendant ma thèse, que les microtubules puissent s'auto-organiser, comme dans un système multi-agents, par l'intermédiaire d'une communication chimique.

Afin de quantifier cette communication chimique, j'ai développé un modèle microscopique de

dynamique microtubulaire permettant de mesurer la façon dont les microtubules modifient leur environnement chimique et sont eux même affectés par ces variations. J'ai rapidement du revoir mes idées sur l'auto-organisation des microtubules au niveau microscopique et remettre ainsi en question la thèse défendue par J. Tabony. Néanmoins, ayant déterminé un des critères importants pour la programmabilité (structurale) potentielle de ces systèmes d'agents fibrillaires, j'ai pu conclure à la possibilité de développer cette approche en utilisant d'autres systèmes plus grands, comme les systèmes artificiels composés de billes de latex de 1 μm propulsées par des comètes d'actine (comme le font les bactéries *Listeria* et *Shigella*).

Au cours de ce travail, j'ai aussi eu l'occasion d'échanger beaucoup avec A. Adamatzky (Univ. of West England, Bristol, UK). Cela a abouti à deux publications, une dans un chapitre de livre et proposée à l'occasion d'un Workshop, l'autre dans un journal de computation non conventionnelle. En 2008, j'ai rencontré O. Bastien, jeune chercheur travaillant au CEA Grenoble et très intéressé par ces idées. Une collaboration s'est vite mise en place sur d'autres thématiques avec une très bonne complémentarité des compétences.

En 2007, j'ai fait une demande de financement pour l'ERC starting grants en tant qu'investigateur principal et unique contributeur. J'ai obtenu la note de 6.38, la barre étant à 8, le projet ayant été bien accepté par les 4 referees qui ont critiqué un calendrier trop serré étant donné les objectifs annoncés, trop ambitieux, et une méthodologie expérimentale insuffisamment détaillée. Les projets que j'ai menés après en ont tenu compte. En 2008, en collaboration avec O. Bastien (PCV, CEA Grenoble) j'ai fait une demande de financement pour un préprojet de l'institut Rhone-Alpin des systèmes complexes (IXXI); nous avons obtenu 5000 euros. Ce financement nous a permis de présenter nos travaux dans 3 conférences en 2009 (UC'09 et CIE 2009 principalement, et ECCS'09). **A l'avenir, je souhaite revenir sur cette thématique.** Une option viable serait de monter une collaboration sur ce projet plus directement avec A. Adamatzky en faisant une demande de financement EPSRC, son équipe étant régulièrement financée par cette voie. Ce dernier développe actuellement des projets dans ce sens (voir *Computing on actin bundles network*, <https://www.researchgate.net/publication/331987049> *Computing on actin bundles network*).

7.1.6. Modélisation de l'organisation du cytosquelette en présence de moteurs moléculaires dans le contexte de l'activation plaquettaire – développement logiciel. Projet Platon 2016.

Le cytosquelette de microtubules s'auto-organise en présence de moteurs moléculaires (kinésine et dynéine) dans le contexte de l'activation plaquettaire [1, 2, 4-6]. Le disque plat formé par les plaquettes sanguines quiescentes est maintenu par les microtubules qui sont organisés selon un anneau périphérique appelé bande marginale (MB) [1]. La transition de la forme disque vers une sphère durant l'activation plaquettaire est induite par une extension de cette bande marginale due au couple microtubules et moteurs moléculaires. Cela provoque un repliement (coiling) de la bande marginale et ainsi un changement de la forme de la plaquette qui devient sphérique [2] ; voir figure 7.1.

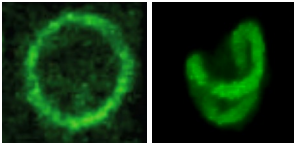


Figure 7.1. Microtubules marqués dans une plaquette sanguine. Bande marginale plate d'une plaquette quiescente (gauche). Bande marginale en train de se replier durant l'activation de la plaquette (droite).

Le projet porté par l'IAB, Grenoble, visait à observer et mesurer ce processus (WP1), en particulier par Imagerie Digitale Holographique Quantitative (DHM [3]) de la bande marginale (ce travail sera principalement assuré par Y. Usson et A. Fertin au TIMC-IMAG). Il s'agissait d'observer la formation de la bande marginale durant l'activation plaquettaire et de réaliser des mesures spatiales et temporelles de l'interaction entre les différents constituants du cytosquelette impliqués (microtubules, actine et moteurs moléculaires).

Un autre aspect de ce projet (WP3) concernait la modélisation de l'auto-organisation du cytosquelette dans les plaquettes. Il s'agissait de réaliser un modèle biomécanique et réactionnel des microtubules en interaction avec un réseau d'actine et des moteurs moléculaires. Nous souhaitons montrer comment se forme la transition de la forme des plaquettes d'un disque à une sphère et identifier les facteurs essentiels de la transition.

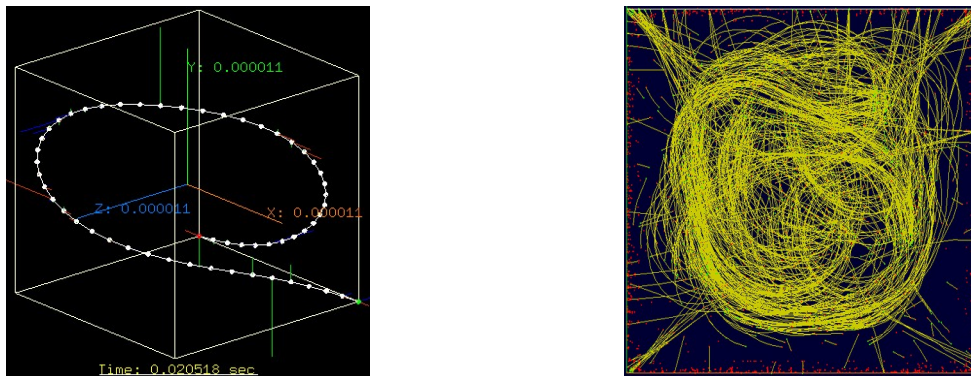


Figure 7.2. Modèle mécanique de microtubules simulés dans un micro-environnement cubique de 20x20x20 μm . Les microtubules simulés s'assemblent ou se désassemblent. Les extrémités croissantes (vertes) et décroissantes (rouges) ont un comportement d'instabilité dynamique. Tandis qu'ils croissent, les microtubules sont mécaniquement contraints par les limites de l'environnement ; cela provoque leur courbure. (Gauche) un microtubule unique en train de se courber. Sa discrétisation en petits sous-éléments est affichée. (Droite) des milliers de microtubules en train de nucléer, croître, se courber, forment ensemble une sorte de faisceau de microtubules circulaire comme observé *in vitro*.

Les perspectives d'un tel travail sont thérapeutiques. En plus d'agir sur l'hémostase (formation du thrombus), les plaquettes interviennent dans le processus inflammatoire, dans la réparation des tissus et dans la progression tumorale. Une connaissance détaillée du rôle et de la fonction du cytosquelette plaquettaire donnerait des pistes pour des stratégies thérapeutiques pour prévenir les pathologies associées. En outre, les mécanismes fondamentaux (et outils développés ici) de la réorganisation du cytosquelette peuvent être transposés dans les cellules nucléées [7] et aider à comprendre certains aspects moléculaires de la migration cellulaire.

Nous avons développé une modélisation du cytosquelette ([8] et figure 7.2), en utilisant un modèle numérique de fibres réactives et ayant des propriétés mécaniques (fondé sur un modèle multi-agent de fibres dont la mécanique est fondée sur des éléments masse-ressorts) et de moteurs moléculaires (kinésine et dynéine) et d'un container déformable qui mime le cortex d'actine. Ce modèle devait permettre d'étudier quel est le rôle des moteurs moléculaires, de l'acétylation des microtubules (qui affecte leur dynamique) et de l'interaction microtubules-cortex contractile d'actine dans ce processus. Nous devons simuler la croissance et le repliement de la bande marginale auto-organisée à partir d'une collection de fibres et de moteurs dans un environnement contraint figé ou déformable. Dans ce projet, nous devons aussi évaluer la dépendance relative du processus vis-à-vis du cortex d'acto-myosine (contractile) et mesurer les forces mécaniques mises en œuvre. Ces mesures de force devaient être comparées à des mesures de déformations et de forces réalisées par microscopie dans le WP2 du projet. Nous devons ensuite déterminer comment un anneau secondaire, plus petit, de nouveau microtubules très dynamiques peut se former à partir d'un support (template) replié formé par la bande marginale.

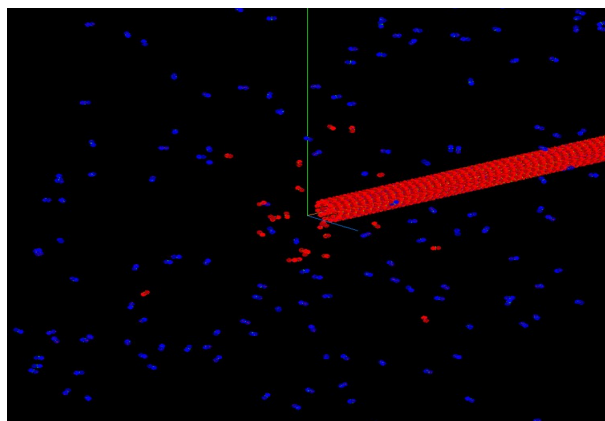
La partie expérimentale de ce travail a été réalisée avec succès. Cependant, le groupe de modélisation de F. Nedelec (Dmitrief et al, PNAS 2016) a produit un modèle bien avant nous, dès le début de notre projet. Nous avons donc mis l'accent sur le développement d'une API logicielle (MESS) destinée à faciliter la modélisation multi-échelle en biologie. Je continue de développer activement cette API.

- 1 White JG, Rao GH. Microtubule coils versus the surface membrane cytoskeleton in maintenance and restoration of platelet discoid shape. *Am J Pathol.* 1998; **152**: 597-609.
- 2 Diagouraga B, Grichine A, Fertin A, Wang J, Khochbin S, Sadoul K. Motor-driven marginal band coiling promotes cell shape change during platelet activation. *J Cell Biol.* 2014; **204**: 177-85.
- 3 Depeursinge C, Colomb T, Emery Y, Kuhn J, Charriere F, Rappaz B, Marquet P. Digital holographic microscopy applied to life sciences. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference.* 2007; **2007**: 6244-7.
- 4 Patel-Hett S, Richardson JL, Schulze H, Drabek K, Isaac NA, Hoffmeister K, Shivdasani RA, Bulinski JC, Galjart N, Hartwig JH, Italiano JE, Jr. Visualization of microtubule growth in living platelets reveals a dynamic marginal band with multiple microtubules. *Blood.* 2008; **111**: 4605-16.
- 5 Sadoul K, Wang J, Diagouraga B, Vitte AL, Buchou T, Rossini T, Polack B, Xi X, Matthias P, Khochbin S. HDAC6 controls the kinetics of platelet activation. *Blood.* 2012; **120**: 4215-8.
- 6 Kalebic N, Sorrentino S, Perlas E, Bolasco G, Martinez C, Heppenstall PA. alphaTAT1 is the major alpha-tubulin acetyltransferase in mice. *Nat Commun.* 2013; **4**: 1962.
- 7 Chao FC, Shepro D, Tullis JL, Belamarich FA, Curby WA. Similarities between platelet contraction and cellular motility during mitosis: role of platelet microtubules in clot retraction. *J Cell Sci.* 1976; **20**: 569-88.
- 8 Glade N. On the nature and shape of tubulin trails: implications on microtubule self-organization. *Acta biotheoretica.* 2012; **60**: 55-82.

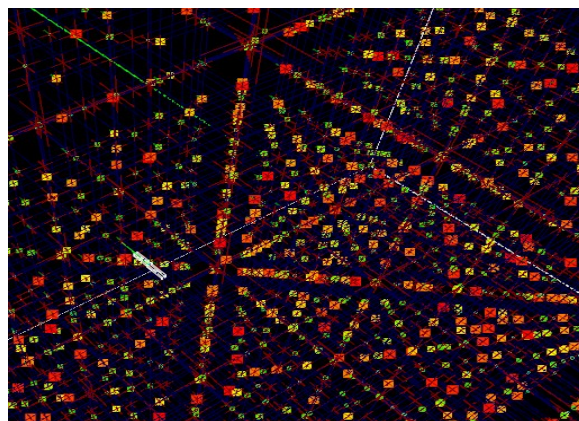
7.1.7. Etude de la diffusion anisotrope des petites molécules au voisinage de régions organisées de la cellule, en particulier au voisinage des faisceaux de microtubules. [*]

Dans ce projet sur lequel je reviens régulièrement, je souhaite mesurer la diffusion anisotrope des petites molécules au voisinage de régions organisées de la cellule, en particulier au voisinage des faisceaux de microtubules (regroupement de plusieurs milliers de microtubules alignés et espacés de quelques dizaines de nm).

Il s'agit d'utiliser le modèle de microtubule, décrit à une échelle microscopique en solvant implicite (toutes les molécules de tubuline sont représentées, voir figures 7.3 et 7.4), pour quantifier la diffusion des petites molécules au voisinage des fibres du cytosquelette lorsque celle-ci sont désorganisées ou, au contraire, lorsqu'elles sont organisées. Il s'agit donc de montrer le caractère anisotrope de la diffusion (dont nous pensons qu'il est bien réel d'après ce que suggèrent des simulations déjà réalisées [8]) et de le mesurer.



Un microtubule et les molécules de tubuline (représentation individuelle de molécules) (microtubule : 0.5 μm)

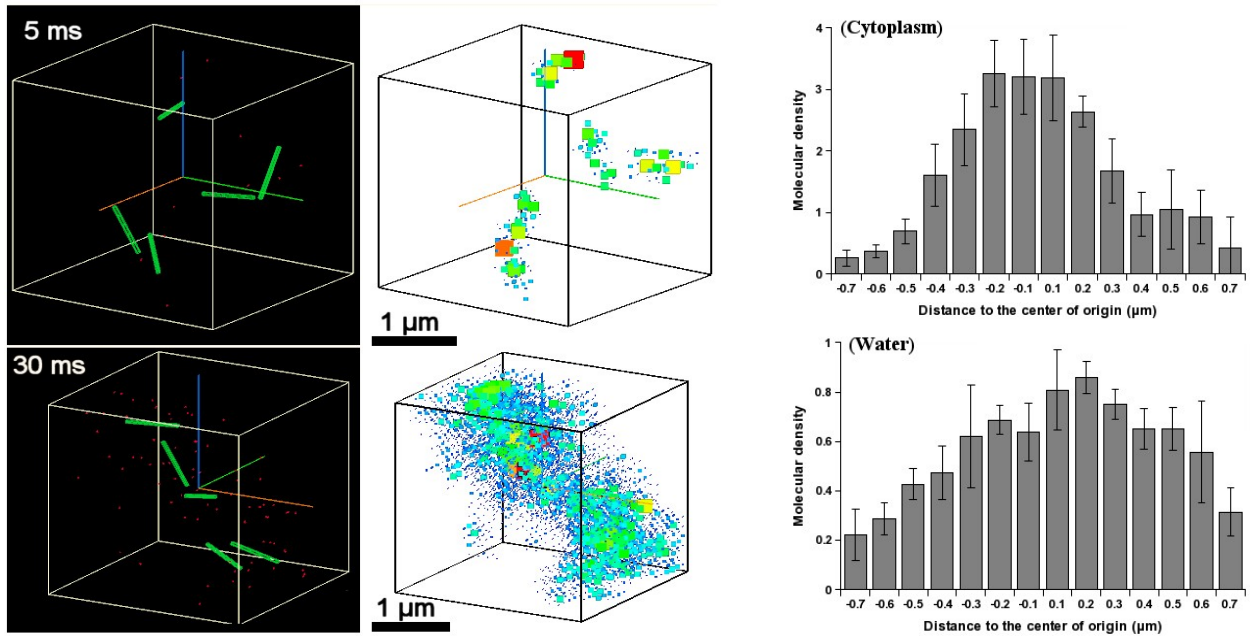


L'environnement chimique (nombre entier de molécules), maillage cubique, avec sa matrice de connectivité

Figure 7.3. Modèle microscopique (multi-échelle) du comportement de solutions de microtubules. Le modèle décrit très finement les processus de diffusion et prend notamment en compte le facteur de forme des objets présents (sous la forme d'une approximation par des ellipsoïdes) pour permettre une description de l'anisotropie de leur diffusion.

En plus de la modélisation, je souhaiterais pouvoir réaliser des expériences *in vitro* pour mesurer ce caractère anisotrope de la diffusion. Ces expériences, dont j'ai une grande expertise (thèse et premier post-doc) consisteront à orienter à différents degrés des solutions de microtubules en utilisant des champs magnétiques intenses lors de leur croissance (en collaboration avec E. Beaugnon, LNCMI, CNRS Grenoble) et de mesurer la diffusion de petites molécules traçantes (ex: rhodamine) cagées (donc activables par une activation lumineuse intense *a posteriori*, une fois la solution assemblée) par microscopie de fluorescence. La microscopie pourra être réalisée au TIMC-IMAG qui possède une plate-forme performante en microscopie de fluorescence notamment.

Le caractère anisotrope de la diffusion des petites molécules au voisinage du cytosquelette est important à prendre en compte car de cela peuvent découler de nouvelles compréhensions du transport passif des petites molécules dans les cellules, ainsi que des pistes thérapeutiques pour la délivrance de molécules à usage thérapeutique agissant sur le cytosquelette ou véhiculée par celui-ci. Par exemple, le site actif du Taxol, un anti-cancéreux se trouve à l'intérieur des microtubules (dans le tube). Ceci implique que la qualité de son action sera déterminée par l'accessibilité des extrémités microtubulaires et par le degré d'organisation des microtubules.



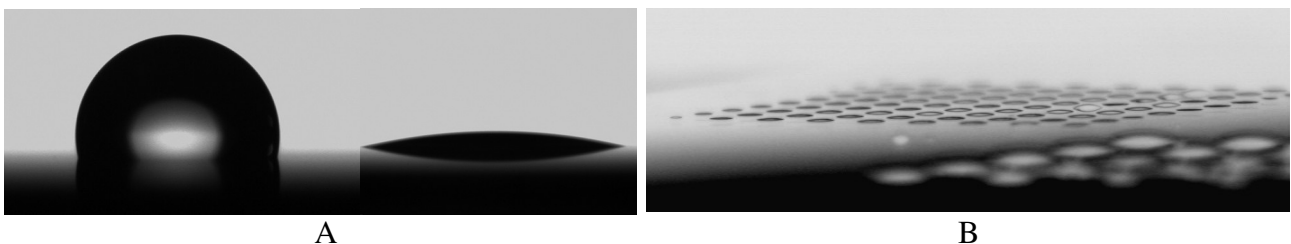
Réaction et diffusion de 5 microtubules libres et mesure de la quantité de tubuline-GDP générée.

Profil de diffusion moyen en tubuline-GDP à l'extrémité des microtubules.

Figure 7.4. Quantification des hétérogénéités de composition dans les solutions de microtubules et de la diffusion anisotrope.

7.2. Second post-doctorat (2003-2004) : post-doc RTB (recherche technologique de base) financé par le LETI (Laboratoire d'Électronique et Techniques de l'Information). Technologies plastiques.

J'ai été en contrat durant 1 an au laboratoire d'optoélectronique dans la section biocomposants plastiques du laboratoire LETI, Laboratoire d'Électronique et Technologies de l'Information sous la responsabilité d'Isabelle Chartier. Mon financement était un financement LETI de type RTB, recherche technologique de base. Durant cette année, j'ai effectué une mobilité thématique complète en micro et nanotechnologies, en particulier physique des surface et traitements de fonctionnalisation.



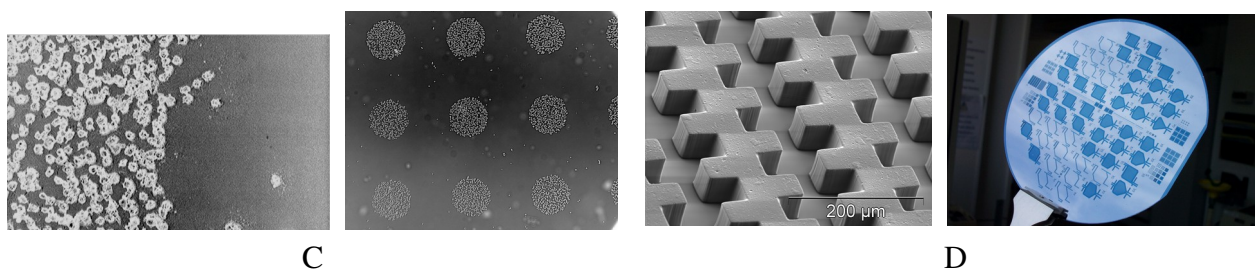


Figure 7.5. Quelques productions faites durant mon postdoc au LETI. (A) vues de coté de gouttes d'eau pure de 15 μl posées sur un film d'OMCTSO déposé par PECVD, à gauche sur la surface brute, à droite après traitement de 10 minutes sous UV+ozone. (B) Vue de coté d'une 'matrice' de gouttes d'eau retenues sur les zones (100 disques de 125 μm de diamètre) après traitement localisé par UV+ozone 10' d'un film d'OMCTSO déposé par PECVD sur un wafer de silicium. (C) Adhésion localisée de cellules HeLa sur des surfaces de polycarbonate (à gauche) et d'OMCTSO (à droite) ayant subi un traitement localisé d'UV+ozone 10'. Les cellules n'adhèrent que sur les parties traitées. A droite, les disques traités font 1 mm de diamètre. (D) photolithographies utilisant le film sec *Ordyl*. A gauche, un gros plan au MEB sur des canaux de microfluidique épais de 75 μm ; à droite un wafer de silicium sur lequel ont été construits avec cette technique de film sec des dispositifs microfluidiques destinés à faire du tri cellulaire.

J'ai travaillé sur 3 problèmes différents (voir figure 7.5). Premièrement, j'ai étudié la primo-adhérence et de la biocompatibilité des cellules sur divers substrats, plastiques notamment, utilisés dans l'élaboration de microsystèmes biologiques. J'ai élaboré et traité moi-même ces substrats et j'ai caractérisé leurs propriétés physico-chimiques. Deuxièmement, dans le même cadre, je suis inventeur d'une nouvelle brique technologique permettant de localiser par traitement de surface global ou localisé n'importe quelle substance liquide ou filmogène sur une surface lisse ou structurée. Cette nouvelle technique permet de réaliser avec précision un grand nombre d'applications très variées, en particulier dans le domaine des réseaux de microélectrodes, des transistors organiques, des piles à combustible, de l'optique et enfin, de la biologie ou de la biochimie. La faisabilité de certaines de ces applications en utilisant cette technique a été testée avec succès en collaboration avec différents laboratoires du CEA Grenoble. Un brevet international a été déposé sur cette technique de localisation d'espèces chimiques ou biologiques. Enfin, en collaboration avec l'Université de Bologne, j'ai mis au point une technique de photolithographie en film sec adaptée pour des applications en micro-fluidique (double bounding) à bas coûts de développement. Sur ces trois sujets, deux articles ont été publiés (dont revue Lab on a Chip) et un brevet enregistré.

Étant donné mon profil de recherche généraliste et mon intérêt pour les traitements localisés de contrôle de l'énergie de surface, j'ai aussi été amené à beaucoup interagir avec les diverses équipes du laboratoire sur des projets très variables (projets biotechnologiques divers, projets DVD, projets de transistors organiques, modélisation d'un processus de lithographie en plan incliné ...).

7.3. Troisième post-doctorat (2005-2006) : Financement ALFA (America Latina, Formación Académica) au Centre de Modélisation Mathématique, Université du Chili, Santiago, Chili : Modélisation du couplage des réseaux métaboliques et génétiques du contrôle de l'homéostasie du fer et du glutathion

J'ai été en contrat post-doctoral 1 an au Centre de Modélisation Mathématique, CNRS UMI2807, rattaché à l'Université du Chili à Santiago, sous la responsabilité du Pr. Rafael Correa. J'étais financé par un programme de la commission européenne, ALFA, de collaboration entre l'Amérique du sud et l'Europe. Là bas, j'ai initié mon projet de recherche concernant l'auto-organisation des fibres dynamiques (microtubules) et les systèmes de calcul non conventionnels fondés sur des complexes moléculaires fibrillaires.

$$X_g(t) = H\left[\left(R(t) - \sigma_X^Y([Y_r], \lambda_X^Y, n_X^Y, 1)\right)\right]$$

$$Z_g(t) = H\left[\left(R(t) - \sigma_Z^Z([Z_r], \lambda_Z^Z, n_Z^Z, 1)\right)\right]$$

$$X_f : \frac{d[X_f]}{dt} = X_g k_X^{(s)} + k_X^{(D-)} [X_f]$$

$$Z_r : \frac{d[Z_r]}{dt} = Z_g k_Z^{(s)} - k_{ZY}^{(cat+)} [X_f] [Z_r] + k_{ZY}^{(cat-)} [Y_r] + k_Z^{(D-)} [Z_r]$$

$$Y_r : \frac{d[Y_r]}{dt} = k_{ZY}^{(cat+)} [X_f] [Z_r] - k_{ZY}^{(cat-)} [Y_r] + k_Y^{(D-)} [Y_r]$$

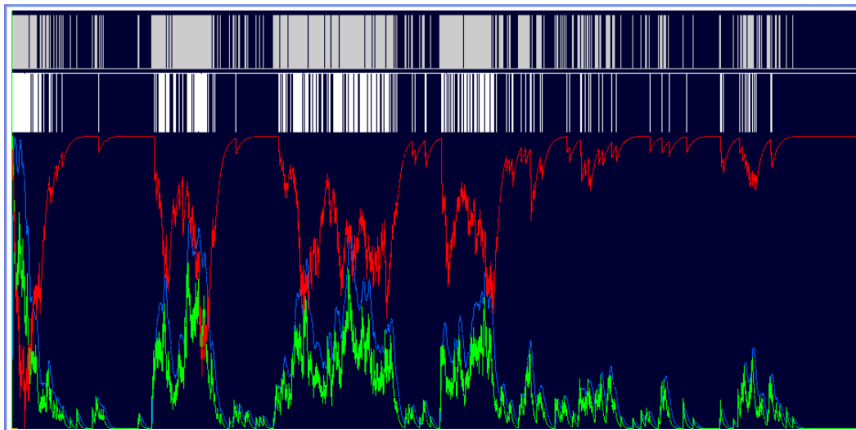
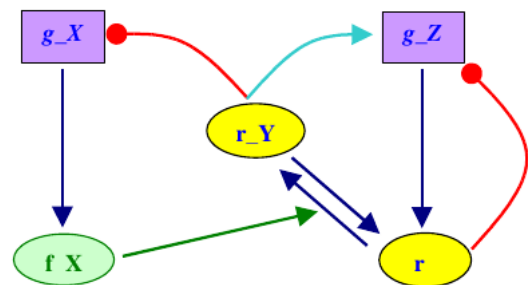
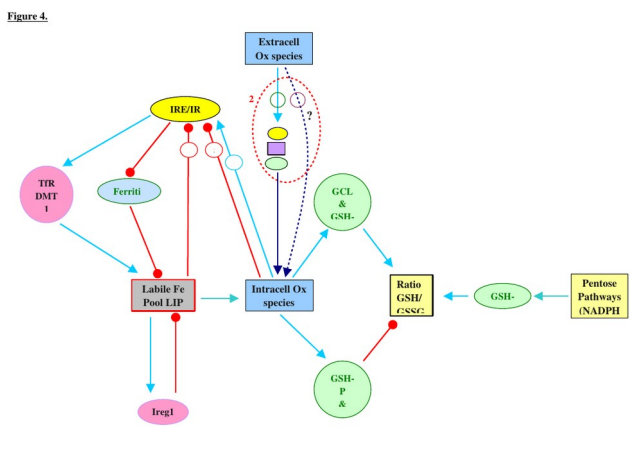
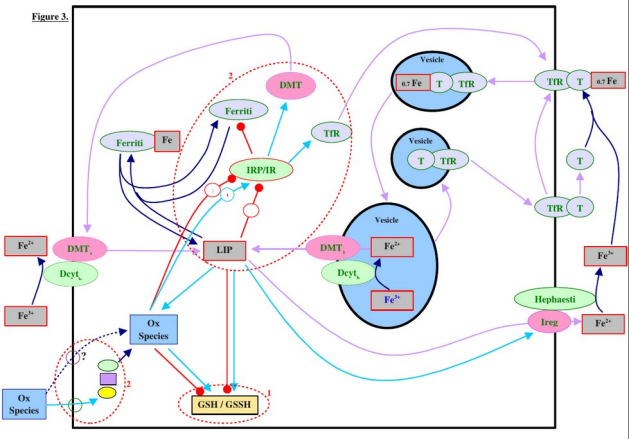
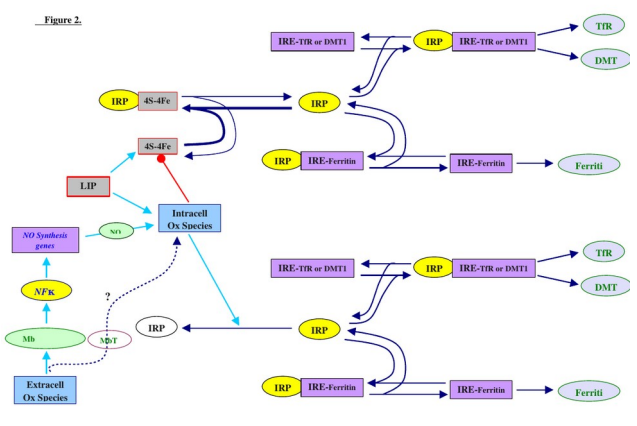
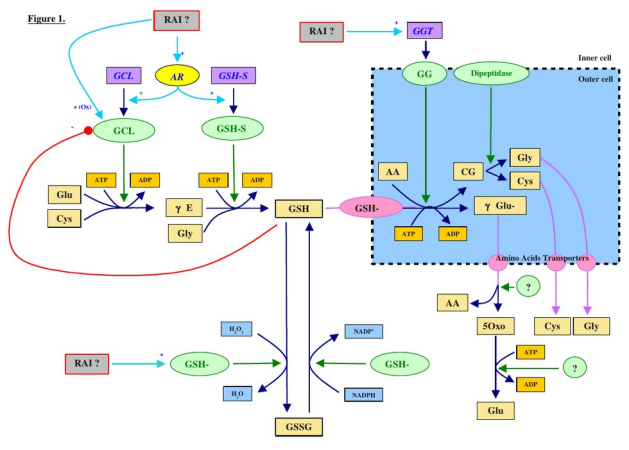


Figure 7.6. Un exemple de système de régulation génétique et métabolique comportant 2 gènes (g_X et g_Z), deux composés (r_Y et r_Z) et une enzyme (f_X), simulé avec un modèle hybride. Les gènes sont modélisés par une fonction heaviside dépendant d'un nombre aléatoire et d'une fonction de saturation de Hill dépendant des composés en présence. A chaque instant, le gène est actif ou inactif (hachures blanches sur la simulation : chaque gène est visible sur une ligne). Les concentrations des composés r_Y , r_Z et f_X sont modélisées par des équations différentielles ordinaires. Les gènes permettent ou pas la synthèse de ces composés. Cette modélisation hybride s'est avéré complexe à mettre en œuvre pour des systèmes plus compliqués.

J'ai aussi collaboré (collaboration à mon initiative) à un travail sur les phénomènes du vieillissement cellulaire, avec les laboratoires CIBYB (Centro de excellencia academica en Ingenieria Bioquimica Y Biotecnologia) du Pr. Asenjo et avec M. T. Nuñez au Department of Biology, Faculty of Sciences and Cell Dynamics and Biotechnology Research Center, de l'Université du Chili à Santiago. L'étude porte sur la réalisation d'un modèle du réseau métabolique et génétique dirigeant l'homéostasie du Fer, couplé à celui du cycle du Glutathion. L'objectif de ce travail consistait à extraire de la littérature un schéma de ces 2 systèmes couplés, puis de démarrer la réalisation un modèle afin de

confirmer qualitativement et quantitativement leur couplage. De plus, il s'agissait à terme d'expliquer pourquoi certaines cellules développent subitement une adaptation aux quantités de fer, survivent et se développent alors que la majorité meurent dans ces conditions. J'ai donc réalisé ce travail d'extraction et obtenu un schéma assez important des réseaux métaboliques et génétiques du fer et du glutathion (figure 7.7). J'ai ensuite introduit un formalisme de modélisation stochastique (d'abord une méthode mixte entre système différentiel pour le réseau métabolique et stochastique pour les gènes, puis une méthode stochastique directement inspirée de la méthode de Gillespie) dans la modélisation des réseaux métaboliques et génétiques permettant la simulation de comportements de cellules individuelles (figure 7.6). Cela permet d'introduire la variabilité observée dans les populations cellulaires (tous les gènes de toutes les cellules n'étant pas dans le même état en permanence). Ce travail s'est arrêté là faute d'avoir à disposition les données nécessaires à injecter dans le modèle (lois de comportement cinétiques, paramètres cinétiques). C'est une difficulté inhérente à ce type de modèles. Néanmoins, le modèle existe (sous forme d'équations) et un document très détaillé a été produit à l'attention du CIBYB de façon à ce que laboratoire puisse poursuivre cette tâche.



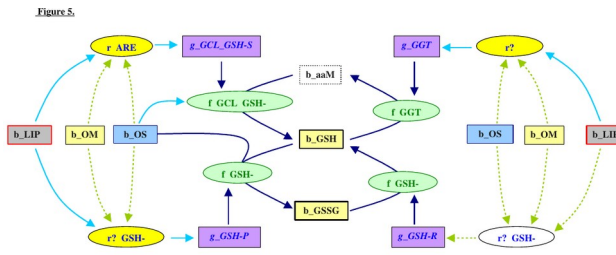


Figure 7.7. Régulation métabolique du fer. L'ensemble de ces diagramme représente un réseau de régulation que j'ai construit à partir de la bibliographie. Sa modélisation avec un modèle hybride s'est révélée trop complexe à mettre en œuvre (à paramétrer).

7.4. Fibration isochrone et décomposition potentiel-hamiltonienne (travaux étalés de 2002 à 2006) [*]

Depuis 2002, nous avons travaillé, J. Demongeot, L. Forest et moi-même sur la décomposition potentielle-Hamiltonienne des systèmes de Liénard et sur leur justification biologique (nous expliquons en quoi la décomposition de modèles biologiques sous la forme potentielle-Hamiltonienne augmente leur compréhension). La décomposition (de Hodge) se fait, de manière non unique, en deux parties polynomiales, représentant les dynamiques de gradient (potentielle) et Hamiltonienne des systèmes étudiés. Nous en donnons la formule explicite au voisinage d'orbites particulières comme les cycles limites de systèmes de Liénard. Nous nous sommes intéressés à plusieurs systèmes biologiques types, pouvant être ramenés à des systèmes de Liénard, (oscillateurs couplés avec phénomènes d'entraînement, type Fitzhugh-Nagumo ou van der Pol) (figure 7.8) et à des systèmes hamiltoniens purs (ex: modèle de Lotka-Volterra) ou potentiels purs retrouvés en biologie pour le contrôle de gènes de morphogénèse dans les végétaux (les systèmes n-switches). J'ai particulièrement étudié ces systèmes n-switch et j'ai cherché à donner du sens biologique à ces travaux. J'ai également réalisé un programme pour calculer finement les trajectoires isochrones de nos systèmes. Ces travaux ont abouti, en 2006, à la publication de 4 articles dans les Comptes Rendus de l'académie des sciences, 3 dans les C. R. Mathématiques et 1 dans les C. R. Biologies.

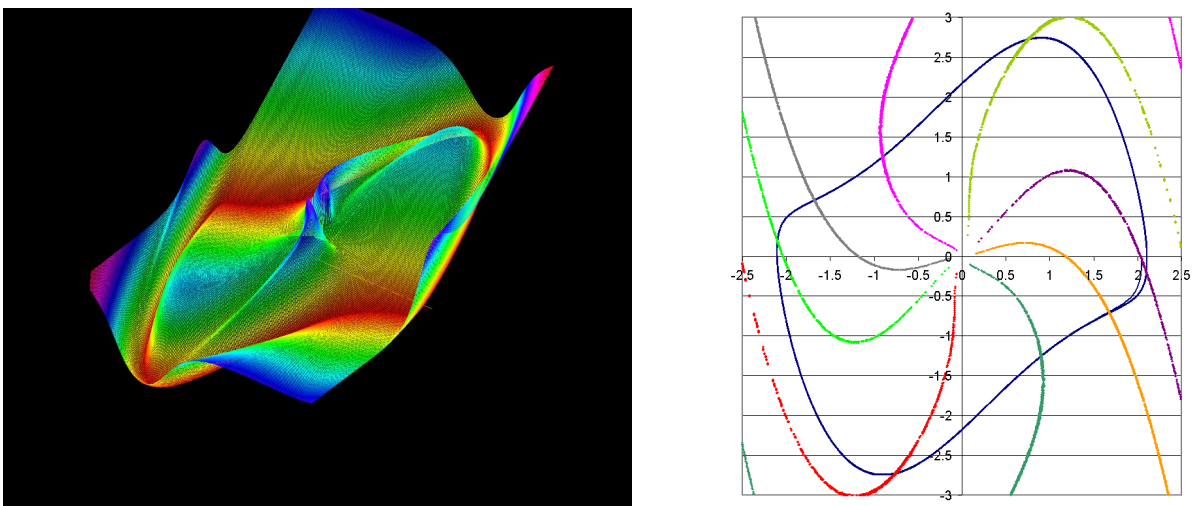


Figure 7.8. Système de van der Pol. A gauche l'intensité du vecteur potentiel-hamiltonien de ce système différentiel. A droite, 8 de ses isochrons.

Avec l'étudiant en thèse (H. Ben Amor) que nous avons co-encadré Jacques Demongeot et moi-même, nous avons repris ce travail, en particulier tout ce qui concerne la fibration isochrone et la synchronie dans les populations d'oscillateurs. Nous avons mis au point des algorithmes efficaces pour calculer la fibration isochrone en 2D, mais aussi dans des systèmes 3D (nous sommes les premiers à l'avoir fait). Ce travail nous a permis d'affiner le comportement des isochrons au voisinage des cycles (figure 7.9). La compréhension de la structuration 2D et 3D des isochrons dans l'espace des phases et leur approximation éventuelle nous a paru utile pour l'étude de l'évocation mnésique notamment pour avoir des accès directs (comme des sortes de pointeurs d'adresse mémoire) vers des régions spécifiques de la trajectoire asymptotique. Nous avons aussi utilisé la cartographie de la fibration isochrone pour connaître les régions dans lesquelles une perturbation peut mener à la synchronie, ou au contraire à une désynchronisation d'une population d'oscillateurs. Pour en savoir plus, je renvoie à la lecture de l'excellente thèse d'Hedi Ben Amor.

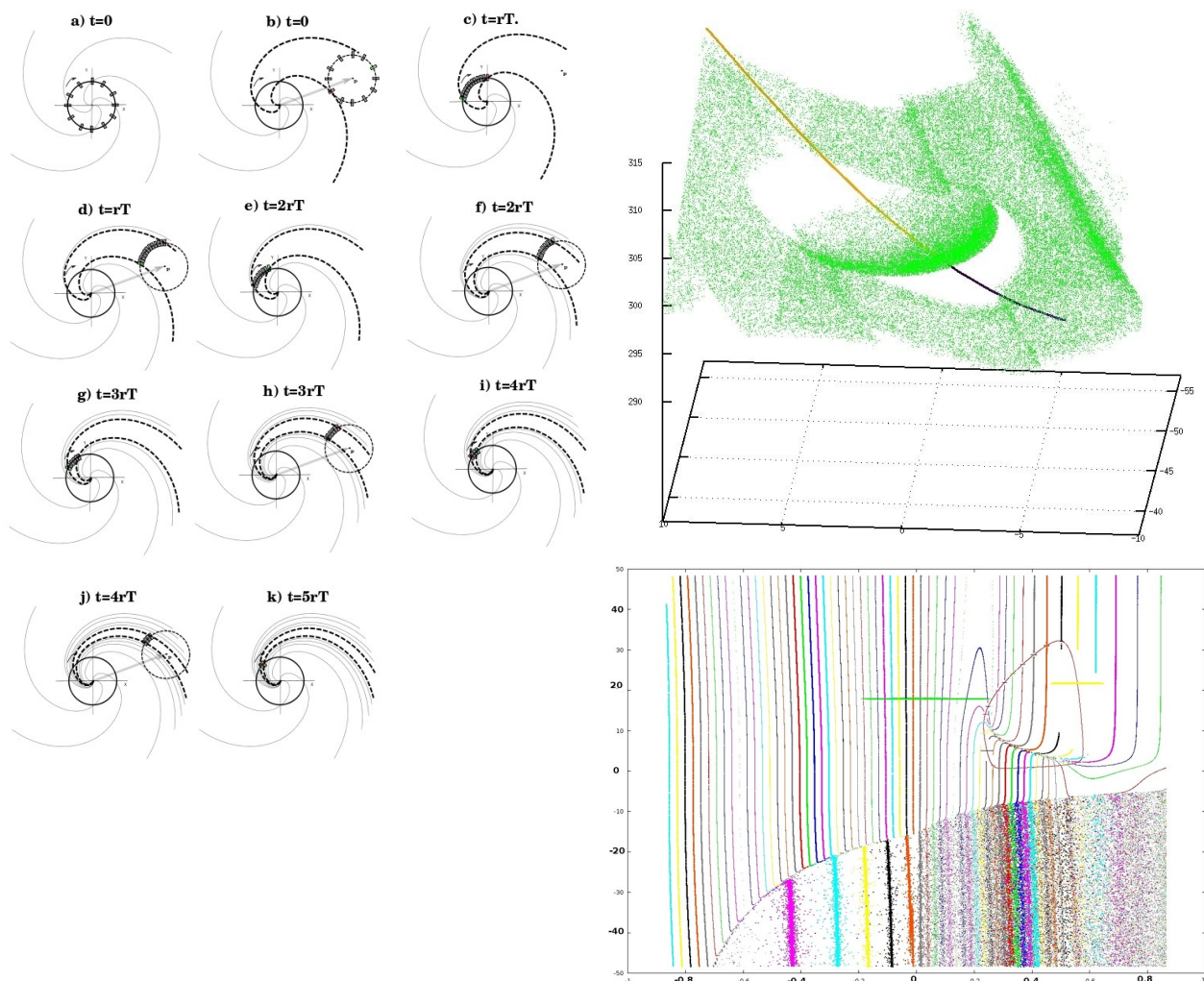


Figure 7.9. Fibration isochrone. A gauche, usage de la cartographie de la fibration isochrone pour synchroniser des populations d'oscillateurs (rectangles gris). A droite en haut, surface isochrone (en vert) au voisinage d'un cycle tridimensionnel. L'intersection est orthogonale. En bas à droite, cycle de l'enzyme Pyruvate-Fructo-Kinase (PFK) et la

fibration isochrone associée. Vers le haut, les isochrons sont clairement bien définis ; ce n'est pas le cas en bas, région dans laquelle la désynchronisation d'une population d'oscillateurs serait maximale.

Je continue régulièrement de réfléchir à ce travail avec l'esprit tourné vers les systèmes physiques, l'espace-temps, ses singularités et la relativité générale. Il y a en effet des similarités entre les attracteurs et les singularités de la physique. Je trouve intéressant d'apporter dans ces objets mathématiques (les systèmes dynamiques) un sens physique, une dépendance énergétique notamment.

Les attracteurs, qu'ils soient point fixes ou cycles limites, sont inatteignables par un système dynamique : le système se relaxe indéfiniment en se rapprochant, dans l'espace des phases, de l'attracteur. Si l'on rajoute une dimension supplémentaire au système, une dimension qui fait plonger la trajectoire dans un puits creusé par l'attracteur, dans ce cas on peut concevoir que chaque avancée du système dynamique vers l'attracteur vu dans la tranche que constitue l'espace des phases, sera compensée par une plongée d'autant plus importante dans la dimension supplémentaire. Ainsi, on comprend qu'accepter qu'une trajectoire soit au voisinage de l'attracteur par une définition à la limite n'a de sens que dans l'espace des phases mais pas dans cet espace d'augmenté d'une dimension. On voit donc aussi que définir un point de phase de l'attracteur (d'un cycle limite) comme une petite fenêtre englobant une petite partie de l'attracteur n'a pas plus de sens dans cet espace augmenté : la fenêtre est dilatée à l'infini dans la dimension supplémentaire. Cette dimension supplémentaire (une dimension potentielle vraisemblablement) me fait penser au changement de fréquence (vers le rouge, c'est-à-dire les basses énergies) que perçoit l'observateur extérieur qui observe un objet se rapprochant effectivement peu à peu d'un trou noir. Remarquons que du point de vue de l'observateur, l'objet n'atteint jamais le trou noir ; l'image qui lui parvient ne fait que changer apparemment de fréquence et l'objet semble s'arrêter alors qu'il atteint sa surface (s'il s'agit d'un trou noir sans rotation et que l'objet tombe en chute libre rectilinéaire dedans). Le point de vue de l'objet est très différent. Celui-ci traverse bien l'horizon des événements.

De plus, je pense qu'il faut voir la relation entre le flot tournoyant d'un système dynamique comme la déformation de l'espace, en particulier lorsque celui-ci s'approche du puits potentiel que constitue l'attracteur. Le système dynamique définit une transformation d'un espace plan de telle manière que cette déformation définit la trajectoire que suivrait une particule tombant vers un attracteur dans ce système. Cela ressemble à la déformation de l'espace-temps par des masses et aux trajectoires que suit la lumière dans de tels espaces. Les isochrons, eux, sont des marqueurs du potentiel d'un système ; associés à un point de phase d'un cycle limite, ils ignorent tout du mouvement de rotation qui fait tourner le champ de vecteurs de plus en plus parallèlement, dans l'espace de phase, à l'attracteur. Ils montrent ce que serait la trajectoire d'un point s'il ne subissait que l'attraction potentielle. Voir ces systèmes comme dotés d'une dimension de plus vers laquelle plongent à l'infini ses trajectoires, et les réfléchir en faisant des analogies physiques, sinon mathématiques, avec le monde de la relativité générale devrait nous suggérer de nouveaux outils pour manipuler les oscillateurs, biologiques notamment, comme nous avons souhaité le faire avec H. Ben Amor.

7.5. Modélisation de l'auto-organisation de la membrane interne mitochondriale (2006)

J. Demongeot, A. Moreira et moi-même avons proposé un modèle simple pour explorer certains aspects de l'organisation particulière de la membrane mitochondriale interne (MMI). J'ai développé un modèle simple bidimensionnel de biomécanique membranaire. L'auto-organisation en crêtes membranaires était suivie par un score d'efficacité énergétique fondé sur la proximité des complexes ATP synthase (qui catalyse la formation d'ATP à partir d'ADP en utilisant l'énergie d'un gradient transmembranaire de protons H⁺) et Adénylate Translocase (ANT, le transporteur de nucléotides adénylates permettant l'exportation de l'ATP produit vers l'extérieur de la mitochondrie, et l'importation d'ADP). Il semble en effet qu'il y ait un lien entre cette organisation particulière et la proximité de ces sites. Un tel repliement de la MMI et un rapprochement spatial des sites des réactions combinées permet un accroissement considérable de l'efficacité métabolique et respiratoire de la mitochondrie. Certaines pathologies liées à une conformation anormale des ATP synthases montrent notamment des structures sans crêtes, mais organisées en pelures d'oignons, confirmant leur rôle dans cette organisation. Nous proposons dans l'article publié dans *Biochimie* un ensemble de pistes pour la modélisation de l'organisation des mitochondries.

7.6. Rôle des microARNs en tant que clefs potentielles de la coévolution dans les écosystèmes viraux (2007-2008)

J'ai initié le sujet suivant dans le contexte d'une réflexion du laboratoire sur le rôle des microARNs dans la régulation génétique. Pour ce travail j'ai passé plusieurs mois à faire une importante bibliographie et j'ai rédigé une demande de financement à l'Institut Rhone-Alpin des Systèmes Complexes (IXXI) (étant déjà porteur d'un autre préprojet IXXI, je ne me suis pas mis porteur du présent projet). Son objectif était d'initier une collaboration entre un laboratoire de virologie (Unit of Virus Host Cell Interaction UMR 5233 UJF/CNRS/EMBL (UVHCI)), le TIMC-IMAG, et le département d'informatique de l'UTFSM à Valparaíso au Chili.

Ce sujet allie biologie des systèmes, bioinformatique et biologie. C'est ma connaissance de ces disciplines, mon intérêt pour ce problème et cette idée originale d'adopter une vue très mésoscopique du problème qui m'ont amené à prendre en charge ce projet. Je n'ai pas souhaité continuer l'aventure car je ne voulais pas me disperser davantage et parce que j'ai pu mettre le projet dans de bonnes mains, à savoir D. Bicot et P. Sabatier, physiciens théoriciens s'intéressant aux virus, dans l'unité EPSP-TIMC, UMR 55-25, Ecole Nationale Vétérinaire de Lyon. Je continue cependant de participer aux réflexions du groupe.

Voici la retranscription du projet IXXI (porté par E. Drouet de l'UVHCI) qui a été accepté :

Objectifs. Dans le présent projet, nous nous intéressons à un réseau biologique d'interactions complexes ayant lieu à plusieurs échelles dans les plans spatiaux (géographique), temporels

(évolution, dynamique d'infection) et fonctionnels (infection): le réseau en question est fait du tissu dense des interactions entre les *Virus* d'une part et les *Espèces Vivantes* d'autre part. Notre hypothèse de travail est que cela se fait par l'intermédiaire de clefs de reconnaissance spécifiques dont les microARNs (miRs) sont de bons candidats. Nous proposons de l'étudier comme un réseau de percolation tel que le décrit Kaufmann [6], réseau dont les noeuds seront les espèces hôtes et les agents viraux et les liens élastiques les connections médiées par les miRs. Un autre objectif, en parallèle, sera l'ébauche d'un réseau semblable – naïf des relations réelles qu'entretiennent virus et hôtes – à partir des génomes hôtes et viraux disponibles et des séquences miRs connues. Cela pourrait constituer un nouvel outil pour comprendre et mettre en évidence de tels 'écosystèmes viraux' liés aux miRs. Enfin, l'ambition de ce travail sera d'apporter une nouvelle vision des virus, considérés dans cette perspective moins comme agents pathogènes que comme agents affectant et contrôlant l'évolution de leurs espèces cibles [9], et connectant les dynamiques évolutives de plusieurs espèces entre elles, dans une stratégie globale 'gagnant-gagnant'.

Contexte. Les miRs sont des petites séquences polyribonucléotidiques d'environ 22 bases et sont les produits spécifiques transcrits de séquences génomiques; Ils ont vraisemblablement un rôle majeur dans les processus de régulation génétique en s'appariant avec des séquences cibles (ARN messagers) qui leur sont complémentaires (au moins sur les 8 premiers nucléotides). En plus de leur mécanisme particulier, ils devraient probablement nous obliger à revisiter les aspects de la régulation génétique les concernant puisque, loin de se restreindre au schéma des régulateurs 'classiques' touchant un nombre très limité de cibles (souvent une seule), il semble qu'ils interviennent davantage comme des 'micromanagers' de l'ensemble de l'expression génique [7] agissant depuis une autre couche de régulation, plus indépendante, nommée le 'milieu miARN'. Un seul miR peut en effet toucher des centaines de cibles ARNm.

Il a été découvert que toutes les espèces cellulaires (eucaryotes et procaryotes) et au moins plusieurs virus à ADN (par exemple *Herpes Virus*) codent pour de nombreux miRs, souvent très spécifiques de l'espèce. Il reste aussi vraisemblablement de nombreux miRs non identifiés, notamment chez les virus mais également chez les espèces cellulaires. De plus il a été montré que plusieurs miRs cellulaires, outre leur fonction physiologique, sont capables de s'apparier spécifiquement avec des cibles virales présentes chez les virus à ADN et à ARN. Ainsi, ils jouent un rôle majeur lors des processus d'infection virale, tant du côté cellulaire en garantissant une certaine forme d'immunité, que du côté des virus qui utilisent aussi des voies de silencing miARN pour contourner ces défenses et réduire les autres systèmes de défense immunitaire [8]. Ainsi, beaucoup de travaux suggèrent maintenant que les miARNs sont des clefs d'une coévolution entre les virus et leurs hôtes [7, 8, 9, 10]. La conservation et la mutabilité de ces séquences détermineraient la liaison unissant ces derniers, dans une dynamique lente à l'échelle des espèces et non à celle de l'individu, au sein d'écosystèmes viraux.

Notre connaissance du monde viral et surtout de sa connexion avec le reste du monde vivant reste très restreinte. En effet, bien que les principales familles de virus aient probablement été identifiées, notre connaissance porte surtout sur les virus qui infectent l'homme ou les mammifères ou ceux qui infectent les espèces vivantes avec lesquelles nous avons des interactions (oiseaux, arthropodes,

moustiques), mais cela ne reste qu'une infime fraction de la diversité des 'espèces' virales existant dans la nature [10]. Une façon de comprendre le monde viral dans sa globalité, à savoir (i) comment ils sont connectés à certaines espèces et pas à d'autres, (ii) comment ils changent d'espèces cibles, (iii) comment ils maintiennent ces connections spécifiques malgré l'évolution de leur hôtes, est de chercher à représenter ces inter-connections. De plus, cela permet de visualiser la manière dont l'ensemble du monde vivant étudié est relié, 'indépendamment' des considérations eco-systémiques classiques.

Modélisation. Outre la recherche de séquences miR et les aspects phylogénétiques, il n'y a encore aucun développement théorique en biologie systémique intégrant les miRs. Nous proposons de réaliser un 'toy model' fondé sur un réseau de percolation où les liens sont les connexions miRs dont la force correspond à la force d'appariement des miRs avec les cibles (on considérera un seuil d'appariement à partir duquel le lien sera considéré comme existant), et où toutes les espèces, hôtes ou virales (de plusieurs types), sont des noeuds ayant quelques caractéristiques, notamment la mutabilité des séquences qu'ils portent (dans le modèle, uniquement des séquences miR et des séquences cibles de miRs qui seront mémorisées sous la forme de listes pour chaque espèce) et la taille de son génome (un nombre donné pouvant évoluer). La mutabilité des séquences sera considérée comme inchangée pour une espèce donnée durant la simulation et tiendra compte de paramètres comme (a) le type d'agent, hôte ou viral (par ex: les virus à ARN font plus d'erreur à cause de la mauvaise fidélité de l'ARN polymérase ARN dépendante), (b) son taux de multiplication par cellule quand il s'agit des virus (en se multipliant beaucoup, selon leur type, ils peuvent produire davantage de quasi espèces), et (c) la taille de son génome. En outre, des duplications ou délétions de séquences miR ou cibles seront possibles. La simulation sera initiée au hasard avec un nombre fixe d'espèces portant des séquences miR et cibles (sauf pour les virus à ARN qui ne portent pas de miRs). Les liens seront ensuite établis en testant la connexion entre chaque virus et chaque espèce cible. Durant la simulation, les séquences miR et cibles pourront évoluer et leur répertoire par espèce changer : ceci conditionnera l'organisation du réseau. On étudiera ainsi la façon dont l'évolution de ces séquences (par mutations des séquences miRs et cibles virales surtout) en fonction des types viraux ('égoïstes' (virus à ARN) ou 'altruistes' (virus à ADN)) affecte la connectivité du réseau ou au contraire permet de maintenir des écosystèmes viraux. On pourra aussi étudier ceci lorsqu'on détruit des espèces hôtes et/ou virales (en éliminant les espèces trop peu connectées par exemple). L'organisation et le comportement du réseau seront aussi comparés à son équivalent construit à partir des données biologiques.

Objet de la demande. Nous souhaitons que ce pré-projet puisse initier ce travail théorique et développer un échange interdisciplinaire entre le TIMC-IMAG et l'UVHCI. Cela permettra aussi de dégrossir cette approche pour pouvoir ensuite l'étendre (PEPS / ERC) et faciliter dans le futur sa prise en main par des étudiants en thèse. Nous prévoyons que ce financement serve à couvrir notamment une ou deux participations au 'RNA 2008 – Thirteenth Annual Meeting of the RNA Society' à Berlin, où nous souhaitons présenter cette nouvelle approche, et éventuellement pour des frais de mission pour travailler avec notre collaborateur chilien Andres Moreira, spécialisé en modélisation bioinformatique.

1. Thelu MA, Drouet E, Hilleret MN, Zarski JP: Lack of clinical significance of variability in the internal ribosome entry site of hepatitis C virus. *J. Med. Vir.* 72 (2004) 396
2. Chassagne S, Laffly E, Drouet E, Herodin F, Lefranc MP, Thullier P: A high-affinity macaque antibody Fab with human-like framework regions obtained from a small phage display immune library. *Mol. Immunology* 41 (2004) 539
3. Mahot S, Fender P, Vives RR, Caron C, Perrissin M, Gruffat H, Sergeant A, Drouet E: Cellular uptake of the EBV transcription factor EB1/Zta. *Virus Res.* 110 (2005) 187
4. Laffly E, Danjou L, Condemine F, Vidal D, Drouet E, Lefranc MP, Bottex C, Thullier P: Selection of a macaque Fab with framework regions like those in humans, high affinity, and ability to neutralize the protective antigen (PA) of *Bacillus anthracis* by binding to the segment of PA between residues 686 and 694. *Antimicrobial Agent Chemotherapy* 49 (2005) 3414
5. Paquin N, Ménade M, Poirier G, Donato D, Drouet E, Chartrand P: Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. *Mol. Cell* 26 (2007) 795
7. Kaufmann S.: Autocatalytic sets of proteins, *J. Theor. Biol.* 119 (1986) 1
8. Bartel DP, Chen CZ: Micromanagers of gene expression: The potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* 5 (2004) 396
9. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P, Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA, *Science* 309 (2005) 1577
10. Witzany G, Natural genome-editing competences of viruses, *Acta Biotheor.* 54 (2006) 235
11. Koonin EV, Dolja VV, Evolution of complexity in the viral world: The dawn of a new vision, *Virus Research* 117 (2006) 1

7.7. Substitution visuelle : amélioration de la perception via un dispositif électrotactile, substitution de la vision en 3D, application à la vision prothétique. (2009-2015)

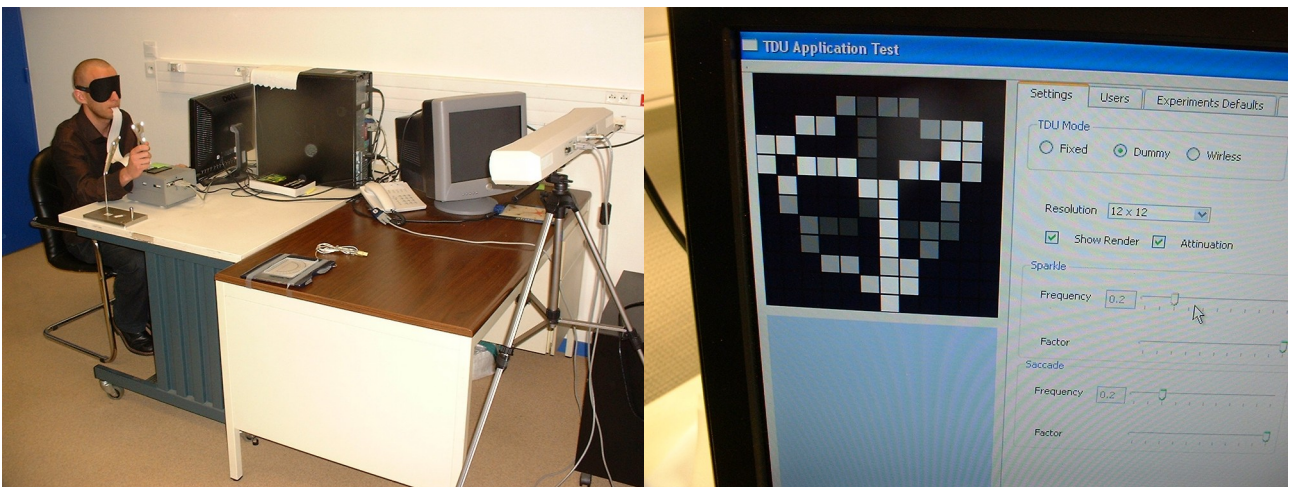
Groupe de travail. A. Chekhchoukh (TIMC-IMAG, CNRS UMR5525 & AGIM, FRE3405), G. Tatur (LGI2P, Ecole Mines d'Ales, Univ Nimes), I. Marc (LGI2P, Ecole Mines d'Ales, Univ Nimes), N. Vuillerme (AGIM FRE3405), C. Scheiber (CHU Bron), Y. Payan (TIMC-IMAG, CNRS UMR5525). Financement CSVSB – UJF 2011 – 25000€.

Dans le cadre de la mise au point et de la validation expérimentale de dispositifs médico-techniques suppléants aux handicaps sensori-moteurs, en particulier la vision prothétique, nous avons conçu un dispositif de substitution de la vision en 3 dimensions. Une plate-forme logicielle nous permet de réaliser des expériences de substitution visuelle 3D en utilisant une matrice d'électro-stimulation linguale (TDU). Une thèse a été soutenue sur ce sujet (soutenance en octobre 2015) ; elle visait à déterminer la manière dont est perçue l'information visuelle 2D ou 3D via ce dispositif (voir figure 7.10). Il s'agissait de quantifier la qualité de cette perception 2D et 3D par des expériences d'identification de formes et des expériences d'immersion 3D couplant modèles 3D d'environnement, géolocalisation et détection de la posture. Ceci permettait de valider pour la vision prothétique plusieurs technologies de traitement de l'information visuelle en se fondant sur le paradigme de substitution visuelle. Ce travail se situait à l'interface entre informatique et sciences cognitives. Il reposait sur le déploiement et l'utilisation de systèmes haptiques, de capteurs et de systèmes de géolocalisation indoors (en laboratoire), et nécessitait la mise au point de plate-formes logicielles intégrant le contrôle des systèmes d'acquisition et de suppléance, et la modélisation de scènes 3D.

L'objectif de ce type travaux est, à terme, de développer un dispositif de substitution visuelle pour restituer à des personnes aveugles ou déficientes visuelles (ex: personnes atteintes de dégénérescence maculaire liée à l'âge) ou fournir des solutions de réalité augmentée pour la perception d'information de guidage et critiques dans un environnement en trois dimensions. Il s'agissait aussi

de transposer les développements faits sur ce dispositif à un système de vision prothétique développé par I. Marc et G. Tatur. Des échanges fructueux ont été fait avec cette équipe. Ce même dispositif est aussi déclinable pour des applications de guidage du geste chirurgical et pour des applications de positionnement spatial et de suivi de trajectoire en astro et aéronautique.

La réalisation de ce projet passait par un travail exploratoire fondamental en neurosciences cognitives: (i) il s'agissait d'obtenir une bonne qualité de la restitution d'informations spatiales tridimensionnelles sur un tissu faiblement discriminant (muqueuse linguale ou palatine) en comparaison de l'œil ; (ii) d'étudier la qualité de perception de signaux vectoriels en 2D ou en 3D projetés sur une seule matrice d'électrodes ; (iii) de s'intéresser à la virtualisation 3D de scènes vectorisées par une projection sur 2 matrices d'électrodes en même temps, l'une (inférieure) sur la langue, l'autre (supérieure) sur le palais ; (iv) enfin, afin de permettre une restitution fidèle de la scène projetée via le dispositif électro-tactile, il fallait que la géolocalisation et la posture du sujet dans l'environnement soient connues et précises. Dans ce contexte, la qualité du couplage haptique a été étudiée. Un TDU haute résolution a été fabriqué par la société SMEE à Voiron. Nous avons travaillé étroitement avec eux pour repenser entièrement ce dispositif (codage numérique, miniaturisation, haute résolution, protocoles de communication). Enfin, nous avons développé (avec C. Scheiber, CHU Bron) une approche anatomo-fonctionnelle fondée sur la détection par IRMf de l'activité des aires dédiées à la vision et au traitement de l'information 3D dans le cerveau au cours de telles expériences d'électro-stimulation linguale. Quatre articles ont été publiés dans les actes du congrès ASSISTH 2011, dans Acta Biotheoretica 2012 et 2 dans la conférence internationale EMBC en 2013. Trois communications orales ont été faites sur le sujet. Bien que tardives, deux autres publications sont encore prévues pour ce travail très en avance dans le domaine, en particulier sur l'effet de la bimodalité et sur des expériences de contradictions et de fusion d'information aux résultats spectaculaires.



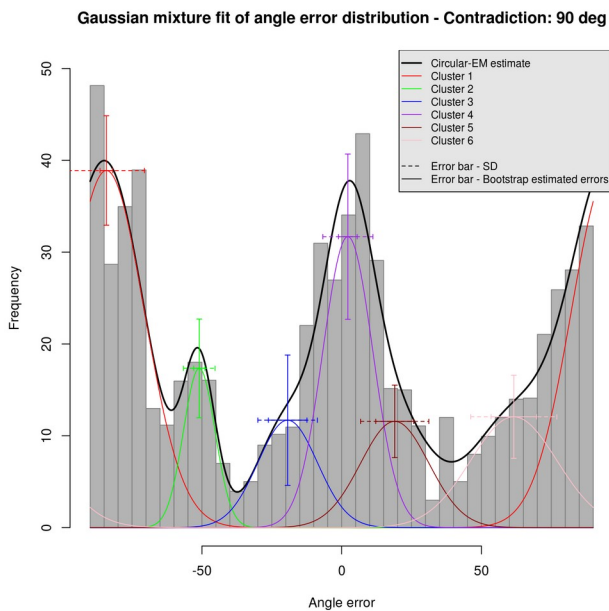


Figure 7.10. Tongue display unit. En haut à gauche, A. Chekhchoukh (thèse) mettant au point une expérience destinée à mesurer l'effet des micro-saccades ou d'un scintillement du signal sur la langue, par l'intermédiaire du TDU. En haut à droite, une forme en 3D (un cube) visualisé sur l'écran en niveaux de gris représentant les niveaux de voltage sur le TDU et en 12x12 pixels (taxels sur le TDU). En bas à gauche, résultats d'une expérience de contradiction d'information : une droite est tracée avec un décalage de 90° sur le TDU et sur le visu ; les sujets doivent indiquer la direction ressentie de la droite. En plus des réponses principales centrées sur 0 et 90° par rapport à la référence visuelle, des modes de fusion intermédiaires apparaissent. En bas à gauche, la matrice en 24x24 taxels du TDU haute résolution.

7.8. Modélisation multi-échelle en épidémiologie (2013-2016) [*]

Durant quelques années, je me suis intéressé à l'épidémiologie des maladies transmissibles. Avec mon troisième doctorant, Hessam Hessami, nous nous sommes particulièrement intéressés d'une part à des indicateurs épidémiologiques semblables à la notion de charge en électricité (un débit de charges électriques multiplié par un temps) et l'identification d'évènements infectieux dans le cadre des systèmes critiques dont l'existence est ponctuée de catastrophes, d'autre part de structures de population.

Nous avons donc initié une description des évènements infectieux dans un contexte où le suivi de l'épidémie peut se faire à l'échelle individuelle, comme c'est le cas en simulation individu centrée ou dans des implémentations en temps continu de modèles différentiels de propagation épidémique. Nous avons étudié l'intégrale temporelle de tels évènements au cours desquels l'épidémie augmente (catastrophes) ou diminue (rescues) et montré que les distributions des évènements épidémiques peuvent suivre des lois de puissance (figure 7.11). Tout du moins constituent-elles des signatures des maladies épidémiques et de leurs conditions, signatures qui peuvent être recherchées dans les systèmes réels (ex : dans le cas de covid dans lequel on dispose d'un suivi très fin). De telles signatures pourraient être utilisées pour confronter les modèles et affiner leurs paramètres. Ils pourraient également servir à tester, en simulation, des conditions pour répartir différemment la charge infectieuse (il vaut mieux plein de petits évènements infectieux disjointes qu'une crise

épidémique), par exemple (et c'est ce que nous venons de vivre pendant la longue période covid qui me permet d'en parler maintenant), le confinement massif, partiel, le port du masque, les gestes barrière, le pass sanitaire, *etc.* En développant ces réflexions, nous sommes inmanquablement arrivé à l'écueil de comment accéder à une telle granularité dans la connaissance de la propagation épidémique, et très vite à celui de comment prendre correctement en compte des éléments essentiels dans la transmission des maladies, en particulier la structure des populations, mais aussi leur composition, leur dynamique propre, les aspects sociologiques, *etc.*

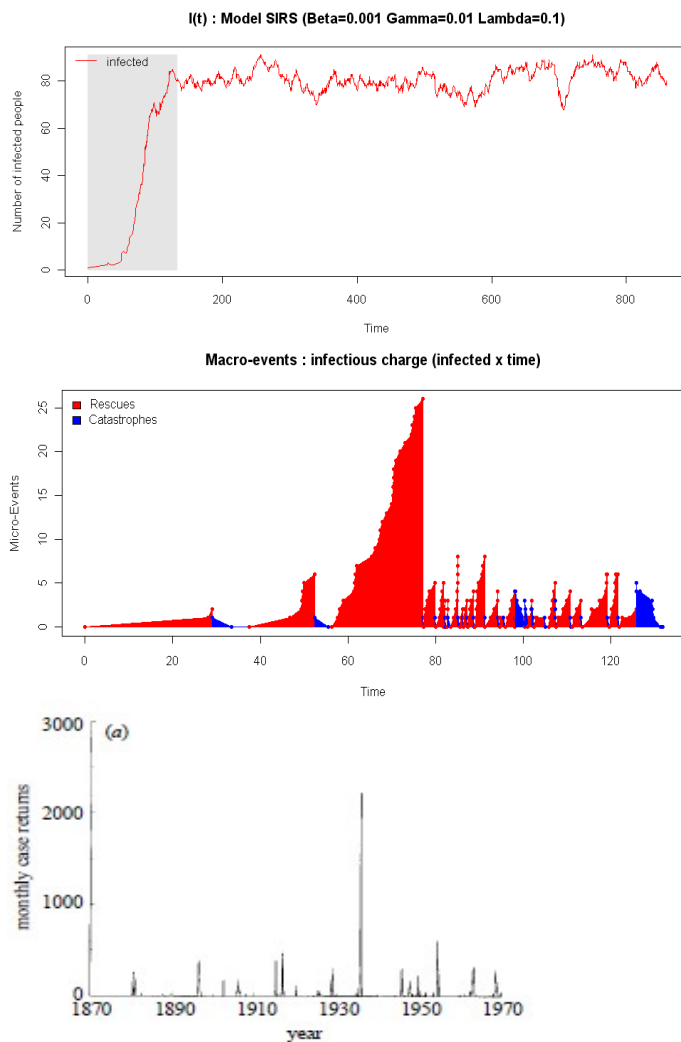


Figure 7.11. Evolution stochastique d'une épidémie et événements épidémiques.

En haut, la simulation montrée, très simple, met à jour de façon stochastique l'effectif d'individus infectés dans une population. On voit qu'au début de l'infection la croissance de cet effectif est très importante, mais pas nécessairement régulière. Elle est effectivement ponctuée de pauses, voire de régressions, accidentelles dans la croissance de l'épidémie.

La figure du milieu montre ces phases de croissance ou de décroissance contiguës identifiées comme des phases notées *rescue* (ici des phases de croissance infectieuse) ou *catastrophe* (ici des phases de décroissance de l'infection). Leur intégration définit la charge infectieuse (si augmentation) ou la charge sanitaire (si décroissance).

Les deux figures du bas montrent un enregistrement des épidémies de rougeole aux Îles Féroé, en effectif à gauche et en diagramme de fréquence à droite. La distribution de taille de ces épidémies suit une loi de puissance.

Nous nous sommes aperçus en lisant la littérature que ce n'était pas le cas ! Les modèles sont essentiellement fondés sur des systèmes d'équations différentielles ordinaires sans inclure de structure de population, en particulier de groupe (c'est dans les groupements d'individus que se transmettent principalement les maladies) ; pire, ces modèles ne prennent pas correctement en charge le rôle d'obstacles à la propagation épidémique des personnes immunisées (naturellement ou vaccinées) !

Nous nous sommes attelés à l'intégration de la notion de groupe dans plusieurs types de modèles,

dans des systèmes d'équations aux dérivées ordinaires pour commencer en réécrivant les équations d'un système SIR comme un système enzymatique. Nous montrons notamment que les individus immunisés interviennent dans l'équation de catalyse (génération de personnes infectées par transmission de la maladie à des personnes saines) comme des inhibiteurs non compétitifs.

Nous avons ensuite créé un modèle de structuration de la population en groupes, fondé sur des lois de probabilité pour l'instant de création d'un groupe, la taille de ce groupe et sa durée de vie. Ces lois dépendent de paramètres que nous avons explorés et qui sont reliables à des valeurs mesurables ou pouvant au moins être estimées dans des systèmes humains (ou animaux) réels, comme la distribution des vitesses de déplacement des individus, la surface effective qu'occupe une personne, les distributions d'attraction ou de répulsion sociale, *etc.* Nous avons implémenté ce modèle de structuration de population dans divers modèles de dynamique infectieuse : un modèle stochastique en espace homogène ou encore un modèle multi-agents spatialisé. Nous avons travaillé sur la convergence des résultats entre modèles (figure 7.12).

L'ensemble de ces travaux a été encore peu publié faute de temps. La publication principale est un important chapitre de livre [Glade, 2015].

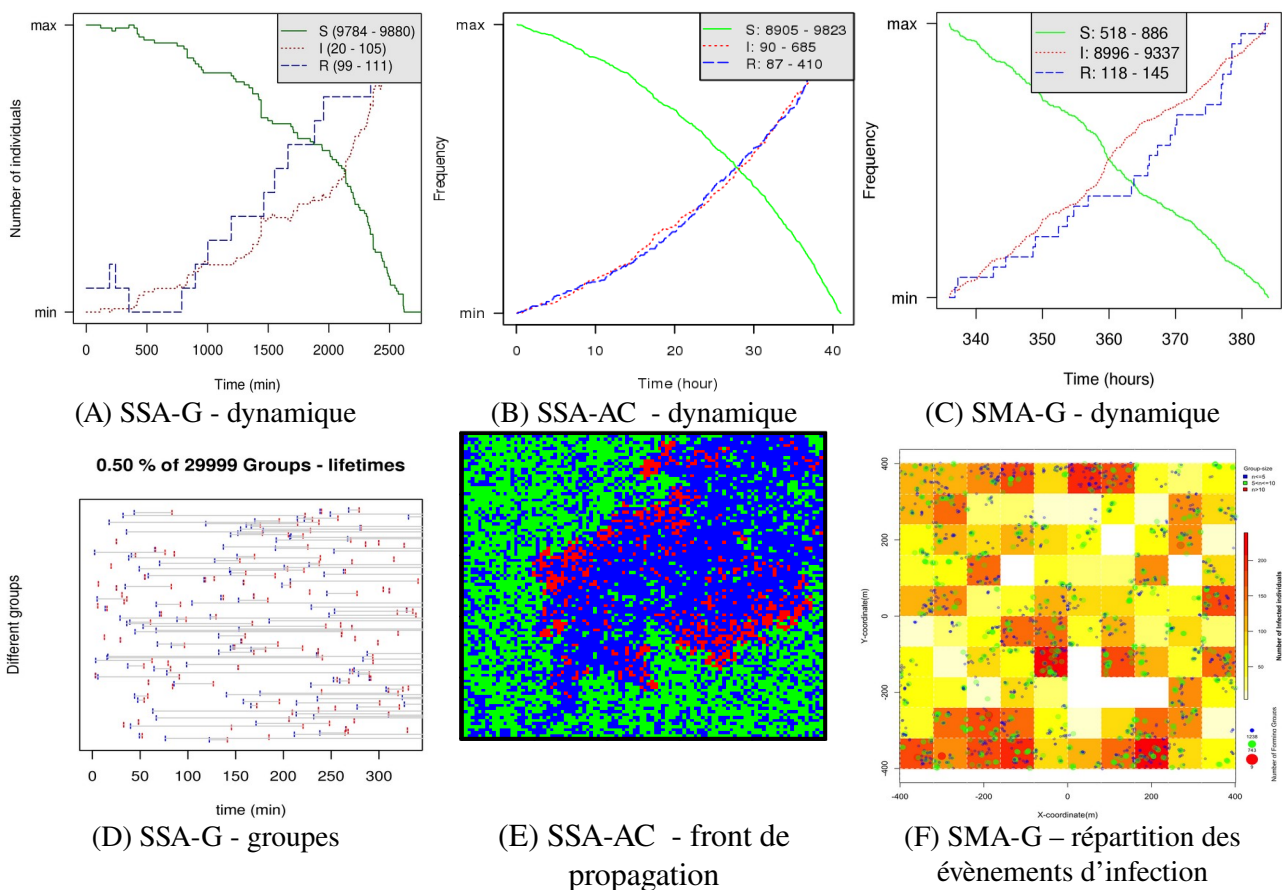


Figure 7.12. Simulations numériques de 3 modèles SIR convergant avec les mêmes paramètres [voir Glade, 2015]. Les 3 sont des modèles de simulation stochastique (SSA) fondés sur l'algorithme de Gillespie [Gillespie, 1976, 1977]. Le premier (A et D) est un modèle SSA simulant le processus infectieux dans une structure de groupe dynamique. Les groupes (D) sont de durée de vie relative à leur taille. Le deuxième (B et E) est un SSA appliqué dans un automate cellulaire ; la dynamique spatio-temporelle de l'épidémie est suivie à large échelle. Le 3ème (C et F) est un SSA avec une structure de groupes dynamiques simulés dans un système multi-agents.

Je vais ajouter une remarque : j'aurais peut-être du en 2020, avec l'arrivée de Sars-Cov-2, relancer franchement cette activité. J'en ai ressenti la nécessité mais j'ai du ranger ma frustration. Dans le contexte 2020-2021, étant donné que mon activité était centrée sur les sujet que je vais présenter dans les chapitres 8 et 9, et que le travail de laboratoire, celle des étudiants du labo notamment, et les enseignements étaient fortement affectés par la pandémie, il eut été difficile de réaliser un travail acceptable en essayant de maintenir difficilement le reste à flot. Cependant, j'ai suivi la pandémie avec attention et effectué une bibliographie très importante et très large spectre allant de la dynamique épidémique, à la formation des variants de ce virus, en passant par la stratégie sanitaire, vaccinale en particulier, à l'action immédiate du virus sur les organismes, aux covid longs, au zoonoses inverses ... Toute dramatique qu'elle soit, cette pandémie est un laboratoire à ciel ouvert ! J'ai mis à jour mes cours d'épidémiologie et compte bien relancer mon activité de recherche dans ce domaine, en particulier m'intéresser à la dynamique de la variabilité virale (son impact mais aussi sa formation) dans les modèles que j'ai proposés.

En 2013-2016, Hessam Hessami et moi avons imaginé (voir sa thèse [Hessami, 2016]) que des sociétés de conseil puisse proposer des services d'optimisation des rythmes et environnements humains (par exemple, conseiller l'architecture d'un bâtiment pour faciliter, voire promouvoir les gestes sanitaires comme le lavage de main), mais aussi de formation pédagogique et de sensibilisation dans les entreprises et institutions, pour limiter les propagations épidémiques (liées à des maladies transmissibles, mais aussi à d'autres formes de troubles transmissibles liés à des dynamiques d'opinions). A l'époque, nous pensions beaucoup à des maladies transmissibles à fort impact sur l'activité des entreprises, comme la gastro-entérite. Les propositions que nous arions faites comme je l'ai évoqué, paraissaient presque de la science fiction chez nous ! C'est ce que j'avais pu mesurer lors d'une conférence que j'avais donnée en tout début d'après-midi devant une soixantaine de personnes en demandant qui s'était lavé les mains avant de passer à table (personne ; la disposition des sanitaires ne facilitait alors pas du tout la démarche). De nos jour, nous mesurons l'importance de ceci avec d'une part les mesures sanitaires individuelles (lavage des mains, distanciation, masque, aération des pièces) mais aussi les mesures sanitaires plus larges échelle comme les files de circulation, les confinements et couvre feux, le télétravail, les passes sanitaires et vaccinaux). L'importance de tout ceci peut être finement estimé avec des modèles, en particulier ceux que nous avons proposés. Avec une pointe d'humour, je faire remarquer que si nous avons créé ce genre de société en 2016, nous serions probablement riches à l'heure qu'il est !

Quoi qu'il en soit, Hessam Hessami a tout de même fini, après un passage dans l'industrie en tant que Data Scientist chez Servier, par créer sa société, *DONEH*, société dont l'activité sera la mise en relation d'études de santé publique avec les personnes se proposant comme sujets d'étude. Cette activité permet notamment des enquêtes épidémiologiques variées. J'y contribue en tant que personne conseil et ai initié et tutoré un projet étudiant consistant à mettre en place une étude fictive (sur la mort fœtale), fondée sur une étude existante, le but du projet étant de spécifier complètement le cadre légal (RGBD, CPP, ...) dans lequel cette société basée en Angleterre pourra proposer ses services.

7.9. Effet des champs externes faibles sur la matière vivante [*]

Groupe de travail. Eric Beaugnon (LNCMI, CNRS Grenoble), Caterina Tomba (Univ. Geneve), Nicolas Glade (TIMC-IMAG, CNRS – UGA Grenoble)

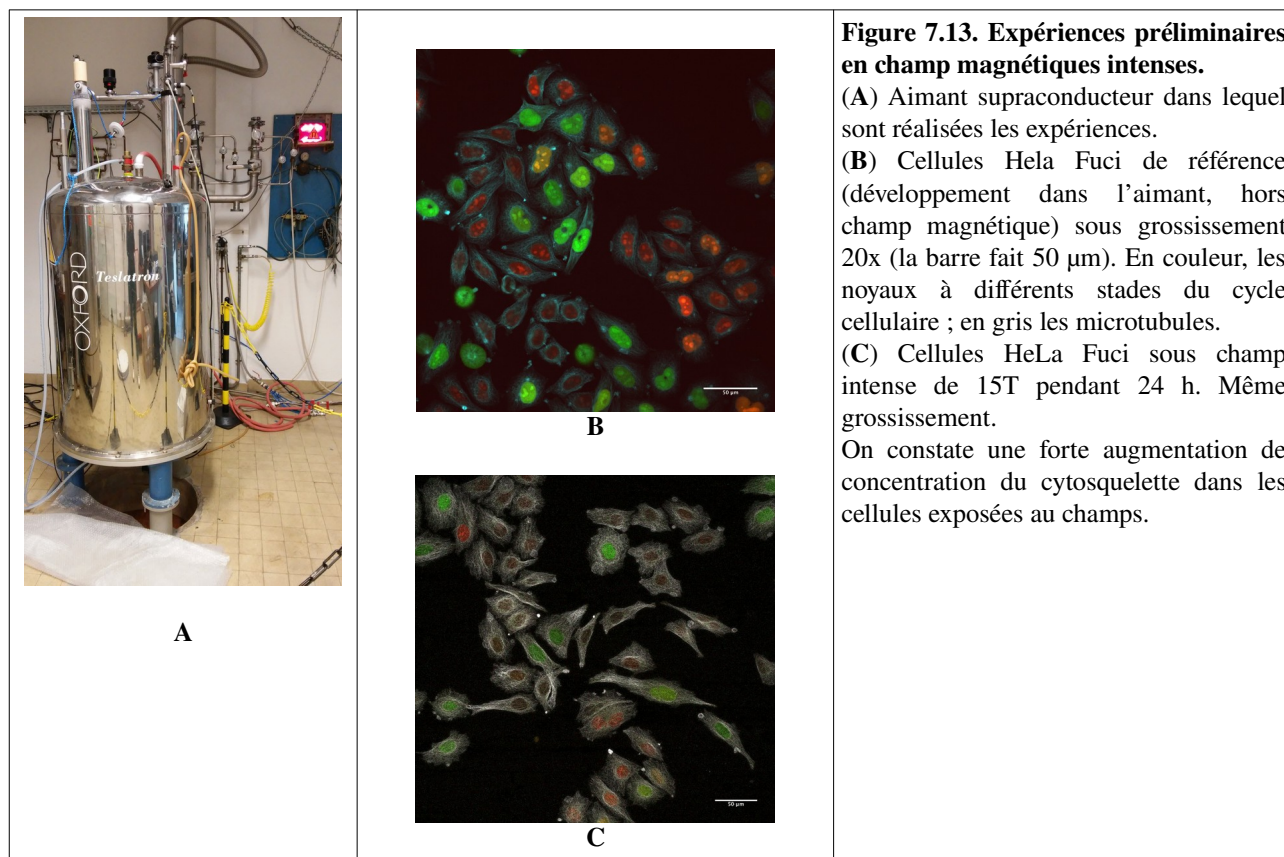
Cette thématique concerne la mesure expérimentale des effets des champs magnétiques intenses sur le vivant principalement, mais également sur l'eau.

7.9.1. Effet des champs magnétiques intenses sur le vivant

Au cours de ma thèse, j'ai étudié l'effet des champs magnétiques intenses sur les solutions de microtubules *in vitro*. Les microtubules, *in vitro*, sont très sensibles à l'effet des champs externes faibles (voir les travaux antérieurs), notamment aux champs magnétiques intenses. Ces travaux ont donné lieu à deux publications [Glade & Tabony, *Biophys. Chem.* 2005 ; Glade et al, *Biophys. Chem.* 2006].

Un de mes souhaits est de réaliser de nouveau ces expériences *in vivo*. J'y reviendrai probablement dans les années à venir. L'objectif est de montrer que des structures cellulaires magnétiquement anisotropes, donc sensibles à l'action d'un champ magnétique intense, comme les microtubules ou l'actine, provoquent des modifications de morphologie et de comportement des cellules lorsque celle-ci sont soumises à un champs magnétique intense. Nous souhaitons mesurer les champs auxquels les cellules sont sensibles et le moment auquel elle le sont. Nous pensons notamment que celle-ci sont sensibles lorsque le cytosquelette se réarrange durant la vie cellulaire, en particulier pendant la phase de division cellulaire. Ces expériences ont un double intérêt, médical d'une part (les appareils d'imagerie par résonance magnétique nucléaire travaillent à des champs magnétiques comparables à ceux que nous utilisons, c'est-à-dire entre 2 et 8 T); instrumental d'autre part : les champs magnétiques intenses, à l'instar des pinces optiques, pourraient s'avérer être un bon moyen de manipulation spécifique des structures cellulaires magnétiquement anisotropes.

Ces expériences sont cependant difficiles à réaliser car il faut mettre en place tout un setup pour les expériences de culture cellulaire dans l'aimant supraconducteur dans lequel nous réalisons ces expériences. Avec l'aide de Caterina Tomba (Postdoc, Univ. Genève), nous avons mis en place un tel setup avec notamment la régulation fine de la température et des gaz nécessaires à la culture cellulaire. Nous avons aussi réalisé des expériences préliminaires (Figure 7.13) sur des cellules HeLa Fuci (marquant les étapes du cycle cellulaire) sous champ. Nous avons réalisé ces expériences plusieurs fois pour notamment résoudre toutes les difficultés rencontrées dans ces expériences (culture, manips sous champ, microscopie dans un autre lieu ...). Nous avons ainsi pu mettre au point un protocole expérimental et souhaitons continuer ces expériences dans les années à venir sous la forme d'un sujet de M2 puis de thèse. Caterina Tomba en a fait une partie de son projet pour candidater au CNRS.



7.9.2. Effet des champs magnétiques intenses sur certaines propriétés de l'eau

Une autre influence possible des champs magnétiques intenses sur le vivant concerne l'eau du milieu cellulaire elle-même. Il a été montré par Otzuka & Ozeki [Otzuka 2006] que certaines propriétés de l'eau, liquide bien plus complexe qu'on ne le pense habituellement et qui se comporte davantage comme un continuum que comme une collection de molécules H_2O agitées et en interaction par l'intermédiaire de liaisons hydrogène, en biologie notamment, puissent être modifiées sous l'action de champs magnétiques intenses (voir le projet PEPS 2016 ci-dessous pour les détails), en particulier sa mouillabilité sur des surfaces métalliques. Un étudiant qu'Eric Beaunon a encadré (et dont j'ai été le tuteur universitaire), Bao Chen, a reproduit les expériences d'Otzuka avec succès, montrant que de l'eau, sous légère agitation, exposée à un champ magnétique intense de 8 teslas, change peu à peu de propriétés de mouillage sur des métaux. Il montre que l'effet est clairement lié à la charge ionique de l'eau. De plus, il montre que ce phénomène a une longue rémanence puisqu'une fois le champ coupé, les propriétés de mouillage de l'eau préalablement exposée continuent d'être affectée pendant plus d'une heure. (Figure 7.14).

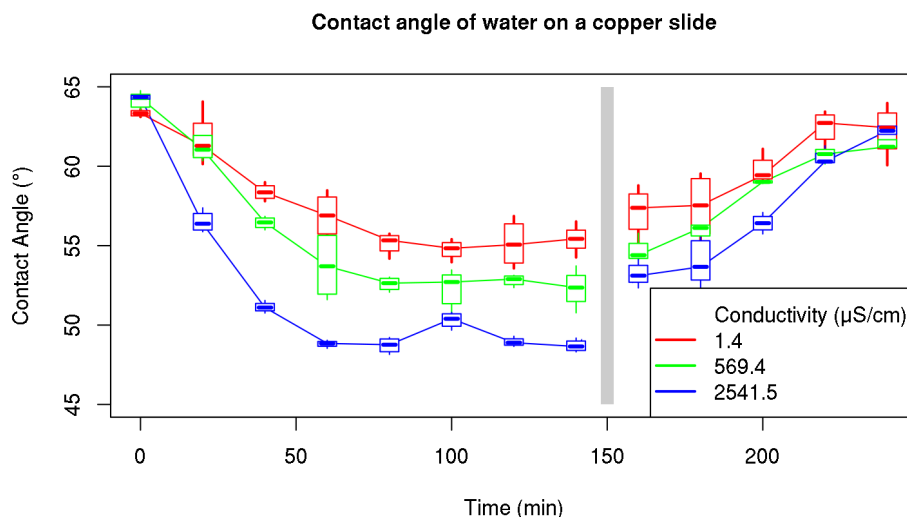


Figure 7.14. Angle de contact de gouttes d'eau (20 μl) déposées sur du cuivre propre. Les 3 courbes, rouge, verte, bleue, représentent respectivement de l'eau ultra-pure (conductivité très faible de 1.4 $\mu\text{S/cm}$), et de charge ionique croissante ; l'eau à 2541.5 $\mu\text{S/cm}$ est de l'Hepar. L'eau est placée dans un flacon dans un champ magnétique de 8 T depuis le temps 0 de l'expérience, le flacon étant maintenu mobile sous champ (léger mouvement de haut en bas ; ce point n'est pas compris mais s'avère indispensable). Toutes les 20 minutes, le flacon est retiré du champ magnétique pour prélever quelques gouttes et mesurer leur angle de contact sur le cuivre. On voit que plus l'exposition au champ a été longue et plus l'eau est chargée en ions, plus l'angle de contact sur le cuivre chute. Au temps 150 minutes (2h30), le champ est coupé. On constate que l'effet est rémanent pendant environ une heure, sa relaxation étant lente.

Nous pensons que ces modifications sont liées à une structuration plus forte de l'eau en clusters. Nous pensons aussi qu'ainsi d'autres propriétés pourraient être affectée comme sa viscosité. Les milieux biologiques étant composés en grande quantité d'eau (solvant), il était légitime de se poser la question de l'effet de telles modifications des propriétés physico-chimiques de l'eau sur la matière vivante. Dans ce cadre, nous avons :

- déposé un projet PEPS en 2016 (voir Annexe 1.3), ainsi qu'une demande à l'ANR. Les deux ont été refusées. Les avis sur le PEPS étaient néanmoins très bons. Je ferai une nouvelle demande lorsque l'occasion (appel à projet + disponibilité des participants) se présentera.
- entamé des expériences pour mesurer par spectroscopie de fluorescence et par pH-métrie d'éventuels effets du champs magnétique sur l'eau. Dans ce cadre, nous avons encadré une étudiante de L3 (Anita Ravat - B. Sc. internship 2016 - Tutors : Eric Beaugnon and Nicolas Glade, *Effets des champs magnétiques intenses sur le potentiel hydrogène (pH) de solutions aqueuses*). Les résultats sont prometteurs car nous avons effectivement observé des modifications du pH (Figure 7.15) ainsi que des changements du spectre de fluorescence. Ces modifications pourraient être imputables à un renforcement des clusters d'eau, organisations éphémères de l'eau pouvant se renforcer autour d'ions en solution (voir l'Annexe 1.3 pour les détails scientifiques de ce projet sur lequel je vais travailler à nouveau à partir de fin 2022).

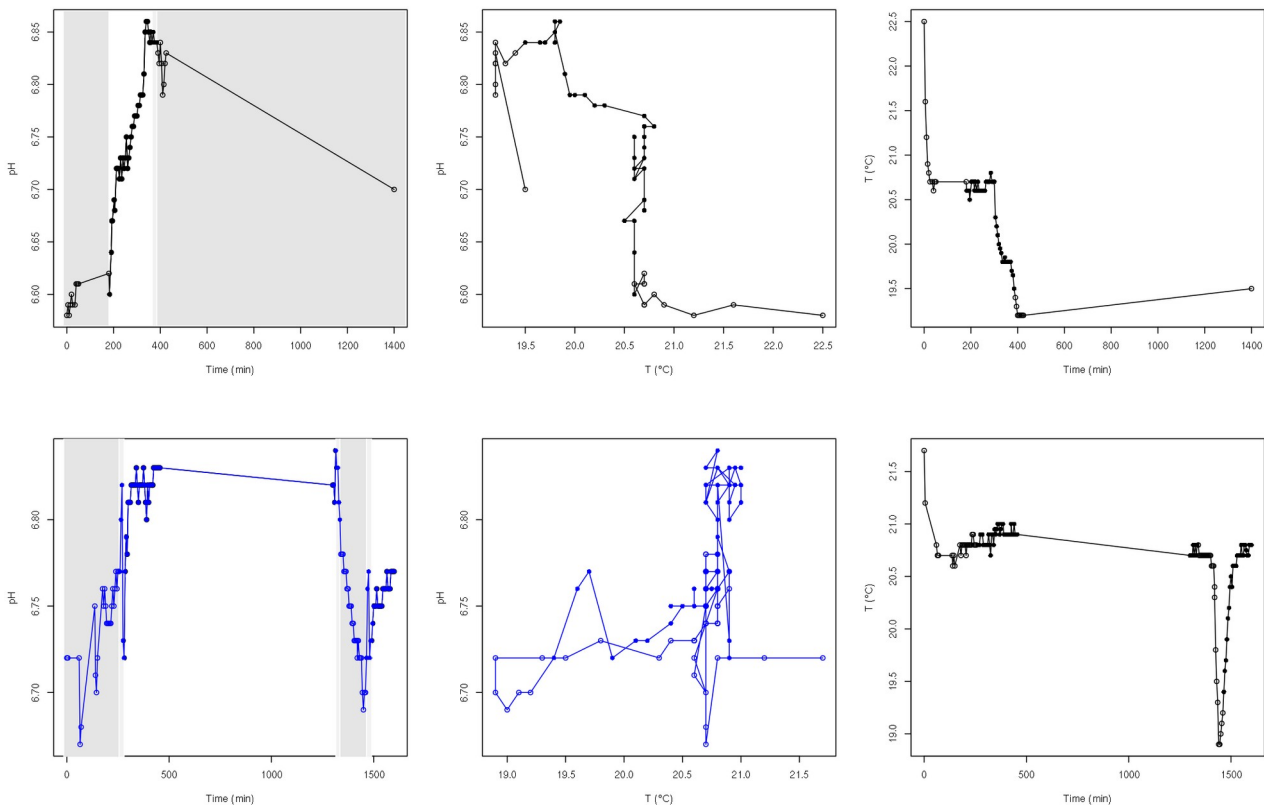


Figure 7.15. Expériences sur le pH de l'eau en champ magnétiques intenses. Chaque ligne de graphiques ci-dessus correspond à une expérience. La sonde d'un pH-mètre et un thermocouple étaient plongés dans un tube fermé d'eau placé dans un champs magnétique intense (8T). Les enregistrements ci-dessus montrent les variations du pH et de la température de l'eau au cours du temps, alors que le champ est coupé (zones grises du premier graphique) ou actif (zones blanches du premier graphique). Ces résultats semblent aller dans le sens d'une variation du pH liée au champ magnétique. Ils sont cependant à prendre avec prudence car les deux instruments mesurent des déplacements de charges via un courant électrique ; dans un champ intense, ceci pourrait être artefactuel.

- mis au point un instrument pour mesurer la viscosité et l'indice de refraction de l'eau sous champ. Ce dispositif consiste, suite à impulsion (pulse symétrique latéral d'air) ou à une série d'impulsions de fréquence donnée, à faire vibrer une goutte d'eau en lévitation magnétique (nous obtenons une lévitation magnétique sous fort gradient de champ, obtenu dans notre aimant supraconducteur à 15T) et à mesurer la relaxation de cette goutte. Nous mesurons la vibration de la goutte à l'aide d'un appareil permettant de mesurer par reflexion d'un spectre lumineux sur la goutte la distance entre l'appareil et la surface de la goutte. Les modes de vibration et la relaxation de la goutte, ainsi que les modes de résonance, nous donnent accès à plusieurs propriétés de l'eau. Après avoir développé un prototype et l'avoir testé avec succès (voir Figure 7.16) nous avons fait construire un appareil de mesure et réalisé des séries de mesures. Nous sommes actuellement en train de développer un modèle de la goutte et de ses déformations pour calculer sa viscosité à partir des mesures faites. Les mesures de divers liquides non sensibles au champ magnétique (ex: alcool) seront faites (nous avons déjà vérifié que les propriétés de l'éthanol sont inchangées), ainsi que celles de

l'eau plus ou moins pure.

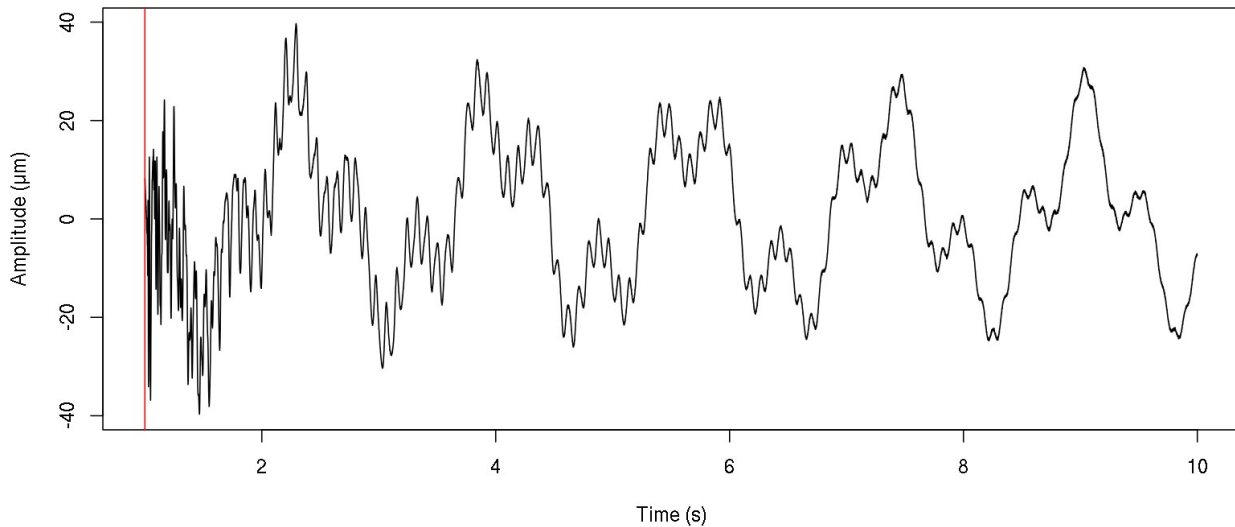


Figure 7.16. Relaxation des vibrations d'une goutte en lévitation magnétique après un pulse symétrique intense. On voit clairement différents modes de vibration. De cette relaxation peuvent être extraites diverses propriétés du liquide, notamment sa viscosité.

- nous avons aussi réalisé des expériences de dynamique enzymatique dans de l'eau placée sous champ (comme décrit dans le projet). Les résultats ne sont pas probants.

8. TRAVAUX DE RECHERCHE ACTUELS

Mon activité de recherche se situe principalement entre la biologie théorique et l'informatique théorique. Toutes les recherches que j'ai entreprises en biologie théorique et qui motivent les développements futurs que je souhaite mener ont un dénominateur commun. Elles posent la question du rôle et de l'importance de l'organisation spatiale et temporelle dans le vivant.

L'objet biologique de mes recherches a souvent varié, fluctuant entre l'organisation spatiale des microtubules, la chimie prébiotique et l'ordre aux origines du vivant, les réseaux de régulation, *etc.* Derrière ces objets, ce qui me fascine c'est le fait que nous (humains) identifions de l'ordre et une complexité dans ces systèmes, que nous comprenions que ceci joue un rôle dans le fonctionnement du vivant. Ce qui m'interroge, c'est dans quelle mesure cet ordre, ces formes, cette complexité sont en relation avec le fonctionnement naturel du vivant. A cette question sont liées d'autres interrogations, en particulier la relation qu'entretient le vivant avec *la machine* ou tout autre processus de calcul (*computational systems*) : Premièrement, le vivant est-il assimilable à une machine ou à un processus de calcul ? Deuxièmement, qu'est ce que ça change de considérer cela comme idée sous-jacente du vivant ?

Ci-dessous, je décris le cadre général des travaux qui constitueront mon activité de recherche à venir, je les décris brièvement sous une forme abrégée, enfin je détaille spécifiquement certaines parties sous la forme d'un programme de recherche.

8.1. Cadre général de mes travaux de recherche actuels.

Le paradigme habituel en biologie est proche d'une vision de la *cellule-machine*, c'est-à-dire dans laquelle ses composants se comportent comme des éléments mécaniques passifs ou comme des éléments actifs comparables à des micro-machines ; une cellule-machine donc, mais dans laquelle on (les chercheurs, les étudiants, les humains) a conscience parfois qu'il y a « du bruit ». Le vocabulaire même employé en biologie est déformé par cette vision particulière de la cellule (et même plus globalement du vivant) faite par les hommes, une espèce douée pour l'ingénierie. Ainsi, au niveau cellulaire, on parle de moteurs moléculaires marchant sur les fibres du cytosquelette, fibres que l'on associe facilement à des poutres et des tendeurs, on parle de centre organisateur, on associe souvent les "unités" de synthèse de l'ATP ou des protéines à des centrales énergétiques de la cellule et à des usines de production. Les contours mêmes des particules biologiques ont dans notre esprit, dans nos représentations, des géométries bien définies, bien nettes : ce sont des fibres, des tubes, des membranes, des globules, *etc.* On en a une vision géométrique, certainement pas protéiforme. Une machinerie bien huilée donc, parfaitement contrôlée par les centres organisateurs et par l'ADN qui a longtemps été au centre de l'organigramme cellulaire (actuellement le centre de gravité du contrôle cellulaire se déplace cependant légèrement vers les « réseaux de régulation », un autre terme emprunté à l'ingénierie), avec un système de signalisation si efficace qu'on le présente aux étudiants à l'université comme un centre de tri postal.

Je parais prendre ici en défaut ce paradigme ; néanmoins, il est évident celui-ci fonctionne plutôt bien puisqu'on sait même l'exploiter pour fabriquer des molécules biosynthétiques (ex: insuline) en contrôlant ou en « reprogrammant » certains gènes. On va jusqu'à « reprogrammer » des réseaux génétiques entiers (les pirater devrait on dire!) pour que des épidermes, à l'échelle d'un organisme (une souris), se comportent comme Turing ou Belousov et Zhabotinski l'avaient prédit, comme Murray l'avait mis en équations [Belousov 1951 & 1985, Turing 1952, Zhabotinski 1967, Murray 1993], à savoir produisent des motifs spatio-temporels stationnaires ou sous forme de vagues de pigmentation se propageant sur l'animal au cours de sa croissance [Suzuki 2003], démonstration impressionnante s'il en est d'ingénierie génétique¹⁴. C'est aussi un paradigme utile pour le modélisateur et le mathématicien qui cherchent à simuler ou à mettre le vivant en équations. Nos représentations (mathématiques ou informatiques) de la connaissance des systèmes biologiques, nos méthodes de calcul, nos algorithmes sont adaptés à ce paradigme. Il est en effet bien plus facile de mettre en œuvre des modèles du vivant fondés sur des objets structurés interagissants entre eux selon des lois bien définies (même si elles comportent une part de hasard) que de fonder nos modèles sur le principe même du désordre ambiant (c'est cependant ce qu'entreprend la physique statistique notamment).

Si le paradigme (ou les théories) en vigueur en biologie tient bon, c'est parce qu'il est utile, et les scientifiques ont raison de continuer à développer leurs recherches dans ce cadre, tant que ses limites ne se font pas trop ressentir. C'est aussi notamment parce que la biologie (cellulaire surtout) est une science encore jeune (la biologie moderne débute vraiment avec la découverte de la structure de l'ADN, du code génétique et de la machinerie de réplication, traduction et transcription, c'est-à-dire entre 1950 et 1970¹⁵) et certainement encore trop conservatrice en comparaison de la physique qui a déjà subi plusieurs révolutions lourdes l'obligeant à accepter le fait que revoir sa copie, même de façon importante, fait partie du processus scientifique. Ce n'est pas encore facilement admis en biologie, mais on ne doit pas oublier qu'il fut tout autant difficile à la physique d'accepter de laisser le monde tout newtonien pour prendre en compte, peu à peu, les propositions faites par les théories physiques relativistes notamment, quantique aussi. C'est cela qui a permis l'apparition de la grande majorité des dispositifs technologiques dont on dispose aujourd'hui. Il eut été impossible de concevoir le LASER ou le GPS dans le contexte des physiques ondulatoire et newtonienne. C'est le besoin, en termes de compréhension (ex : c'est l'échec de la compréhension de la catastrophe ultraviolette à la fin du 19^{ème} siècle qui poussera à la recherche d'une nouvelle physique et qui permettra de retenir la théorie quantique de Max Planck) comme de production (l'industrie, celle liée à la guerre notamment au 20^{ème} siècle, a poussé la physique dans ses retranchements et à contribué à de nombreuses découvertes dans des temps sombres toutes très courts : radar, chiffrement/déchiffrement, calcul numérique, matériaux modernes, atome ...), qui a forcé et retenu ces transitions. La physique newtonienne n'est pour autant pas mise au rebut puisqu'elle est tout à fait

14 NB. Au demeurant, la modification génétique faite reste relativement simple car ne touchant que peu de gènes, mais l'ensemble de l'expérience demeure impressionnant car l'effet se manifeste très nettement, de façon macroscopique, à l'échelle d'une souris vivante.

15 C'est en tout cas là que je décide de la situer dans ce manuscrit.

efficace dans la majorité des applications humaines (construction, mécanique ...) qui fonctionnent à des échelles modérées de temps, d'espace et d'énergie.

En biologie, c'est aussi le besoin et les nécessités d'accès à des échelles de temps (très courtes), d'espace (très petites) et d'énergie (très faibles) qui pourrait bien commander ces transitions.

Le consensus de la cellule-machine vient de faits observationnels indiscutables et vraisemblablement de la naissance quasi simultanée de la biologie moderne (années 50) d'une part, de la théorie du contrôle (années 40), de l'informatique et de la cybernétique (années 40-70) [Triclot 2008, Chapoutier 2011]. D'abord, la biochimie et la microbiologie (agroalimentaire notamment), puis la biologie moléculaire ont montré que, dans plusieurs systèmes biologiques ou biochimiques, l'on pouvait contrôler avec succès enzymes ou gènes pour fabriquer des substances d'intérêt (par exemple, la synthèse d'insuline). Notons toutefois que d'une part de nombreux autres systèmes biologiques visant à être contrôlés sont passés à la trappe car n'aboutissant à aucun résultat exploitable ou intelligible ou même simplement digne d'intérêt pour l'industrie et que, d'autre part, le forçage effectué sur ces systèmes (ex: doses de substrats à transformer par les enzymes ou les cellules) est colossal, sans commune mesure avec leur fonctionnement naturel. En réalité, la plupart des systèmes naturels fonctionnent en forte limitation (quasi absence souvent) de substrat. Pour un système vivant, tout apport d'énergie ou de matière est potentiellement, rapidement (au delà d'une dose pouvant s'avérer faible), une perturbation de son homéostasie. Les systèmes vivants vivent donc dans les milieux qui leur sont les plus adaptés, c'est-à-dire les milieux dans lesquels les apports d'énergie et de matière (lumière, chaleur, composés organiques, ...) sont adaptés à leur homéostasie, donc à leur capacité naturelle à traiter ces apports en les tamponnant d'abord, puis en transformant matière et énergie en matière biologique et en éliminant l'excédent. Ensuite, il y a le biais observationnel qui revêt au moins 2 aspects : on ne sait observer (au microscope, RMN, diffraction de la lumière ...) que ce qui est structuré, organisé temporellement ou spatialement, pas ce qui nous apparaît comme « du bruit ». L'observateur, ingénieur par nature (c'est-à-dire qu'il est efficace à la recherche de structures et qu'il pense « machine »), en déduira que le vivant fonctionne grâce au ballet bien agencé de ces micro-structures fonctionnelles. Enfin, il y a le biais expérimental qui déforme notre compréhension de ce que sont les systèmes naturels dans leur milieu. Par exemple, lorsqu'on cherche à résoudre la structure des protéines, c'est-à-dire à connaître leur structure tridimensionnelle, on impose un double biais à notre compréhension de ce que sont en réalité les protéines. D'abord on part du principe qu'elles sont forcément structurées (en 3 dimensions) pour pouvoir fonctionner. Cette présupposition nous apparaît naturelle et provient d'une part de la nécessité présupposée de structures 3D pour la formation de super-assemblages (ex: microtubules) et de la façon dont nous pensons que les mécanismes biologiques fonctionnent, calqués sur le modèle de nos machines, par exemple le modèle clef-serrure enzymatique (et de ses variantes allostériques) utilisé pour décrire le « mécanisme » enzymatique, dans lequel l'enzyme et son ligand, pré-structurés, présentent respectivement un site actif bien formé correspondant à la forme du ligand, également bien définie ; une alternative dynamique (non pré-structurée) serait pourtant envisageable.

Ensuite, faire de la cristallographie des molécules biologiques¹⁶ implique de forcer les objets biologiques à se ranger comme des sardines dans leur boîte, c'est-à-dire à les forcer à s'agencer en cristal en les mettant dans un tube d'échantillons à des concentrations extrêmement élevées, des concentrations qui n'ont rien de naturelles et qui ne sont qu'extrêmement rarement rencontrées dans les cellules. On pourrait ici continuer à énumérer les très nombreux biais liés à la mesure et à l'observation qui déforment notre vision du monde biologique microscopique.

Pourtant, depuis quelques années, des études de plus en plus nombreuses tendent à montrer que de nombreux objets biologiques, dans leur mode naturel de fonctionnement, ne sont pas si structurés qu'on ne le pense (voire ne le sont pas du tout). C'est par exemple le cas de la protéine *Tau*, dont le rôle est essentiel dans la régulation de la stabilité des microtubules (naturellement instables). Ainsi il a été ainsi montré que les microtubules décorés, des structures pourtant assez organisées, présentent vraisemblablement un aspect de fil de laine angora [Roll-Mecak, 2015]. Cela pose la question de la façon dont se font les interactions entre protéines associées (par exemple les moteurs moléculaires) et microtubules dans un cadre où les objets en présence ne sont pas très structurés. C'est aussi le modèle clef-serrure lui-même qui commence maintenant à être revisité [Kulkarni, 2020], une meilleure vision du processus étant que, si sur une très grande quantité d'enzyme et de substrat (comme dans un bioréacteur agroalimentaire) le processus correspond en moyenne à ce que donnerait un modèle clef-serrure, à l'échelle des particules biologiques individuelles, le substrat comme l'enzyme ne sont pas forcément très structurés (ils peuvent se trouver dans un très grand nombre possible de conformations) mais se replient localement (ce n'est pas forcément la totalité de la particule qui se replie) au contact, et se déplient à nouveau aussitôt libérés ; il s'agit d'un modèle bien plus dynamique du vivant en somme.

Allons plus loin et posons la question de la pertinence des réseaux d'interaction biologique. De tels réseaux sont une abstraction nous permettant de ne mettre en exergue que les processus les plus fréquents et les plus observables dans une cellule, mais en aucun cas une réalité. Les interactions, des plus spécifiques (donc entrant dans la logique de tels réseaux) au moins biologiquement affines, sont d'une grande diversité. Certaines sont rares, d'autres sont fréquentes ; certaines sont fortes, d'autres faibles ; certaines durent longtemps, d'autres sont éphémères ; certaines font intervenir la structure et l'ordre, d'autres sont plus distribuées ou font intervenir un agrégat de particules subcellulaires, ce que sous-tendent respectivement les idées d'hyperstructures et de FDS (pour *Functioning-Dependent Structures*) des structures qui se forment à la demande d'une fonction par le système cellulaire, proposées par [Amar 2004, Norris 2002, 2007, Thellier 2004, 2006] ... ce qui nous amène naturellement à poser la question des origines du vivant. S'il est concevable

16 En quelques mots, la cristallographie des molécules biologiques consiste à 1) former un cristal de la molécule biologique d'intérêt en concentrant fortement cette molécule notamment, 2) à faire passer une source de lumière à travers et à observer la figure de diffraction qui en résulte. De cette figure de diffraction, par transformation de Fourier, on reconstruit le réseau cristallin d'atomes (des molécules en question) et on remonte ainsi à la structure des molécules étudiées. Le rayonnement utilisé est principalement des rayons X car ceux-ci possèdent une longueur d'onde plus courte que celles de la lumière visible, c'est-à-dire entre 0,001 nm et 10 nm, c'est-à-dire du même ordre de grandeur que les dimensions atomiques et les espaces inter-atomiques, permettant ainsi de résoudre finement la structure des molécules d'intérêt.

actuellement d'associer les processus du vivant actuel à une logique mécaniste dans un cadre structuré complexe, il reste à résoudre comment le magma chimique prébiotique a pu atteindre ce stade organisé, à commencer par l'émergence d'un code génétique et des premiers métabolismes. Les réflexions et recherches sur le vivant à ses origines sont un bon exercice pour tenter de cerner la nature même de la matière vivante ... car se situant justement à la limite de ce qui est vivant.

La très bonne correspondance entre ce que l'on pense du vivant et ce que l'on observe effectivement dans le vivant, dans le cadre du paradigme actuel de la théorie cellulaire, ne doit pas laisser occulter que justement, nous ne percevons et n'interprétons que dans le cadre d'une théorie actuelle du vivant : les projections mentales de nos observations ne se font que sur le paradigme du vivant actuel¹⁷. Sans remettre en question ce paradigme et son utilité, mes principaux collaborateurs (O. Bastien (LPCV CEA Grenoble), A. Stephanou (TIMC), I. Cheddadi (TIMC), E. Fanchon (TIMC), P. Ballet (LATIM, Brest), L. Trilling (TIMC), S. Ivanov (IBISC, Evry), E. Beaugnou (LNCMI, CNRS Grenoble) ...) et moi-même proposons de nous intéresser à une vision alternative du vivant microscopique dans laquelle la cellule peut fonctionner dans un état naturellement très peu organisé ou l'ordre n'est que transitoire et où persistent certaines structures plus ordonnées que d'autres qui biaisent (dirigent) le fonctionnement de la cellule vers certains comportements particuliers. Dans ce contexte peuvent être réinterprétés de nombreux aspects de la biologie comme des observations de cinétique enzymatique rapide non compatibles avec les modèles actuels, mais aussi la bio-toxicité des nanoparticules artificielles (nanotubes de carbone, fibres amiante) qui, si elle n'est absolument pas comprise à l'échelle microscopique (elle est pensée comme un processus cellulaire inflammatoire, sans qu'on en sache bien plus) pourrait être vue comme le résultat d'un biais fonctionnel du à un forçage structurel permanent dans la cellule (trop d'ordre maintenu trop longtemps). La diversité des réseaux de régulation génétiques ou métaboliques et leur plasticité dans une même cellule pourraient bien constituer également une nouvelle compréhension de la manière dont les processus cellulaires sont maintenus, permettant à la cellule de continuer d'assurer ses fonctions vitales.

Nous souhaitons tester la viabilité de ces assertions par la modélisation (mathématique et fondée sur des modèles numériques) d'une part mais également récolter davantage d'éléments d'observation et de mesure pouvant supporter nos propositions, soit dans la littérature soit à partir de nos propres expériences..

Passons maintenant à la description détaillées des thématiques de recherche que j'ai choisi de présenter dans ce document.

¹⁷ Le lecteur trouvera dans [Chalmer 1987, Kuhn, 2008, Barrau 2016] les discussions explicitant clairement à la fois l'intérêt et la contrainte des paradigmes en science, c'est-à-dire leur nécessité pour cadrer des recherches et éviter l'absence de noyau dans lequel peut se construire et évoluer scientifiquement un consensus, et inversement le biais cognitif qu'ils forcent en fixant sans équivoque l'ensemble des faits scientifiques, mécanismes, lois (issus d'un consensus dont on sait qu'il est valable un certain temps à une époque donnée) selon lesquels toute nouvelle observation doit être interprétée.

8.2. Origines du vivant

W. Martins affirme, à juste titre, que les recherches sur les origines du vivant sont une conjecture infalsifiable et que le mieux que nous puissions espérer est de disposer d'une histoire convaincante (... *research on the [origins of Life] topic is "unfalsifiable conjecture" — the best we can hope for is a convincing story.* W. Martins, from J. Whitfield, *Nascence Man*, [Whitfield 2009]). Je pense que cette question, au-delà des nouvelles attentes actuelles à ce sujet, en particulier en exobiologie, est un excellent cadre de départ pour réfléchir sur la nature même du vivant. Nous allons en tout cas contribuer à fournir quelques éléments plausibles de cette histoire.

Pour amorcer cette histoire, je vais commencer, dans ce chapitre, par contextualiser cette thématique ancienne car fondamentale pour l'homme, mais qui bénéficie d'un regain d'intérêt important à cause des développements importants de l'astronomie et de la conquête spatiale. Les recherches que je vais présenter ici focalisent sur l'établissement d'une génétique primordiale. Je montrerai comment des recherches expérimentales et théoriques (fondées sur des simulation numériques) peuvent être faites dans ce cadre.

8.2.1. La question des origines du vivant – Contexte général

Une question fondamentale ancienne. Le problème des origines de la vie est ancien, très ancien ! Il est évidemment relatif à une question, *La Question* probablement la plus fondamentale pour l'humanité, celle qu'on ne peut que mal formuler tant elle est protéiforme, celle qui prend mille formes bien que tout le monde la comprenne : Pourquoi la vie ? Comment ? Pourquoi l'homme ? ... la *Question des Origines*. Ce que je trouve le plus incroyable dans l'affaire, ce qui nous surprend le plus et qui, évidemment, fait que l'on trouve *La Vie* et *La Question des Origines* mystérieuses, les plaçant ainsi en position centrale des humanités, c'est que justement nous soyons conscients de cela et que cela nous fasse réagir.

Cette question des origines du vivant est ainsi, pour moi, fondamentale. C'est pour ça que j'ai choisi de traiter ce sujet en premier dans mon manuscrit. Cette question donc m'apparaît centrale principalement pour deux raisons : d'abord parce qu'évidemment elle se réfère à la question générale des origines (en biologie, en physique, en théologie ...), mais aussi parce que, plus pragmatiquement, on se demande comment, de l'inerte, on est passé à l'organique, de l'organique au vivant et, bien évidemment, cela entre en résonance avec la façon dont nous, des êtres vivants doués de conscience, créons l'outil puis par extension la machine asservie puis autonome et visons à nous diriger tout doucement vers la machine consciente. Dans cette résonance, j'y vois d'une part le fait que nous transposons la façon dont nous concevons ou démontons ces machines, par notre ingénierie, dans notre compréhension du vivant (le vivant-machine), mais aussi le souhait pour l'homme de répondre à une angoisse, l'angoisse de solitude, d'unicité, angoisse qui, je le pense est

la motivation fondamentale de l'humanité à avancer toujours plus. J'invite le lecteur à relire ma partie introductive sur le rare, le faible et l'amorphe, au début de ce document et en particulier sa digression sur ce qu'est la science pour moi.

D'Haeckel à Pasteur - Une génération spontanée de la Vie ? On trouvera toujours une autre manière de faire commencer cette histoire, mais je trouve intéressant de partir d'Haeckel¹⁸. C'est ainsi que je vais commencer, par une histoire biologique des origines du vivant, histoire qui nous amènera au problème qui m'intéresse. *NB* : pour ce récit, je vais beaucoup m'appuyer sur ce livre que j'ai trouvé dans une obscure librairie parisienne aujourd'hui disparue, *Écrits sur la vie* [Ecrits sur la vie : Livre 1964], qui est un recueil de textes anciens, des grecs anciens à nos jours, relatifs à la vie et à ses origines.

Haeckel donc, croit que les mers du globe contiennent une substance vivante primitive qu'il nomme *monère*. Les monères sont, selon lui « *les plus simples des organismes observés, [...] les plus simples des organismes imaginables.* ». A-t-il vu des bactéries et les a-t-il associées à son concept de monères ? Entre la paréidolie, dont même Kepler dont on vante le modèle en astronomie fut victime (puisqu'il a passé une bonne partie de son existence à décrire les habitants de la lune, leur vie, leurs mœurs, à partir de leurs constructions qu'il pensait observer dans sa lunette), et les idées préconçues fortement ancrées y compris en sciences, par exemple la génération spontanée à son époque, Haeckel ne sera pas le premier à voir ce qu'il avait envie de voir et à penser librement sans trop se confronter à la méthode scientifique. Toujours est il que, bien que ses travaux ne reposent que sur peu de bases scientifiques seulement, ce que Haeckel défend c'est tout de même la vie primitive, très simple, et le fait que toute vie fut à l'origine monère. Il dit, je cite : « *Ces organismes, nullement différenciés [...] ressemblant par la simplicité de composition de leur particules aux cristaux inorganiques, [...] ont pu être les primitifs ancêtres de tous les autres organismes.* ». Mieux encore, il soutient que ces mêmes monères, organismes vivants primordiaux, puissent se former à partir « *d'un plasma colloïde* » se formant lui-même à partir de « *composés carbonés plus simples* » (Citation complète en note de bas de page¹⁹).

La base de son raisonnement est fausse ! Haeckel croit, parce que les monères qu'il observe sont « *des organismes sans organes* », « *privés d'organes, constitués par un simple composé chimique et doués pourtant de la faculté de croître, de se nourrir et de se répandre* », ne présentant « *ni*

18 Ernst Haeckel (1834-1919), biologiste et philosophe allemand qui contribue à diffuser la théorie de l'évolution de Darwin.

19 Citation complète 'Écrits sur la vie' p47-48 : « *En effet, puisque chez ces organismes, il n'y a ni organisation, ni différenciation quelconque de parties hétérogènes, puisque, chez eux, tous les phénomènes de la vie sont accomplis par une seule et même matière homogène et amorphe, il ne répugne nullement à l'esprit d'attribuer leur origine à la génération spontanée. S'agit-il de plasmagone ? Y a-t-il déjà un plasma capable de vivre ? alors ce plasma a simplement à s'individualiser, comme le cristal s'individualise dans une solution mère. S'agit-il, au contraire, de la production de monères par véritable autogonie ? alors il est nécessaire que le plasma susceptible de vivre, la substance colloïde primitive, se forme d'abord aux dépens de composés carbonés plus simples.*»

organisation, ni différenciation » et ne pouvant ainsi assurer leur reproduction, que ces êtres qu'il attribue pourtant au vivant, naissent spontanément. Comme beaucoup à cette époque, Haeckel est un défenseur de la génération spontanée !

Ce que je trouve intéressant toutefois, c'est que l'idée de relier une chimie complexe, carbonée, que l'on est « *en mesure de produire artificiellement dans nos laboratoires chimiques* »²⁰ à l'émergence d'un plasma colloïdal, autrement dit une forme primitive d'organisation (plus qu'une soupe diffuse de molécules), menant à une forme de vie primitive, les monères, n'effraie pas Haeckel. Il suppose même que, parce qu'on sait produire ces composés carbonés complexes en laboratoire, alors « *rien n'empêche [...] d'admettre que, dans la libre Nature, des conditions favorables à la formation de ces composés puissent aussi se présenter.* ».

On sait que ce raisonnement ne tient pas dès lors que la vie a colonisé la Terre : les processus de la vie sont infiniment plus efficaces et ainsi plus rapides que toute réaction chimique en milieu naturel. Le vivant occupe toutes les niches depuis longtemps et ne laisse aucune chance à une chimie de type prébiotique de voir le jour de nouveau. La chimie du monde terrestre est complètement contrôlée par la vie. Néanmoins ce raisonnement correspond à celui que nous avons actuellement sur la façon dont nous voyons l'apparition de la vie sur Terre et potentiellement ailleurs dans l'Univers. Pour caricaturer un peu, pas tant que ça en réalité, nous espérons effectivement qu'à partir d'une soupe de molécules simples, la matière s'assemble, se complexifie, se combine, jusqu'à faire émerger des processus chimiques qui, eux-mêmes combinés entre eux, aboutissent à la matière vivante primordiale : les premières cellules vivantes. En réalité, je pense, en lisant Haeckel, qu'il ne croyait pas vraiment que la génération spontanée puisse se faire aujourd'hui, car il pose la question de la possibilité que « *peut-être [...] parmi les monères actuellement connus, y a-t-il une espèce qui, aujourd'hui, continue à naître par génération spontanée* ». Vraisemblablement, il associait davantage son raisonnement à la question de l'apparition de la vie et non à la vie actuelle. Notons au passage que ce qui pose problème quant à la 'génération spontanée' n'est pas tant son principe qu'aujourd'hui on détaillerait dans le cadre de l'émergence, mais la transposition que certains, comme Pouchet²¹, ont faite à la vie actuelle et qui ont créé de vives polémiques que Pasteur s'est attelées à invalider. Haeckel cherche d'ailleurs, comme il le dit, à « *combler la lacune existant entre la cosmogonie de Kant et la théorie de la descendance de Lamarck* », autrement dit à résoudre le passage de l'inerte au vivant. Il conclut : « *L'origine des premières monères par génération spontanée nous semble être un phénomène simple et nécessaire du mode d'évolution des corps terrestres. J'accorde à ce phénomène, tant qu'il n'a pas été directement observé ou reproduit, qu'il est et demeure une simple hypothèse, mais je le répète, cette hypothèse est indispensable à l'enchaînement tout entier de l'histoire de la création.* ». Je pense qu'Haeckel se servait surtout de ce 'principe de génération spontanée' comme d'un cache misère pour pallier cette lacune qui nous

20 NB : il faut rappeler que depuis le milieu du 19ème siècle, jusqu'au milieu du 20ème siècle, la chimie est une activité de pointe. Je pense qu'Haeckel tient un discours influencé par une époque où la chimie est grandement mise en avant. D'autres vraisemblablement aussi ; Oparine (voir note 22) notamment. Il faut bien comprendre que comme l'informatique actuellement, la chimie de l'époque semble être en mesure de résoudre tous les problèmes.

21 Felix Archimède Pouchet (1800-1872), médecin français, farouche défenseur de la génération spontanée et opposant à Pasteur.

gène toujours actuellement, à savoir l'émergence (spontanée) des premiers métabolismes, de la première génétique, *etc.*

Haeckel encore, parle clairement de la nécessité d'isoler la substance vivante dans une membrane : « *Un des premiers actes d'adaptation qu'accomplissent les monères nées par génération spontanée est l'épaississement de leur couche superficielle, qui devient pour la molle substance du centre, une membrane protectrice contre les attaques du monde extérieur.* ». C'est le besoin de microcosme ; le même que celui que Langton a cherché à prendre en compte dans son système d'automate auto-reproducteur (vie artificielle), appelé maintenant *boucles de Langton* [Langton 1984, Rennard 2002]. C'est celui aussi, cher à Oparine²² qui s'attela notamment à obtenir chimiquement ce qu'on aimerait qualifier de protocellules (mais dont les fonctionnalités sont insuffisantes pour mériter ce statut), des coacervats.

Pasteur²³ prend la suite et, avec lui, le problème devient plus chimique et ce pour longtemps puisque les travaux des chimistes, biologistes et biochimistes sur les origines du vivant vont continuer ce mouvement : de Pasteur à Miller, en passant par Oparine, on cherche à savoir comment se forment les molécules du vivant, les sucres, les bases azotées, les acides aminés et leurs polymères. Pasteur, par ses travaux sur la déviation de la lumière polarisée par des énantiomères (molécules non superposables en miroir), ceux d'acide tartrique en particulier, a montré que le vivant ne produit pas de mélanges racémiques de molécules, c'est-à-dire des mélanges à part égales de molécules déviant la lumière dans une direction ou une autre (les énantiomères lévogyres ou dextrogyres). Les processus du vivant, au contraire, produisent un seul énantiomère d'acide aminé (acides aminés L) ou encore un seul énantiomère pour les sucres (les sucres D). Pasteur aimerait pouvoir contrôler ce processus de synthèse, en chimie. Et il s'interroge, notamment sur l'origine de cela : « *quelles sont donc les forces dissymétriques qui président à l'élaboration des principes immédiats naturels ? [...] la dissymétrie, je la vois partout dans l'Univers. L'Univers est dissymétrique. [...] c'est là, suivant moi, un des liens entre la vie à la surface de la Terre et du cosmos, c'est-à-dire l'ensemble des forces répandues dans l'Univers.* ». Il détaille : « *Je pressens même que toutes les espèces vivantes sont primordialement, dans leur structure, dans leurs forces extérieures, des fonctions de la dissymétrie cosmique. [...] La solution consisterait d'une part, dans la découverte de la génération spontanée, si tant est qu'elle soit en notre pouvoir ; d'autre part, dans la formation de produits dissymétriques à l'aide des éléments carbone, hydrogène, azote, soufre, phosphore, si, dans leur mouvements, ces corps simples pouvaient être dominés, au moment de leur combinaisons, par des forces dissymétriques. Voudrais-je tenter des combinaisons dissymétriques de corps simples ? Je ferais réagir ces derniers sous l'influence d'aimants, de solénoïdes, de lumière polarisée elliptique, ... enfin, de tout ce que je pourrais imaginer d'actions dissymétriques.* »

Ce problème pour lequel Pasteur propose des solutions (l'action de facteurs externes pour briser la

22 Alexandre Ivanovitch Oparine (1894-1980), biochimiste russe, pionnier des travaux expérimentaux sur les origines du vivant. Son activité expérimentale concerna principalement la chimiosynthèse de biomonomères.

23 Louis Pasteur (1822-1895), physico-chimiste français, pourfendeur de la théorie de la génération spontanée, c'est un des pères principaux de la microbiologie. Il est aussi connu par les chimistes pour ses travaux précieux qui ont donné naissance à la stéréochimie. Il a aussi mis au point le vaccin contre la rage.

symétrie) est toujours d'actualité. D'une part par les chimistes et biochimistes qui visent à comprendre comment l'énantiométrie s'est mise en place entre la "soupe prébiotique" et la première cellule, mais également à la reproduire. Par exemple, on sait que la silice (SiO_2), minéral omniprésent dans la nature, permet la condensation des acides aminés [Costa 2008], c'est-à-dire la formation de polymères d'acides aminés, les peptides entre autres. On ne sait pas en revanche comment former des peptides énanti-sélectifs ! [Dudjak 2002 ; Sakhmo 2019]. D'autre part, l'énantiométrie étant l'une des signatures du vivant sur Terre, cela en fait un bon critère pour déterminer la vie ou les traces fossiles de vie ailleurs dans le système solaire (Mars, Titan, ou les lunes glacées des géantes gazeuses). L'analyse de la lumière polarisée (et de son spectre) provenant des corps célestes pourrait permettre cela si elle devenait suffisamment sensible ; c'est encore loin d'être possible : l'analyse de la lumière cendrée de la Lune (c'est-à-dire la lumière de la Terre réfléchiée sur la Lune) ne permet pas de mettre en évidence la vie sur Terre, ni par son analyse spectrale ni par la polarisation de sa lumière. En revanche, les instruments embarqués sur les sondes spatiales et les rovers et autres landers à destination respectives des petits corps comme les astéroïdes ou les comètes et des autres planètes ou lunes du système solaire (Mars et Titan), et capables de prélèvements, pourraient exploiter cette signature.

Avec Pasteur surtout naît une première approche de la biochimie moderne. Il reste qu'on ne sait pas à ce moment comment ces molécules peuvent être produites par des processus physico-chimiques abiotiques, et on ne sait pas non plus comment fonctionne la cellule elle-même.

La chimie prébiotique. En 1953 donc, Miller et Urey réalisent une expérience désormais célèbre [Miller 1953, Miller & Urey 1959]. Destinée à mettre en évidence une éventuelle origine chimique de l'apparition de la vie sur Terre, elle consistait à simuler les conditions que ces auteurs ont supposé régner originellement après la formation de la croûte terrestre. Nous verrons que si nos connaissances sur les conditions physico-chimiques de la Terre originelle ont évolué et ne font pas complètement consensus, elles s'affinent actuellement et surtout se complètent avec la connaissance des compositions des sols et atmosphères des autres planètes et lunes du système solaire, des modèles de formation et des observations récentes des disques protoplanétaires, en fonction de la distance de leur étoile, des modèles de formation des planètes, et se diversifient, par rapport à la vision limitée de la flaque sale des origines, avec le fait que l'on considère une plus large gamme de milieux dans lesquels la (les) vie(s) a (ont) pu apparaître : au fond des océans près des fumeurs noirs, près des lacs volcaniques, *etc.* L'expérience de Miller et d'Urey avait pour fondement l'hypothèse (non vérifiée) d'Haldane et d'Oparine : les composés organiques (briques moléculaires simples du vivant) peuvent se former à partir de composés inorganiques (méthane, dihydrogène, eau ...) dans les conditions régnant sur Terre peu après sa formation. L'expérience proprement dite (voir figure 8.1) consistait à soumettre pendant un temps suffisamment long (7 jours dans leur cas) un mélange de gaz (méthane CH_4 , ammoniac NH_3 , hydrogène H_2) et d'eau à des arcs électriques reproduisant les décharges des éclairs dans l'atmosphère terrestre et apportant l'énergie nécessaire à la dissociation (ionisation) des molécules, favorisant ainsi leurs réactions. L'ensemble de cette soupe (dont le nom inspira le concept de *soupe primitive*) subissait en outre des cycles de vaporisation (par

chauffage) et de condensation (par refroidissement). Cette expérience a été maintes fois reproduite.

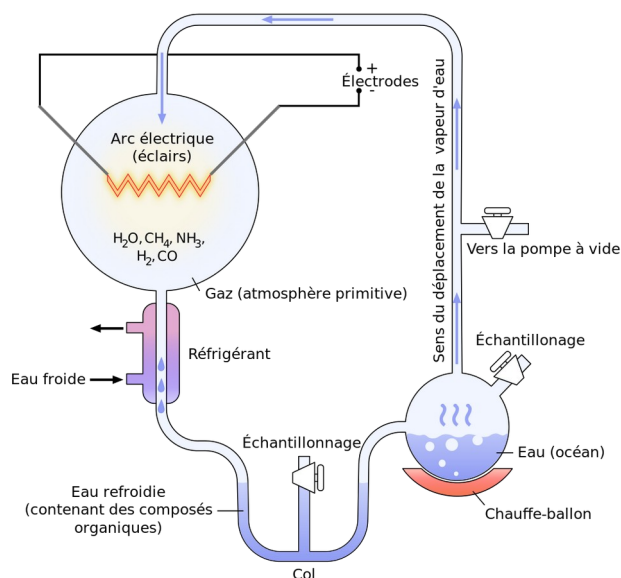


Figure 8.1. Schéma descriptif de l'expérience de Miller et Urey. Il s'agit d'un circuit dans lequel, un fluide (ici de l'eau et des composés inorganiques et organiques) et sa phase vapeur, circulent sous l'action d'un cycle d'évaporation-condensation induit par une source de chaleur et un circuit de refroidissement. Les composés en phase vapeur subissent en outre l'action de décharges électriques, source d'énergie facilitant la catalyse de certaines réactions chimiques. L'échantillonnage correspond aux prélèvements faits pour analyse.

Miller et Urey ont obtenu une solution de molécules organiques contenant de l'urée (CON_2H_4), du formaldéhyde (H_2CO), de l'acide cyanhydrique (HCN), des composés que l'on sait précurseurs de plusieurs autres composés du vivant, l'urée en particulier, mais surtout des bases azotées (qui entrent dans la composition des acides nucléiques) et des acides aminés (constituants des protéines). Miller et Urey ont identifié un certain nombre de composés suffisamment concentrés (certains en quantité supérieure à 2%) dont 7 acides aminés. Les solutions de Miller et Urey ayant été conservées, on a identifié depuis dans ces mêmes échantillons un nombre bien plus important de composés non détectés à l'époque, le nombre d'acides aminés montant ainsi actuellement à 40.

Depuis, l'astrochimie a considérablement contribué à détailler la diversité chimique produite dans le milieu interstellaire, dans les disques proto-planétaires, dans les glaces cométaires également. On montre notamment qu'une part très importante de molécules prébiotiques se forment dans ces milieux et ont pu alimenter le milieu prébiotique naissant sur Terre.

Un regain d'intérêt venant de l'exobiologie. Si des réflexions et recherches concernant les origines du vivant ont toujours eu lieu, elle bénéficient depuis quelques années d'un véritable regain d'intérêt, presque d'un besoin pressant, avec le développement des recherches en astronomie et d'exploration du système solaire.

Le développement des techniques d'observation astronomiques a été en effet spectaculaire depuis

ces 2 dernières décennies et a permis de considérables découvertes, en particulier dans le champ des recherches des planètes extra-solaires. En 1995 était découverte la première exoplanète par la méthode des transits²⁴. Cette exoplanète n'a pourtant rien d'accueillant puisqu'il s'agit d'une planète de type *Jupiter chaud* ; 51 Pegasi b, c'est son nom, fait environ 0.5 masses de Jupiter pour environ 1.56 rayons Joviens et orbite à seulement 0.05 UA²⁵ de son étoile, une naine ultra-froide²⁶, c'est-à-dire une étoile à l'activité extrêmement instable produisant régulièrement d'importantes bouffées de rayonnement ionisants (UV en particulier), tournant autour d'elle avec une période de 4 jours seulement. Cette exoplanète a ainsi une température de surface de l'ordre de 1000°C ce qui dans tous les cas ne permet pas d'envisager la vie. Depuis, entre 1995 et 2020, près de 4000 exoplanètes ont été détectées et, désormais tous les jours, de nouvelles planètes voire lunes de toutes tailles sont identifiées ou entrent dans la liste des candidates à confirmer. En imagerie infrarouge, on observe aussi les disques protoplanétaires creusés de sillons par les planètes naissantes de ces systèmes. Parmi les exoplanètes détectées, la population des planètes de taille comparable à la Terre ou de seulement 2 à 5 masses terrestres explose. On estime maintenant, en se fondant sur les fréquences observées, à 5 à 10 milliards le nombre de planètes de type Terre ou Super-Terre dans une galaxie comme la Voie Lactée [Hsu 2019]. On sait de mieux en mieux établir les bornes des zones habitables en se fondant en particulier sur le type et l'âge des étoiles. De plus, depuis quelques années seulement, la chimie des atmosphères de certaines exoplanètes peut être sondée par spectrométrie lorsque ces planètes passent devant leur étoile. Ajoutons à cela l'exploration du système solaire (Titan et Mars, mais aussi comètes + astéroïdes) qui va croissante et nous avons là un terrain fertile dans lequel la question de l'apparition de la vie revient souvent sur le devant de la scène, est débattue publiquement et surtout dans les milieux scientifiques.

C'est notamment dans ce contexte qu'a pu être lancé le programme transdisciplinaire (CDP) de l'IDEX Grenoblois, *Origine du vivant et des planètes* (programme COOL).

8.2.2. Théories sur l'origine du code génétique

S'il y a bien quelque chose qui saute aux yeux du premier étudiant en biologie venu, c'est évidemment la complexité de la matière vivante. On peut toujours, à grands renforts d'approximations importantes et surtout d'analogies avec les systèmes artificiels (machines, systèmes symboliques, programmes, systèmes fluidiques, compartiments...), décrire dans les grandes lignes un processus cellulaire, une cellule, un organisme (figure 8.2). Cependant, dès que l'on rentre dans les détails, c'est une avalanche qui nous attend. Tout fourmille d'interactions

24 La méthode des transits est l'une des méthodes de détection des exoplanètes. Elle consiste à identifier et mesurer la diminution d'intensité liée au passage d'une planète extra-solaire devant son étoile, et la durée de ce transit.

25 1 UA, pour Unité Astronomique, correspond à la distance moyenne entre la Terre et le Soleil, c'est-à-dire 150 millions de km. Quelques ordres de grandeur : Jupiter se trouve à 5.2 UA du Soleil. La sonde Voyager 2, l'objet le plus lointain construit par l'homme se trouve à 147.6 UA du Soleil à l'heure où j'écris ce texte (en novembre 2019).

26 Une naine ultra-froide est une étoile à la limite entre les naines rouges, objets stellaires de masse comprise entre 0.075 et 0.4 masses solaires (MS) ayant déclenché la réaction de fusion nucléaire, et les naines brunes, des objets substellaires, de masses inférieure à 0,08 MS (13 à 75 MJ) et qui sont sous la limite de fusion thermonucléaire de l'hydrogène.

complexes, multiples, de subtilités qui font que l'écart entre les organisations les plus complexes de la matière non vivante et une simple cellule vivante actuelle, quelle qu'elle soit, est simplement impensable. Lorsqu'on s'intéresse aux origines du vivant, c'est dans ce fossé infranchissable que l'on se jette.

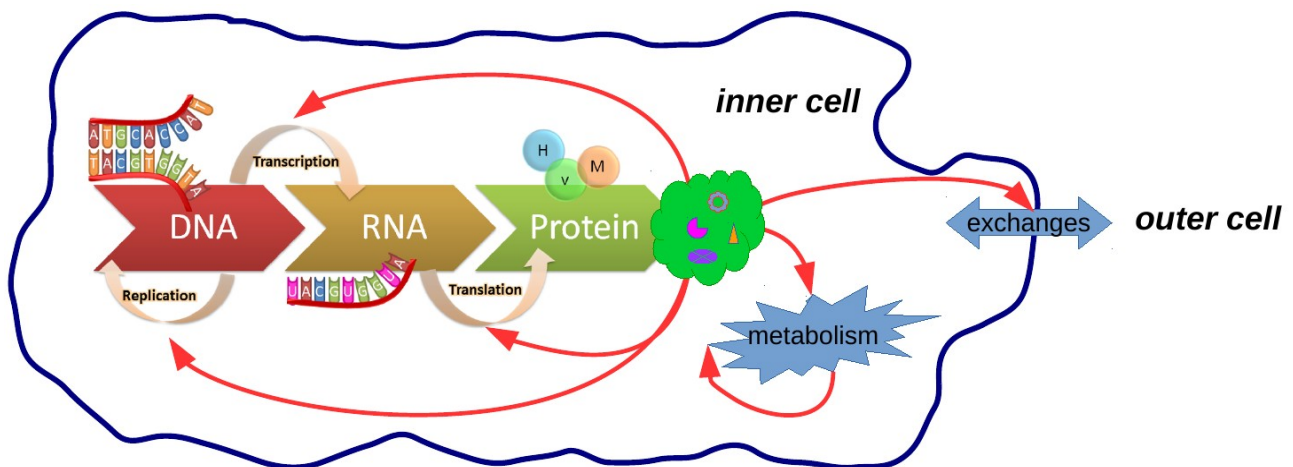


Figure. 8.2. Schéma général du fonctionnement d'une cellule vivante. Une cellule vivante est un système capable de répliquer son patrimoine génétique pour se reproduire, et possède un métabolisme pour maintenir son état vivant en échangeant de la matière avec le milieu extérieur. Ce métabolisme est notamment mis en œuvre par tout un bestiaire de protéines (mais aussi des ARN et de nombreuses petites molécules), produites par traductions des ARNs transcrits.

Un des éléments clef du vivant est bien sûr le code génétique et toute la machinerie qui permet son exploitation (transcription en ARN puis, éventuellement traduction en protéines) et la répllication du génome permettant la reproduction cellulaire. L'ensemble des éléments qui constituent cette machinerie et leur 'harmonieux ballet' est lui-même extrêmement complexe (figure 8.3). Le problème²⁷, c'est que tous ces éléments, le système de traduction en particulier, semblent absolument nécessaires à la bonne exploitation/répllication du génome. L'apparition d'un code génétique et son évolution semblent intimement liés à l'apparition et l'évolution du système de traduction lui-même. Comment dès lors imaginer initier une génétique précurseur de la vie sur Terre ? On se retrouve en effet avec un des problèmes les plus difficiles de la biologie, rien de moins que *Le Problème de Qui de l'œuf ou de la poule ... ?*

27 « [...] the problem appears extremely and unusually hard. »[Koonin 2017 a]

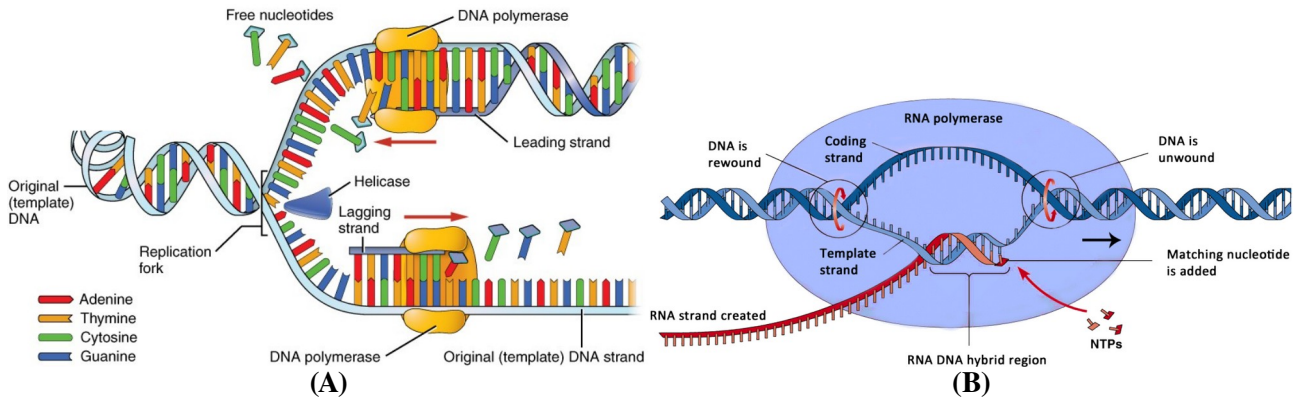


Figure 8.3. (A) *Processus de réplication de l'ADN* : une fourche est ouverte dans la double hélice par une hélicase et permet aux enzymes ADN polymérasés de faire croître deux brins complémentaires à partir de désoxyribonucléotides libres. (B) *Transcription de l'ADN en ARN* : la double hélice est dé-torsadée par des gyrases; une ARN polymérase et plusieurs facteurs et protéines associées se fixent sur le début d'une région codante (gène), le promoteur ; un brin d'ARN est alors synthétisé par cette enzyme qui lit le brin codant d'ADN et polymérise des ribonucléotides libres. (C) *Traduction de l'ARN en protéines* : Le brin d'ARN codant (ARN messager : ARNm) est pris en charge par un lourd complexe riboprotéique, le ribosome ; celui-ci favorise l'appariement d'ARN de transfert (ARNt) portant des acides aminés avec la chaîne codante (succession de codons), et leur rapprochement permettant la formation d'une liaison peptidique ; les chaînes polypeptidiques sont ainsi polymérisées ; les protéines sont leur produit de maturation.

Il y a actuellement 3 théories majeures proposant une possible explication à l'apparition du code génétique. Koonin en donne une description et un regard critique très détaillés dans ses excellentes revues de 2009 et 2017 [Koonin 2009 & 2017 a]. J'incite celui qui souhaite entrer plus en détail dans cette problématique de l'origine du code génétique et disposer de plus de références (que je donne pas ici) à lire ces deux revues dans l'ordre de publication.

Ces trois théories²⁸ sont nées très tôt après la découverte du code génétique au début des années 1960 [Nirenberg, 1961 ; Matthaei, 1962 a et b ; Leder, 1964] :

- la **théorie de l'origine stéréochimique**
- la **théorie adaptative** (d'optimisation ; de minimisation d'erreur).
- la **théorie de la co-évolution**

1ère base Codon (5')	2ème base Codon				3ème base Codon (3')
U	C	A	G		

28 Koonin, 2017 (a), souligne à juste titre que le terme *théorie* fréquemment utiliser pour parler de ces hypothèses et cadres conceptuels est un peu abusif du point de vue épistémologique.

U	5'UUU3'-- 3'AAA5'	Phe	5'UCU3'-- 3'AGA5'	Ser	5'UAU3'-- 3'AUA5'	Tyr	5'UGU3'-- 3'ACA5'	Cys	U
	5'UUC3'-- 3'AAG5'	Phe	5'UCC3'-- 3'AGG5'	Ser	5'UAC3'-- 3'AUG5'	Tyr	5'UGC3'-- 3'ACG5'	Cys	C
	5'UUA3'-- 3'AAU5'	Leu	5'UCA3'-- 3'AGU5'	Ser	5'UAA3'-- 3'AUU5'	Stop	5'UGA3'-- 3'ACU5'	Stop / Trp	A
	5'UUG3'-- 3'AAC5'	Leu / init	5'UCG3'-- 3'AGC5'	Ser	5'UAG3'-- 3'AUC5'	Stop	5'UGG3'-- 3'ACC5'	Trp	G
C	5'CUU3'-- 3'GAA5'	Leu	5'CCU3'-- 3'GGA5'	Pro	5'CAU3'-- 3'GUA5'	His	5'CGU3'-- 3'GCA5'	Arg	U
	5'CUC3'-- 3'GAG5'	Leu	5'CCC3'-- 3'GGG5'	Pro	5'CAC3'-- 3'GUG5'	His	5'CGC3'-- 3'GCG5'	Arg	C
	5'CUA3'-- 3'GAU5'	Leu	5'CCA3'-- 3'GGU5'	Pro	5'CAA3'-- 3'GUU5'	Gln	5'CGA3'-- 3'GCU5'	Arg	A
	5'CUG3'-- 3'GAC5'	Leu / init	5'CCG3'-- 3'GGC5'	Pro	5'CAG3'-- 3'GUC5'	Gln	5'CGG3'-- 3'GCC5'	Arg	G
A	5'AUU3'-- 3'UAA5'	Ile	5'ACU3'-- 3'UGA5'	Thr	5'AAU3'-- 3'UUA5'	Asn	5'AGU3'-- 3'UCA5'	Ser	U
	5'AUC3'-- 3'UAG5'	Ile	5'ACC3'-- 3'UGG5'	Thr	5'AAC3'-- 3'UUG5'	Asn	5'AGC3'-- 3'UCG5'	Ser	C
	5'AUA3'-- 3'UAU5'	Ile	5'ACA3'-- 3'UGU5'	Thr	5'AAA3'-- 3'UUU5'	Lys	5'AGA3'-- 3'UCU5'	Arg	A
	5'AUG3'-- 3'UAC5'	Met / init	5'ACG3'-- 3'UGC5'	Thr	5'AAG3'-- 3'UUC5'	Lys	5'AGG3'-- 3'UCC5'	Arg	G
G	5'GUU3'-- 3'CAA5'	Val	5'GCU3'-- 3'CGA5'	Ala	5'GAU3'-- 3'CUA5'	Asp	5'GGU3'-- 3'CCA5'	Gly	U
	5'GUC3'-- 3'CAG5'	Val	5'GCC3'-- 3'CGG5'	Ala	5'GAC3'-- 3'CUG5'	Asp	5'GGC3'-- 3'CCG5'	Gly	C
	5'GUA3'-- 3'CAU5'	Val	5'GCA3'-- 3'CGU5'	Ala	5'GAA3'-- 3'CUU5'	Glu	5'GGA3'-- 3'CCU5'	Gly	A
	5'GUG3'-- 3'CAC5'	Val	5'GCG3'-- 3'CGC5'	Ala	5'GAG3'-- 3'CUC5'	Glu	5'GGG3'-- 3'CCC5'	Gly	G

Figure 8.4. Code génétique. L'association codons - acides aminés se lit en choisissant dans l'ordre (1,2,3) correspondant au sens de lecture de l'ARN (de l'extrémité 5' vers l'extrémité 3'), chacune des bases du codon, pour déterminer la correspondance avec un acide aminé. Dans le tableau sont également indiqués de manière surlignée les anticodons (sens 3' → 5'), c'est-à-dire les bases portées par l'ARN de transfert portant l'acide aminé correspondant dans le code génétique. Enfin, un code couleur est ajouté pour indiquer le type de polarité et la charge des acides aminés : en jaune les acides aminés apolaires, en vert les acides aminés polaires, en bleu les acides aminés basiques et en rouge les acides aminés acides. Les codons UAA, UAG et UGA sont des signaux de terminaison (ils indiquent la fin de la traduction). [Graphique modifié à partir de https://fr.wikipedia.org/wiki/Code_g%C3%A9n%C3%A9tique]

Avant de les présenter, disons le tout de suite : aucune ne domine ! Aucune ne fait consensus et surtout il est possible d'opposer des contre-arguments parfois forts à chacune d'entre elles. Mais elles sont aussi non-exclusives : elles peuvent avantageusement être considérées combinées, permettant d'entrevoir des scénarios plus plausibles pour l'origine du code génétique.

Rappelons rapidement ce qu'est le code génétique et ses propriétés. Le code génétique associe chacun des 64 triplets de nucléotides (nommés *codons*) de quatre types possibles, l'Adénosine (A), la Thymidine (T), la Cytosine (C) et la Guanosine (G), à un des 20 acides aminés²⁹. Ce tableau des associations est représenté figure 8.4. Parmi ses propriétés, la première, frappante, est une tolérance à l'erreur importante (bien que non optimale) : à cause d'une redondance évidente de codons pour certains acides aminés (il y a 64 codons et 20 acides aminés ; plusieurs codons peuvent donc coder pour un même acide aminé ; on dit que le code est dégénéré mais non ambigu), une erreur sur une base, en particulier la 3ème, a peu d'effet ; dans ce type d'erreur (un ARN de transfert inadéquat se présentant devant un codon), soit ce sera le même acide aminé qui sera présenté, soit un acide aminé d'un type physico-chimique proche (polaire, apolaire ..., longue chaîne, ...).

Peu après la découverte du code génétique (complet en 1966), Francis Crick proposa que cette

²⁹ NB : il existe d'autres types de nucléotides plus rares, ainsi que d'autres acides aminés, mais on ne parlera que des 20 acides aminés classiques. Globalement, le code génétique associe les triplets A, T, G, C aux 20 acides aminés courants.

configuration puisse être principalement le résultat d'un *accident figé* (une configuration aléatoire donc) qui s'est maintenu dans les systèmes vivants par la suite [Crick 1968 ; Ellington, 2000]. C'est ce que l'on nomme théorie du « *Frozen accident* » (que je traduis par *accident figé*). La structure clairement non-aléatoire, correctrice d'erreurs, du code génétique, ne plaide *a priori* pas particulièrement en faveur de ce scénario, pris tel quel. Il faut, je pense, accoler aux trois théories proposées ci-dessus, la **théorie de l'adaptateur**, proposée dès 1955 par F. Crick [Crick 1955], ajoutant un mécanisme sous-jacent à l'accident figé : ce sont des molécules adaptatrices, des ARN de transfert primordiaux, dès l'apparition d'une forme de génétique, qui permettent l'interface entre le monde à ADN et le monde à protéine [Hopfield 1978].

Pour expliquer d'une part la formation d'un système génétique (molécule codante, système de réplication, système de transcription/traduction en protéines) et la structure particulière du code génétique actuel (et de ses quelques légères variations chez certaines espèces), trois théories toujours en vigueur, ont donc été proposées. Je commencerai par brièvement décrire les deux dernières, la théorie adaptative et la théorie de la coévolution, qui s'appuient sur la théorie de base proposée par Crick, l'accident figé, en cela qu'elles décrivent comment s'éloigner d'une configuration initiale aléatoire à la configuration actuelle. Je discuterai ensuite de la théorie stéréochimique qui, elle, propose un scénario alternatif à l'accident figé. Enfin, je présenterai le modèle récent proposé par Koonin [Koonin 2017 a] alliant plusieurs de ces théories dans un scénario plus raisonnable que chacune d'entre elles séparément. Son modèle est fondé sur la théorie de l'adaptateur primordial.

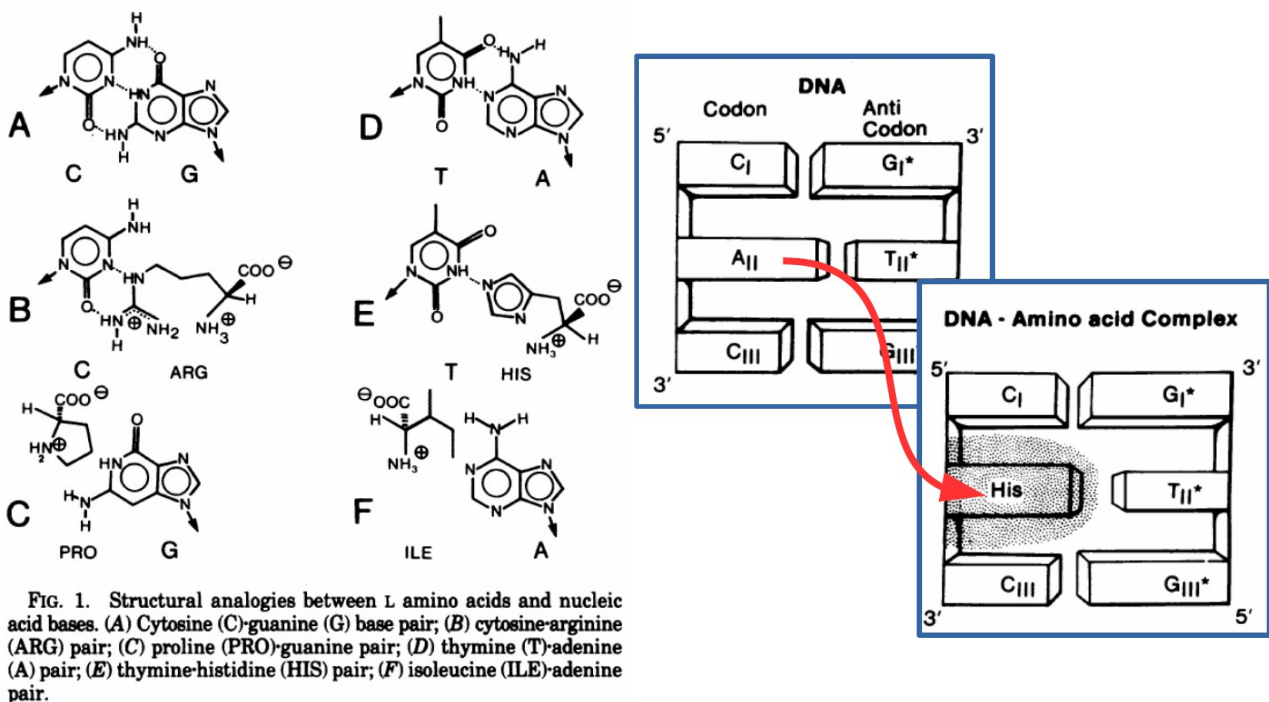
La **théorie de la coévolution** [Di Giulio 2004, Wong 2005 & 2007] postule que la diversité en acides aminés a augmenté au fur et à mesure que les métabolismes se sont perfectionnés, et que l'allocation des codons aux amino-acides pourrait avoir été guidée par des voies métaboliques. Les voies métaboliques doivent être vues comme des forces influençant la mise en correspondance des codons et des acides aminés. Le mécanisme de traduction n'est toutefois pas explicite dans le cadre de cette théorie.

La **théorie adaptative** postule que le code génétique s'est structuré sous l'action de forces sélectives qui tendaient à rendre le code le plus robuste possible, c'est-à-dire ayant la propriété de minimiser les erreurs de structure et de fonction des protéines synthétisées en cas d'erreur de lecture ou de mutation. Pour corroborer cette théorie, il s'agit de chiffrer la robustesse relative du code génétique standard face à un autre code, potentiellement aléatoire ou au contraire calculé pour être le plus optimal possible, parmi les 10^{84} codes possibles [Haig 1991, Gilis 2001, Freeland 2000 & 2003, Goodarzi 2004]. Travailler avec ce nombre astronomique de codes est impossible ; des méthodes de Monte-Carlo ont donc été employées. Dans le cadre de cette théorie, les efforts ont beaucoup porté sur la recherche d'une fonction de coût permettant de calculer le degré d'optimisation. La métrique la plus couramment utilisée est fondée sur la nécessité du respect de la polarité (*Polar Requirement scale*) des acides aminés pour le bon fonctionnement de la protéine analogue. D'autres métriques plus précises prenant en compte la non-uniformité de la position des codons et les biais de transitions dépendants des bases considérées, ont permis d'établir la matrice de substitution des

acides aminés couramment utilisée [Freeland 1998]. D'autres raffinements ont été apportés à cette théorie, mais les conclusions évoluent peu. Ainsi, il a été estimé que les chances d'obtenir le code génétique standard parmi les possibles sont d'environ 1 sur 1 million [Freeland 1998]. Il a été toutefois montré que le code génétique standard est loin d'être le plus optimal ; on estime que sa robustesse se situe à mi-chemin entre un code aléatoire non optimal et le code le plus optimal possible [Novozhilov 2007].

Encore une fois, cette théorie ne fournit pas d'explication quant au mécanisme de traduction nécessaire.

Faisons à ce stade une remarque : si le code génétique actuel avait correspondu au code le plus optimal possible, cela aurait signifié que les forces influençant le choix d'un code dans l'espace de tous les codes génétiques possibles (fitness landscape) auraient été si fortes qu'elles auraient été en mesure de vaincre toutes les barrières de potentiel. Ce n'est pas le cas ; pourtant on peut montrer que le code génétique actuel ne correspond pas particulièrement à un maximum (local) d'optimisation. Tout processus d'optimisation fort provoquant une modification du code génétique aurait du au moins amener le code génétique à un maximum local d'optimisation pour la robustesse. Dans le cadre d'une théorie d'optimisation, on peut supposer que des mécanismes ont freiné ces déplacements dans le paysage des codes génétiques, par exemple un mécanisme de verrouillage du processus d'optimisation du code par transferts horizontaux, comme cela est soutenu par les travaux théoriques de Vetsigian et al [Vestigian 2006]. Ou alors on peut considérer que les mécanismes d'optimisation ne sont pas à l'oeuvre et que ce sont des processus fondés sur la biochimie (théorie de la coévolution ou théorie stéréochimique) qui ont davantage forcé l'obtention d'un tel code.



(A)

(B)

Figure 8.5. Théorie stéréochimique - Substitution de bases par des acides aminés. Ces deux figures sont issues de

l'article [Hendry 1981]. Elles illustrent (A) l'analogie entre des résidus d'acides aminés comme l'arginine, l'histidine, la proline et l'isoleucine, et les bases azotées puriques (G et A) et pyrimidiques (T et C). Les auteurs proposent (B) que l'on substitue la 2^{ème} base azotée du codon par l'acide aminé correspondant dans le code génétique. Partant de cette substitution, ils estiment le degré de correspondance (stéréochimique) entre le codon modifié et l'anticodon. Le détail chimique de la substitution est montré figure 8.6. La carte des adéquations est donnée figure 8.7.

Je vais maintenant détailler davantage la **théorie stéréochimique** puisque c'est celle-ci sur laquelle mon étude va principalement porter.

Cette théorie est étroitement liée à la découverte de la structure en double hélice de l'ADN. Dès l'année qui a suivi, G. Gamow proposait un modèle (le code diamant) liant cette structure aux séquences d'acides aminés des polypeptides : selon lui, les acides aminés devaient se confiner dans les poches formées par 2 paires de nucléotides en vis-à-vis (voir l'introduction, chapitre 3.1.1 et notamment la figure 3.1), permettant le rapprochement nécessaire à leur polymérisation.

Le processus, bien que faux, avait fait toutefois germer chez plusieurs personnes l'idée d'une interaction directe entre acides aminés et chaînes nucléotidiques. Cette idée, simple mais assez attrayante car explicative à la fois de la sélection d'un code génétique particulier et de la co-émergence possible entre chaînes peptidiques et chaînes nucléiques, fut plus clairement définie par Pelc, Welton et Dunnill [Pelc 1965, Pelc & Welton 1966, Dunnill 1966] ; elle est la suivante : les acides aminés présentent une affinité physico-chimique plus importante pour leurs anticodons respectifs dans le code génétique que pour les autres. L'interaction se ferait principalement grâce à la 2^{ème} base, nommée base discriminante (cette base étant celle qui influence le plus la détermination du type physico-chimique des acides aminés). La raison d'une interaction se faisant plutôt avec l'anticodon qu'avec le codon est soutenue par la remarque faite par Johnson et Wang [Johnson 2010] que le ribosome, structure accueillant les résidus d'acides aminés lors du processus de traduction dans des poches formées à majorité d'anticodons, est la plus ancienne structure moléculaire commune à toutes les cellules et que cette structure est centrale au processus de traduction. Vu ainsi, le ribosome serait le résultat de l'évolution de structures plus anciennes, associant acides aminés et acides nucléiques par l'intermédiaire d'une affinité directe acide aminé – anticodon. De plus, Hopfield fait remarquer que dans l'ARNt d'*Escherichia coli*, le nombre de bases de l'ARNt provoque le positionnement du triplet de l'anti-codon très près de l'extrémité 3', couvrant ainsi le site de fixation pour les amino-acides [Hopfield, 1978](figure 8.6). Vu ainsi, le triplet anti-codon et l'acide aminé porté par l'ARNt semblent jouer un rôle de maintien mutuel de la structure. Cela renforce la proposition d'hypothèse stéréochimique du code génétique.

Cette théorie est la seule véritablement testable, expérimentalement du moins, mais aussi par des approches de simulation.

Hendry *et al* [Hendry 1979, Hendry 1981 a et b] montrent notamment, par simulations numériques, que la chaîne latérale de nombreux acides aminés s'apparente, du point de vue stéréochimique et des propriétés physico-chimiques, à la structure purique ou pyrimidique des bases azotées portées par les nucléotides. En particulier, ils proposent que si l'on remplace une base azotée (la deuxième) par

la chaîne latérale d'un acide aminé, avec un positionnement du carbone alpha de l'acide aminé au niveau de l'azote N9 d'une purine ou de l'azote N1 d'une pyrimidine, la chaîne latérale peut se trouver dans une conformation qui mime la base azotée ainsi substituée (voir figures 8.5 et 8.6). Ils réalisent numériquement les substitutions des bases azotées centrales (2ème base) des codons, dans tous les codons possibles, par tous les acides aminés parmi les 20 classiques, testent les 1280 combinaisons possibles et mesurent l'affinité polaire entre codons substitués et anticodons. Ils obtiennent une matrice des affinités (figure 8.7) qui correspond de façon surprenante à la correspondance acides aminés – anticodons.

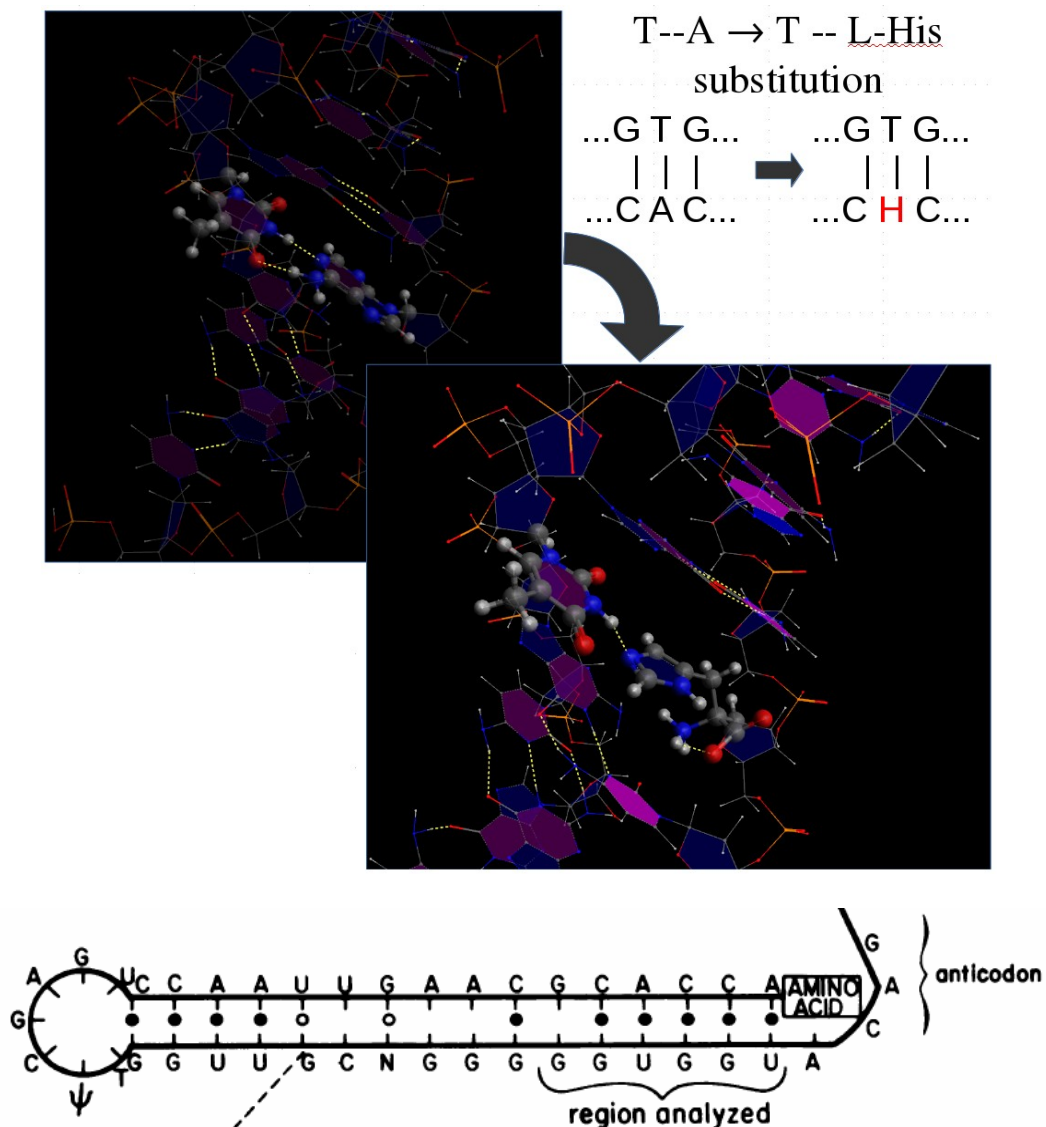


Figure 8.6. Théorie stéréochimique – affinités acides aminés – anti-codons

(Haut) Substitution d'une base azotée par un acide aminé. On substitue ici la base azotée de l'Adénosine (donc l'Adénine) par une Histidine dans un codon CAC. En haut à gauche, on voit en modèle de boules et bâtonnets les bases azotées de la Thymidine (donc la Thymine), à gauche, et de l'Adénine, à droite, chacune reliées à leur brin d'ADN respectif. Les liaisons hydrogène sont en pointillés jaune. En bas à droite, l'Adénine a été remplacée par l'acide aminé Histidine. La substitution se fait en branchant le carbone alpha de l'acide aminé sur le carbone 1' du ribose qui portait

l'Adénine. Cette modélisation a été faite sous le logiciel Avogadro. (Bas) ARNt d'*E. coli*, extrait de la figure 1 de l'article [Hopfield, 1978]. L'anti-codon couvre l'acide aminé porté par l'extrémité 3' de l'ARNt.

Certaines mesures expérimentales corroborent également partiellement cette théorie [Saxinger 1971], à commencer par Woese et al [Woese 1966] dont les mesures fondées sur la solubilité différentielle d'acides aminés pour des pyridines et ses analogues font entrevoir la possibilité d'interactions polaires directes entre bases nucléiques et acides aminés. La plupart des expériences ont cependant porté sur l'affinité entre des acides aminés et des pelotes de chaînes nucléotidiques contenant plusieurs fois les séquences codon et anticodon, molécules nommées *aptamères* et proposées comme adaptateurs initiaux entre chaînes codantes et acides aminés [Knight 1998 ; Yarus 1998, 2000, 2005 & 2009]. Ces mesures montrent d'une part que des séquences courtes et aléatoires d'ARN sélectionnées par fixation d'acides aminés présentent un enrichissement non négligeable en codons ou anticodons spécifiques des acides aminés correspondant dans le code et ayant servi à la sélection de ces séquences (séquences ainsi désignée comme *aptamères*). De plus, l'analyse de 337 sites de fixation indépendants à montré une affinité préférentielle entre cinq (Arg, His, Ile, Phe, Trp) parmi huit (Arg, Gln, His, Ile, Leu, Phe, Trp, Tyr) acides aminés et leurs triplets nucléotidiques correspondants.

La théorie stéréochimique souffre cependant de nombreux contre-arguments opposables. Koonin conclut dans sa revue de 2009 que les interactions mesurées entre acides aminés et triplets de nucléotides associés sont dans tous les cas faibles. A cet argument, il est possible de répondre que cela peut suffire comme biais initial dans un modèle où des peptides, au lieu de se former aléatoirement, se formeraient plus spécifiquement selon la séquence indiquée par des successions de triplets de nucléotides portés par une chaîne informative nucléique. Mais le principal contre argument à cette théorie et aux expériences d'affinité réalisées est cependant que les acides aminés employés pour ces expériences d'affinité entre acides aminés et *aptamères* sont des acides aminés qui sont apparus tardivement !

En effet, il y a désormais un consensus sur le fait que la composition en acides aminés disponibles aux origines du vivant était considérablement réduite par rapport aux 20 acides aminés (qu'on désignera comme acides aminés canoniques) : cela vient d'une part du constat que seuls 10 acides aminés sont systématiquement produits avec une abondance raisonnable dans les expériences de chimie prébiotique ; ce sont ces mêmes acides aminés qui sont identifiés dans les météorites. Il s'agit de Gly, Ala, Asp, Glu, Val, Ser, Ile, Leu, Pro et Thr, dans l'ordre de leur niveaux d'abondance [Pizzarello 2006, Zaia 2008, Cleaves 2010, Burton 2012]. Leur niveau d'abondance dans les météorites est inversement corrélé avec l'énergie libre nécessaire à leur production [Higgs 2009]. De plus, Jordan et al [Jordan 2005] ont montré qu'au cours de l'évolution récente de divers organismes, la proportion de ces acides aminés précoces diminue tandis que celle des acides aminés plus récents augmente. Cela va dans le sens d'un remplacement progressif au fur et à mesure que les acides aminés nouveaux ont été incorporés dans le code génétique.

De ce consensus, on peut dès lors raisonnablement supposer que le code génétique, ne codant que pour 10 acides aminés, n'était pas aussi étendu qu'aujourd'hui ; pour coder 10 acides aminés, 16

codons de 2 lettres suffisent. Il n'est pas raisonnable de penser que les molécules informatives (ADN et ARN) portant les codons soient passées d'un code à 2 lettres à un code à 3 lettres ; cela nécessiterait une très complexe modification de l'ensemble des séquences codantes pour insérer entre chaque codon une nouvelle lettre. Cependant, il est imaginable de penser qu'une des lettres du code (un nucléotide du codon) n'était pas utilisée, n'était pas indispensable ou avait un rôle de stabilisation du système de lecture/traduction ; dans tous les cas, on peut la penser non-informative.

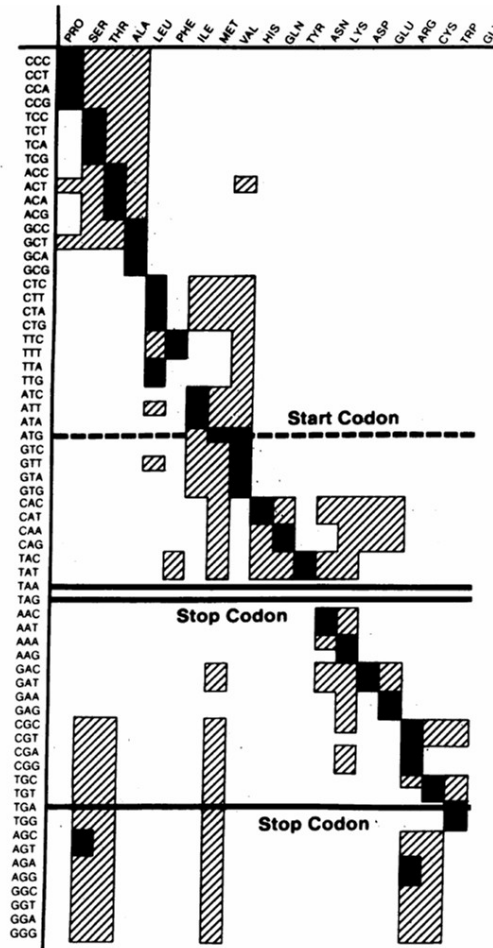


FIG. 5. Preliminary stereochemical fits of L amino acids into the 64 possible cavities formed by removal of the middle base in DNA as in Figs. 2-4. Solid areas indicate that the amino acid fits into cavities formed from its codons (5'→3'); hatching indicates alternate fits of amino acids into cavities not directly associated with their codons. Open areas indicate that the amino acid does not fit into that cavity.

Figure 8.7. Théorie stéréochimique – Carte des adéquations stéréochimiques Anticodons-Acides aminés. Cette figure issue de l'article [Hendry 1981] montre l'adéquation estimée entre les codons dont la 2ème base est substituée et les anticodons. Les cases noires indiquent une forte adéquation. Les cases grises, une adéquation partielle.

La troisième d'un codon est le candidat idéal. En effet, elle joue déjà plus ou moins ce rôle dans le code génétique actuel. Cette troisième base nucléotidique, lorsqu'elle est mal transcrite/traduite, ne génère que pas ou peu d'erreur dans la protéine synthétisée. De plus, cette base est à l'origine du phénomène de *wobble* : un même ARN de transfert peut reconnaître des codons différents codant pour le même acide aminé ; le degré de liberté se situe au niveau de la 1ère base de l'anticodon (porté par l'ARNt) et permet un appariement bancal (*wobble*) avec la 3ème base du codon. Cela

considéré, il est très raisonnable de penser que le code génétique ne nécessitait initialement que 2 lettres pour 10 acides aminés codés (figure 8.8), puis que ce code a subi une expansion progressive avec l'incorporation de nouveaux acides aminés, nouvellement accessibles par les voies métaboliques de complexité croissante [Novozhilov 2009]. Il est à noter que, même s'il est réduit, ce code présente déjà un niveau de robustesse significatif et surtout met d'emblée en place le futur schéma de polarité des acides aminés dans le code génétique actuel. Cette expansion a dû se faire sous l'influence de facteurs soit liés au métabolisme (théorie de la coévolution), soit sélectif et reposant donc sur des critères d'optimisation (théorie adaptative). Rien n'exclut toutefois que les affinités directes trinuécléotides – acides aminés (théorie stéréochimique) soient à l'œuvre pour l'inclusion de nouveaux acides aminés dans le code génétique. En effet, si l'association acides aminés – codons dans le code génétique à 2 lettres était déjà le résultat de, ou du moins dépendait de l'affinité directe entre polynucléotides et acides aminés, cette même association dans le code à 3 lettres n'était pas contradictoire car les nouveaux acides aminés inclus dans le code génétique moderne sont proches dans leurs caractéristiques de polarité des acides aminés du code ancestral à 2 lettres.

1ère base Codon (5')	2ème base Codon							
	U		C		A		G	
U	5'UU*3'-- 3'AA*5'	Leu/ ???	5'UC*3'-- 3'AG*5'	Ser	5'UA*3'-- 3'AU*5'	???	5'UG*3'-- 3'AC*5'	???
C	5'CU*3'-- 3'GA*5'	Leu	5'CC*3'-- 3'GG*5'	Pro	5'CA*3'-- 3'GU*5'	???	5'CG*3'-- 3'GC*5'	???
A	5'AU*3'-- 3'UA*5'	Ile	5'AC*3'-- 3'UG*5'	Thr	5'AA*3'-- 3'UU*5'	???	5'AG*3'-- 3'UC*5'	Ser/ ???
G	5'GU*3'-- 3'CA*5'	Val	5'GC*3'-- 3'CG*5'	Ala	5'GA*3'-- 3'CU*5'	Asp	5'GG*3'-- 3'CC*5'	Gly

Figure 8.8. Code génétique ancien. Ce code génétique à 2 lettres, associe les 2 premiers nucléotides d'un codon à leurs acides aminés respectifs. Ce code supposé ancestral ne met en correspondance que 9 acides aminés sur les 10 supposés ancestraux. Les symboles '*' indiquent une troisième base libre. Les '???' sont là pour signifier l'incertitude quant à la correspondance codons - acides aminés. Le code couleur est le même que celui de la figure 8.4. On remarque que la distribution du caractère polaire des acides aminés est globalement inchangé par rapport au code génétique actuel.

[Graphique modifié à partir de la figure 2 de l'article [Koonin 2009]]

Cela amène Koonin et Novozhilov [Koonin 2017 a] à proposer un modèle intéressant, prenant un peu de chacune des théories et restant acceptable du point de vue de la complexité des éléments (molécules) à générer. Koonin propose que le code originel, à 2 lettres, est le résultat d'un accident figé comme le proposait Crick. Ce point est bien entendu discutable ; des arguments provenant de la théorie stéréochimique pourraient en effet soutenir l'idée que cette association initiale puisse être le résultat d'affinités préférentielles entre polynucléotides (en l'occurrence 2 nucléotides pour un code à 2 lettres ; à noter que la 2ème base, la plus importante dans la théorie stéréochimique comme de façon générale dans le code génétique, est maintenue entre les codes à 2 et à 3 lettres) et acides aminés primordiaux. Les expériences que je mène ont notamment pour objectif de tester ce point. Koonin [Koonin 2017 a et b] propose que des adaptateurs en ARN, ancêtres des ARN de transfert, se présentant sous la forme de chaînes polynucléotidiques assez courtes et présentant une affinité spécifique, en une région particulière de leur séquence (un anticodon donc), pour la séquence

codante (le codon) puissent reconnaître spécifiquement dans une autre région de leur séquence, dans une poche de nucléotides, des acides aminés. Ces adaptateurs auraient ensuite évolué, permettant l'intégration d'acides aminés proches dans leurs propriétés et permis leur intégration progressive dans le code génétique (voir figure 8.9).

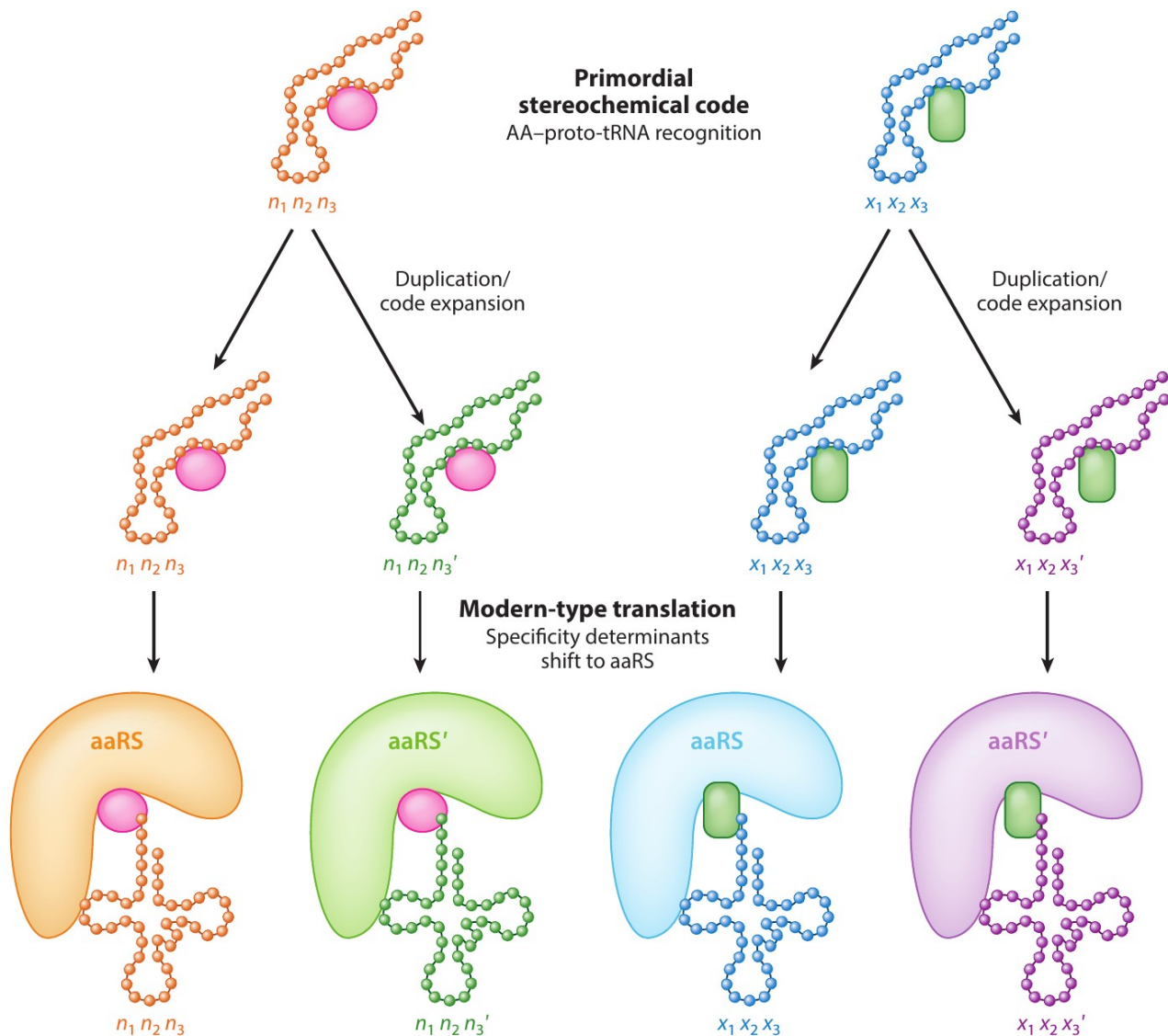


Figure 8.9. Evolution des adaptateurs de traduction. Cette figure est tirée de [Koonin, 2017 (a)]. Des séquences particulières, en nombre restreint au départ et capables de reconnaître un acide aminé (ou plusieurs acides aminés de même type) par reconnaissance stéréochimique (non corrélée avec le code génétique) se diversifient et permettent une expansion du code génétique car reconnaissant plusieurs triplets différents pour un même acide aminé (ou famille) reconnu. Ces proto-adaptateurs évoluent ensuite peu à peu sous la forme des couples ARN de transfert – aminoacyl synthases.

De deux choses l'une : (i) soit la reconnaissance directe entre anticodons (ou codons) et acides aminés ne peut pas se faire, ou en tout cas n'obéit pas aux mêmes correspondances que dans le code génétique, et dans ce cas, seule la présence d'adaptateurs entre la séquence codante (codons) et les acides aminés permet cette correspondance ; (ii) soit cette reconnaissance directe entre anticodons

(ou codons) et acides aminés est valide (théorie stéréochimique du code génétique valide) et dans ce cas, on peut penser le modèle de Koonin comme une 2ème étape dans l'évolution du processus de traduction au cours de laquelle, des molécules adaptatrices faites d'ARN auraient progressivement été insérées entre codons et acides aminés.

Dans le premier cas (i), une préférence d'affinité entre des polynucléotides (pas nécessairement les codons ou anticodons) et des acides aminés est quand même nécessaire. Cela ne correspond pas à la théorie stéréochimique du code génétique mais à une de ses variantes : la région de reconnaissance des ARNt primordiaux doit effectivement présenter une affinité importante entre plusieurs nucléotides et l'acide aminé qu'ils portent. Dans ce cas, les expériences de mesure d'affinité entre polynucléotides et acides aminés que nous menons devraient appuyer ce modèle. Notons que dans ce cas, les associations trouvées pour les acides aminés testés pourraient être celles déterminées par Yarus et al [Yarus 2009] qui a recherché les séquences d'appariement dans de nombreux polynucléotides pour divers acides aminés (sélection d'ARN aléatoires par affinité pour des acides aminés).

Dans le deuxième cas, faisons une remarque importante : Koonin [Koonin 2009, 2017 a et b], comme beaucoup, pensent le processus de traduction, y compris ancestral, comme nécessairement efficace [Hopfield 1978] et surtout à haute fidélité (peu d'erreurs). Pour argumenter ce point, Koonin s'appuie par exemple sur l'évolution apparemment précoce des aminoacyl ARNt Synthases, enzymes qui permettent de lier un acide aminé spécifique à son ARNt [Artymiuk 1994, Anantharaman 2002, Aravind 2002 a et b]. Remarquons toutefois qu'on parle là d'un système finalement déjà évolué qui a vraisemblablement franchi le passage du non vivant au vivant (même pré-LUCA). D'une part, je pense qu'il faut toutefois résoudre ce problème en se posant la question de comment faire dialoguer le monde des acides nucléiques et le monde des peptides, depuis l'ère prébiotique jusqu'à l'apparition des premières formes proto-vivantes puis vivantes. D'autre part, je pense qu'il est envisageable qu'un système dit « vivant » puisse fonctionner même s'il n'est doté que de systèmes très imparfaits, lents, et à faible fidélité de réplication.

Dans ce schéma général de l'origine d'un code génétique, il manque toutefois un aspect qui me semble essentiel : la polymérisation et l'activité catalytique des petits peptides, ainsi que le couplage de ces processus avec les premières ébauches d'un code génétique, d'une part, avec le métabolisme émergent d'autre part.

8.2.3. Émergence d'une génétique : catalyse des petits peptides, typogénétique et auto-organisation.

S'il est un fait avéré qui pèse d'un poids important sur les réflexions concernant les ères prébiotiques, c'est bien la facilité avec laquelle sont obtenus, en laboratoire mais aussi dans plusieurs milieux naturels (sources chaudes, milieu interstellaire ...[Miller 1954, Muller 2019]) les petites molécules simples comme les acides aminés ou certains composés intervenant dans le

métabolisme énergétique, notamment le cycle de Krebs, mais également des polymères de ces molécules, notamment des peptides (surtout lorsque le milieu contient un support matériel comme des silicates, des métaux ou d'autres surfaces minérales [Imai 1999, Costa 2008, Sakhno 2019, Bu 2019], la condensation sur ces surfaces jouant notamment un rôle déshydratant, facilitant ainsi la formation de la liaison peptidique) parfois dotés d'une activité catalytique, l'ensemble des ces molécules mises ensemble pouvant générer des cycles métaboliques proches de certains cycles métaboliques du vivant actuel comme le cycle de Krebs [Smith 2004, Muchowska 2017 & 2019].

Ce qui est assez clair aussi, c'est qu'il est bien plus difficile de synthétiser en conditions prébiotiques les éléments (oses comme le ribose, bases azotées, squelette ribophosphate) nécessaires à la fabrication des nucléotides et de leurs polymères (ARN et ADN) que dans les systèmes biologiques actuels [Robertson 2012]. Bien que des progrès significatifs aient été fait sur la synthèse abiotique de ces éléments ces 20 dernières années, il reste difficile de concevoir la synthèse de nucléotides phosphatés tant les réactions décrites pour leurs composants sont séparées. De plus, même en considérant effective la formation des nucléotides phosphatés, leur assemblage sous forme de polynucléotides est très peu probable car la constante de formation de dimères de nucléotides est seulement de l'ordre de $1 \text{ mol}^{-1} \cdot \text{L}$. Comme pour les peptides, il est envisageable toutefois de faciliter cette polymérisation par utilisation de *séquences modèles (templates)*, notamment d'autres polynucléotides ou du moins polymères présentant une succession de bases azotées. Plus simples à former effectivement, ces polymères, en particulier ceux fondés sur un squelette d'acides aminés comme les PNA (Peptide Nucleic Acids)[Egholm 1992, 1993, Wittung 1994 ; Knight, 2000] ou les ANA (Alanyl Nucleic Acid) [Diederichsen 1996 et 1997], pourraient être apparus avant les polymères d'ARN et d'ADN et même servir de brins modèles (*templates*) pour la formation de ces derniers.

On peut ainsi s'interroger sur qui du métabolisme, éventuellement fondé sur les peptides, ou du monde à ARN est apparu en premier, ou plus précisément, lequel de ces mondes chimiques a provoqué la structuration et a mis en œuvre l'autre ? Robertson et Joyce débattent en détail de cette question dans leur excellent article de revue vers lequel je renvoie mes lecteurs [Robertson 2012]. Il est très tentant de penser ce processus qui mène à la formation et à l'utilisation (code génétique) des acides nucléiques comme une suite d'étapes allant de la génération des composants élémentaires (acides aminés et bases azotées notamment), en passant par une activité métabolique émergente se renforçant avec l'apport progressif de peptides présentant une activité catalytique (*NB* : A. Brack et B. Barbier [Barbier 1987 & 1992, Brack 1989 & 1990, Perello 1991] ont eu une activité de recherche très prolifique à ce sujet, montrant comment des peptides peuvent se former à partir de simples acides aminés en solution, comment amplifier leur formation ; ils ont aussi montré que ces peptides, dont certains présentant une structure secondaire en hélice α ou en feuillet β , présentaient diverses activités catalytiques [Barbier, 1984 & 1992], en particulier sur les polymères d'acides nucléiques, par exemple les poly(Leu-Arg-Arg-Leu) ou les poly(Arg-Leu) capables d'hydrolyser des chaînes poly-A, ou l'influence sur la polymérisation de précurseurs de l'alanine (2-MeImpA) par des poly(Glu-X-Glu)) et leur auto-amplification, puis l'émergence d'une génétique avec d'une part la formation de PNA ou équivalents, aléatoires en partie, mais aussi possiblement contraints par la

stéréochimie, en particulier celle imposée par les peptides les plus fréquemment rencontrés. En effet, des boucles métaboliques émergentes et auto-amplifiées auront ainsi tendance à sélectionner les composés qui y prennent part, soit en tant que produits, soit en tant que catalyseurs. C'est un des principes de l'autopoïèse défendue en premier par Maturana et Varela [Varela 1974] et implémentée mathématiquement puis informatiquement par exemple dans les systèmes typogénétiques développés par Hofstadter puis Varetto [Hofstadter 1979, Varetto 1993] ou encore d'autres systèmes de vie artificielle fondés sur l'émergence de fonctions à partir d'ensemble d'instructions élémentaires, comme dans les *CoreWars* et ses dérivés *Tierra*, *Avida* ... [Ray 1992 ; Adami, 2000](pour une découverte assez exhaustive de ce monde, j'encourage la lecture du livre *Vie Artificielle* de J.-P. Rennard [Rennard 2002]). Le caractère primordial, voire inéluctable de l'émergence de métabolismes, en particulier de boucles métaboliques auto-amplifiées, trouve d'ailleurs un soutien dans plusieurs travaux théoriques, parfois même pratiques, soulignant le caractère inéluctable de l'apparition de la vie, produit émergent des flux d'énergie physico-chimique et de matière d'une part, de processus auto-organisés d'autre part, qui se mettent inmanquablement en place dans les systèmes naturels abiotiques traversés par des flux de chaleur [Turing 1952, Rovinsky 1993, Smith 2004, Ho 2005, Cornish-Bowden 2007, Morowitz 2007, Muchowska 2017 & 2019].

Les composés actifs comme des peptides dont l'activité catalytique intervient dans l'amplification de telles boucles métaboliques auront ainsi tendance à être sélectionnés parmi la foule de composés formés aléatoirement par des phénomènes d'amplification dans lesquels le 'dialogue' entre chaînes peptidiques actives et chaînes nucléiques informatives jouerait un rôle majeur. Dans le cadre de la théorie stéréochimique du code génétique, de tels peptides, successions d'acides aminés, pourraient servir de *templates* pour des triplets de bases azotées, inscrivant 'en dur' leur séquence dans des séquences de bases azotées (ex : PNA ou ANA, voire même RNA), comme mémoires de leur activité ; cela s'apparente à une sorte de traduction inverse. Réciproquement, les chaînes informatives nouvellement formées peuvent servir de modèles (stéréochimiques) pour l'encodage des peptides correspondants.

Il est absolument évident que si un tel phénomène a pu avoir lieu, il devait être toutefois très inefficace (la fiabilité de traduction devait être très faible). Cependant, il suffit que cet effet (affinité stéréochimique préférentielle) provoque un très léger biais dans le transcodage Peptides \longleftrightarrow Chaînes codantes pour initier un système proto-vivant (au moins du point de vue de la génétique et du métabolisme) fonctionnel. En effet, les processus auto-organisés comme les boucles autopoïétiques présentent cette propriété qu'une fluctuation initiale se retrouve très rapidement amplifiée et maintenue dans le système [Turing 1952, Varetto 1993 & 1998] ; de même qu'un biais appliqué à ce type de système provoque chez eux un devenir différent. De telles boucles, amplifiées par une machinerie enzymatique ou ribozymique, constituent probablement le moteur initial par lequel les formes proto-vivantes se sont distinguées de l'inerte et ont évolué vers des formes acceptablement vivantes.

Des systèmes autopoïétiques ont été plusieurs fois modélisés dans des systèmes de vie artificielle

fondés sur des agents ou sur des systèmes symboliques. Parmi les plus intéressants à mentionner, il y a les systèmes typogénétiques. Ces systèmes symboliques proposés initialement par Douglas Hofstadter dans son livre célèbre, le *GEB* [Hofstadter 1979] (voir figure 8.10), et implémentés par Louis Varetto [Varetto 1993 & 1998], modélisent des systèmes de génétique artificielle. Leur objectif est principalement de répondre au problème de l'auto-référence et de l'auto-réplication. Ce problème, qui a été soulevé dans le chapitre introductif (chapitre 3) est celui qu'a rencontré von Neumann [von Neumann 1966, Codd 1968] qui comprend très tôt la nécessité, pour un système vivant et, de façon générale, pour une machine auto-répliquante, de disposer à la fois d'un plan de montage (l'ADN pour les organismes vivants ; un plan de montage pour toute autre machine, par exemple les boucles de Langton ou son automate auto-répliquant) et d'une machinerie permettant de rendre ce plan opérationnel, autrement dit de le traduire en éléments fonctionnels.

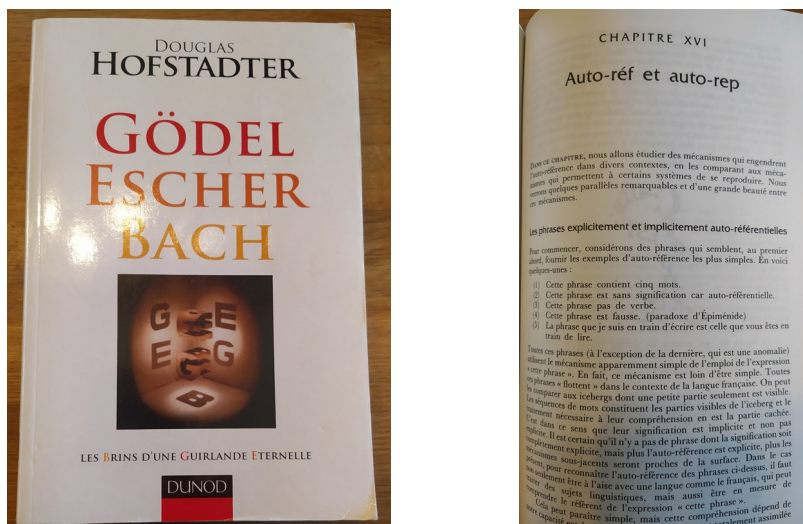


Figure 8.10. Le problème de l'auto-référence et de l'auto-réplication posé par D. Hofstadter dans le GEB [Hofstadter 1979], un livre très particulier, mais qui pose de très nombreuses questions sur la nature de la matière vivante et son fonctionnement.

Dans un modèle typogénétique, des chaînes codantes utilisant 4 lettres comme dans l'ADN, mais un code à 2 lettres (codons de 2 lettres au lieu de 3 comme dans le code standard), permettent, par traduction implicite de typo-gènes, la synthèse de chaînes dites peptidiques, nommées typo-enzymes. Ces typo-enzymes ont une structure primaire, c'est-à-dire une succession de typo-acides aminés, mais aussi une structure secondaire. La structure primaire indique quelle succession d'opérations (ex : *cut*, *copy*, *move left* ...) est réalisée par le typo-enzyme ; la structure secondaire, encodée par des successions particulières de typo-acides aminés, indique les sites du typo-ADN reconnus et modifiés par le typo-enzyme. Les typo-enzymes agissent uniquement sur le typo-ADN. On a ainsi des boucles de traduction typo-ADN → typo-enzyme et d'actions typo-enzymes → typo-ADN (figure 8.11).

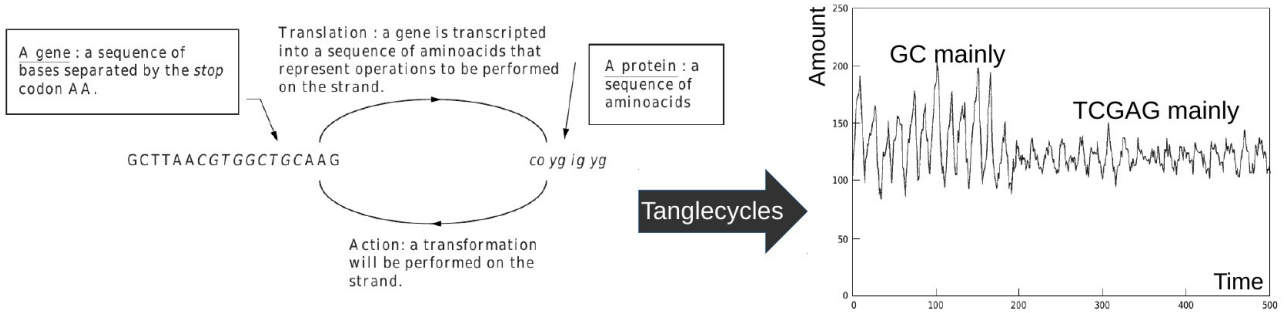


Figure 8.11. Fonctionnement général d'un modèle typogénétique (d'après [Varetto 1993 & 1998]).

Louis Varetto a montré [Varetto 1998] qu'il était possible d'obtenir des boucles auto-répliquantes typo-gènes ↔ typo-enzymes à partir de séquences typogénétiques particulières, nommées répliqueurs. C'est ce même type de formes quasi-autonomes que T. Ray [Ray 1992] obtient dans *Tierra* : des programmes auto-répliquants. Plusieurs recherches sur l'ARN, dont on sait depuis longtemps que certains peuvent présenter une activité catalytique, ont porté sur l'obtention de molécules d'ARN auto-répliquantes. Certaines ont abouti : le ribozyme *Azoarcus Intron group 1* [Drapier 2008, Dolan 2008, Joyathilaka 2008] en particulier est capable de se reconstruire à partir de ses éléments (figure 8.12) (NB : il ne s'agit pas d'une reconstruction autonome toutefois, à savoir pas à partir des composants élémentaires, ni même des nucléotides !).

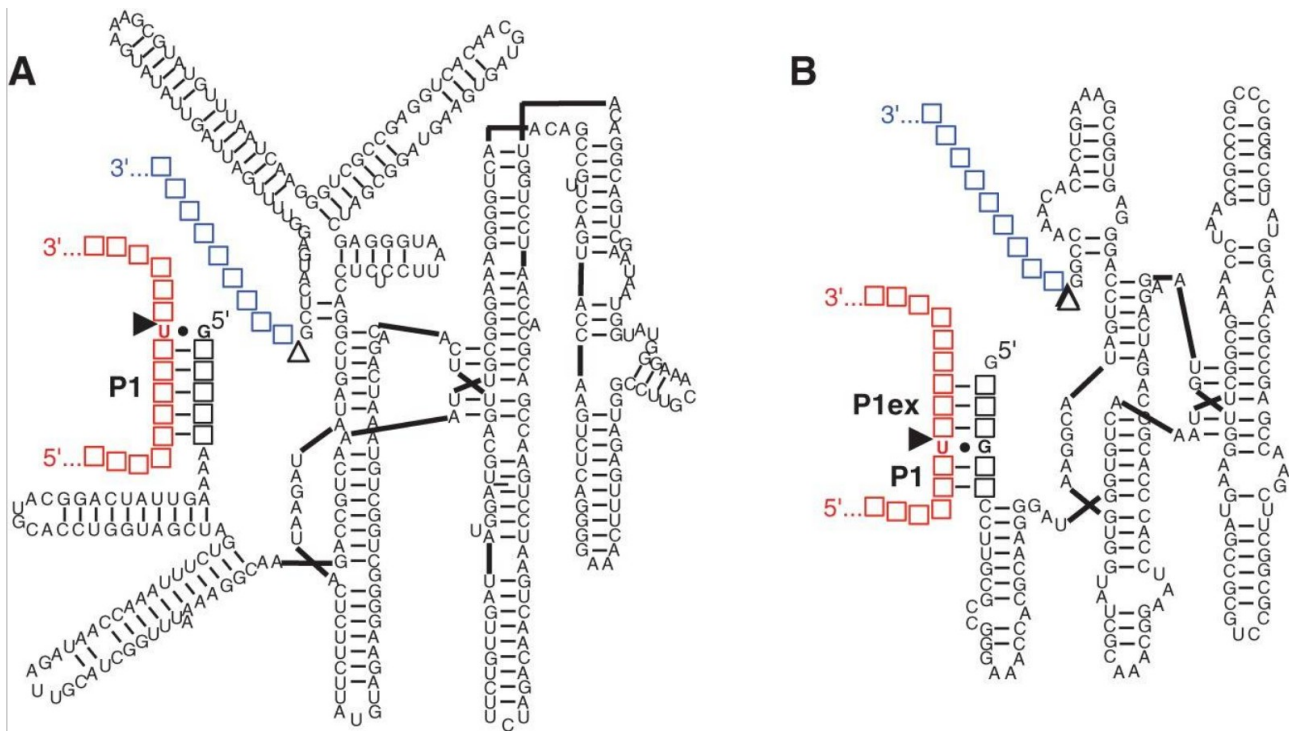


Figure 8.12. Ribozyme Intron du groupe I (d'après [Dolan 2008]) des espèces (A) *Tetrahymena* et (B) *Azoarcus*

Il est toutefois difficile d'imaginer obtenir ce type de molécules, très complexes, dans des conditions prébiotiques hasardeuses, surtout si ces répliqueurs sont le résultat de l'auto-organisation de plus

petites molécules (comme c'est le cas pour les ribosomes actuels) [Sayer 2007]. A noter qu'une autre hypothèse est que les molécules complexes bio-fonctionnelles puissent au contraire s'être formées à partir de simplifications de très grosses molécules de type hydrocarbures lourds. Les travaux récents sur les milieux liquides trouvés dans le système solaire, en particulier les mers d'hydrocarbures de Titan, lune de Saturne, et les aerosols trouvés dans sa ionosphère, les thiolins, suggèrent fortement la possibilité d'un tel mécanisme.

Il reste donc que ces processus (réplicateurs typogénétiques, programmes auto-réplicateurs sous *Tierra*, ribozymes introns du groupe I, ...) ne sont pas autonomes. Ils ont besoin d'une machinerie pour les traduire (dans les systèmes typogénétiques, la traduction *typo-gènes* → *typo-enzymes* est implicite) et/ou permettre leur fonctionnement (un ordinateur est nécessaire pour permettre l'activité des programmes auto-réplicateurs).

Encore une fois, nous pouvons raisonnablement penser que ce dialogue *chaînes codantes* ↔ *molécules effectrices* a pu d'une part émerger, mais également s'auto-amplifier à partir de l'activité de petits peptides formés spontanément et d'affinités stéréochimiques faibles et que, de là, un système proto-génétique a été sélectionné, puis a évolué vers le système génétique actuel composé de chaînes polynucléotiques informatives (car possiblement traduites par un code génétique), et d'effecteurs enzymatiques et ribozymiques. Dans la section suivante présentant les objectifs de cette étude, nous reviendrons sur ces idées et montrerons comment les tester avec des systèmes typogénétiques modifiés.

Jusqu'ici, nous avons longuement peint un tableau décrivant la façon dont des systèmes fonctionnels, métaboliques, potentiellement renforcés par l'activité de peptides, peuvent se mettre en place, ainsi que l'émergence et l'évolution d'un système proto-génétique, permettant de mémoriser dans des chaînes codantes (ARN, PNA ...) certains acteurs essentiels du fonctionnement cellulaire : les protéines.

Une remarque que je trouve intéressante à soulever : on peut constater que l'encodage génétique permet de mémoriser ce qui est difficile à former *spontanément*, c'est-à-dire des peptides, des protéines, dotés de fonctionnalités utiles pour un système vivant. Rien n'est prévu par contre, y compris dans les cellules actuelles, pour mémoriser les systèmes plus simples, à savoir les voies métaboliques, leurs composés (pyruvate, malonate, citrate ...), leurs voies et leurs régulations internes, ces composés qui peuvent se former spontanément dans les conditions de chimie prébiotique, principalement abiotique. On pourra opposer que cette remarque est un non-sens étant donné que cette mémoire n'a pas de raison d'être, le système fonctionnel cellulaire (protéines) étant la machinerie capable de tout reconstituer dans les cellules actuelles, qui se servent des apports de molécules extérieures et les transforment. Je pense cependant qu'il est essentiel de se rappeler ce qu'ont pu être les débuts du métabolisme : plusieurs métabolismes primordiaux, des boucles auto-entretenues couplées à des apports extérieurs, ont pu s'établir sans recours à l'activité protéique ; les

travaux de Muchowska *et al* [Muchowska 2017 & 2019] notamment sont là pour nous en convaincre. Mais on peut ajouter à cela la capacité qu'ont certains composés du métabolisme à faire preuve de comportements auto-organisés complexes : par exemple, la réaction de Belousov-Zhabotinski ou ses variations [Belousov 1958 et 1985, Zhabotinski 1967, Rovinsky 1993, Vanag 2003] reposent sur l'oxydation de molécules comme l'acide malonique et génèrent des motifs chimiques spatio-temporels auto-organisés. Ces petits composés, peuvent se former par des processus abiotiques et disposent en eux de la capacité à faire émerger des comportements complexes (facilement) ; nul *besoin* d'encoder cela. La fin de cette phrase est anthropomorphique, mais il est possible de la retourner : il est en effet possible de voir l'existence de systèmes indispensables à la pérennité de systèmes compliqués à synthétiser (ex : code génétique) et procurant un avantage important dans le développement et l'évolution d'un système vivant (ex : protéines) comme le produit de l'émergence de tels systèmes soumis à la sélection naturelle (Darwinienne). Je pense que si de tels systèmes (protéines spécifiques ayant une fonction indispensable au système vivant) s'avéraient faciles à synthétiser (spontanément, sans recours à un quelconque stockage d'information génétique), l'encodage de l'information génétique n'aurait pas eu lieu d'être. De même, les systèmes (bio-)chimiques s'auto-organisant spontanément ne sont pas nécessairement soumis à la sélection Darwinienne car il n'y a aucun avantage évolutif à être en mesure de stocker dans un support génétique un processus ainsi émergent.

Cette dernière remarque nous amène sur le terrain de l'auto-référence et de l'auto-réplication d'une part, de l'auto-organisation d'autre part. Il commence aussi à être question de complexité, question qui sera assez centrale dans le thème de recherche suivant, « Réseaux biologiques ».

En effet, comme je le disais plus haut, von Neumann comprend que pour faire une machine auto-répliquatrice [von Neumann 1966, Codd 1968], il faut un plan (et j'ajoute à cela *pour encoder ce qui est compliqué à fabriquer dans le système, c'est-à-dire ce qui ne peut émerger spontanément d'un processus auto-organisé*) et il faut une machine, un ensemble d'effecteurs synchronisés, pour donner vie à ce plan, c'est-à-dire le traduire en quelque chose de fonctionnel, à savoir une nouvelle machinerie ... et un nouveau plan, tous deux identiques au plan et à la machinerie initiale. En première approximation, le vivant est globalement fait ainsi : le plan est stocké dans l'ADN et une machinerie cellulaire s'en sert pour répliquer l'ensemble, ADN + machinerie.

Le problème apparaît lorsque l'on essaie d'imaginer une machine qui soit et qui contienne elle-même son plan. Une telle machine doit contenir l'information nécessaire pour savoir synthétiser/usiner les nouvelles pièces pour la nouvelle machine qu'on réplique, mais aussi les positionner comme il faut, à l'identique, pour que tout marche. Là, normalement, un plan, c'est plutôt pratique. L'autre option, sans plan, c'est de s'appuyer sur l'auto-organisation de la machine une fois que les pièces ont été fournies. Si ces pièces sont *faciles* à produire spontanément, par exemple les petites molécules issues de la chimie prébiotique comme les acides aminés, alors le problème de la formation autonome d'une machine à partir de ces pièces se déplace sur la capacité d'auto-organisation de ces pièces en une machine complexe. Remarquons qu'ici on ne peut pas parler de réplication car les pièces (ex : acides aminés) ne sont pas initialement issues de l'activité de la

machine. Nous avons cependant dit précédemment que des boucles métaboliques auto-amplifiés pourraient apparaître et se renforcer à l'aide des produits initialement formés ; il est donc tout à fait imaginable de concevoir un processus auto-réplicateur fonctionnant sur le principe de l'auto-organisation, y compris dans le contexte prébiotique. Le problème se pose quand la machine, l'organisme, le processus à atteindre est compliqué ! Dans ce cas, il va falloir que l'auto-organisation en question soit capable de porter en elle une complexité très élevée, autrement dit que les pièces qui constituent ce système portent en elles l'information nécessaire pour s'organiser avec les autres comme il faut et de façon autonome ; dit encore autrement, il faut que les pièces portent en elles, de façon distribuée, toute la complexité de la machine à faire émerger !

Ainsi, le problème de von Neumann n'en n'est pas vraiment un ; cela dépend seulement de la complexité de la machine à répliquer. Les boucles auto-répliatrices de Langton [Langton 1984] et les variations comme Evoloop, ou plus simple encore, les boucles de Byl et celles de Chou-Reggia [Byl 1989, Reggia 1993, Sayama 1999, Rennard 2002] qui en ont été faites sont une bonne illustration de ce que nous venons de dire : certaines structures auto-organisées peuvent donc être à la fois elle-mêmes (la structure-machine) et une représentation d'elles-mêmes (le plan de fabrication). Langton laisse de côté le caractère de répliateur universel du *kinematon* de von Neumann (et celui de Codd) pour un automate plus simple. Il définit une structure d'automate cellulaire en forme de boucle ouverte sur une queue, possédant 2 catégories de cellules (figure 8.13) : des cellules de membrane (état 2) qui recouvrent de part et d'autre un filament qu'on nommera ADN-machine de la boucle. Ce filament est un automate (chaque cellule voit ses voisines et applique ses règles propres en fonction de son état et de celui de ses voisines ; le parallèle avec le monde biochimique est direct : une molécule interagit avec ses voisines et exerce son activité catalytique sur celles-ci) dans lequel l'information circule, permettant sa propre répliation et la répliation de la membrane.

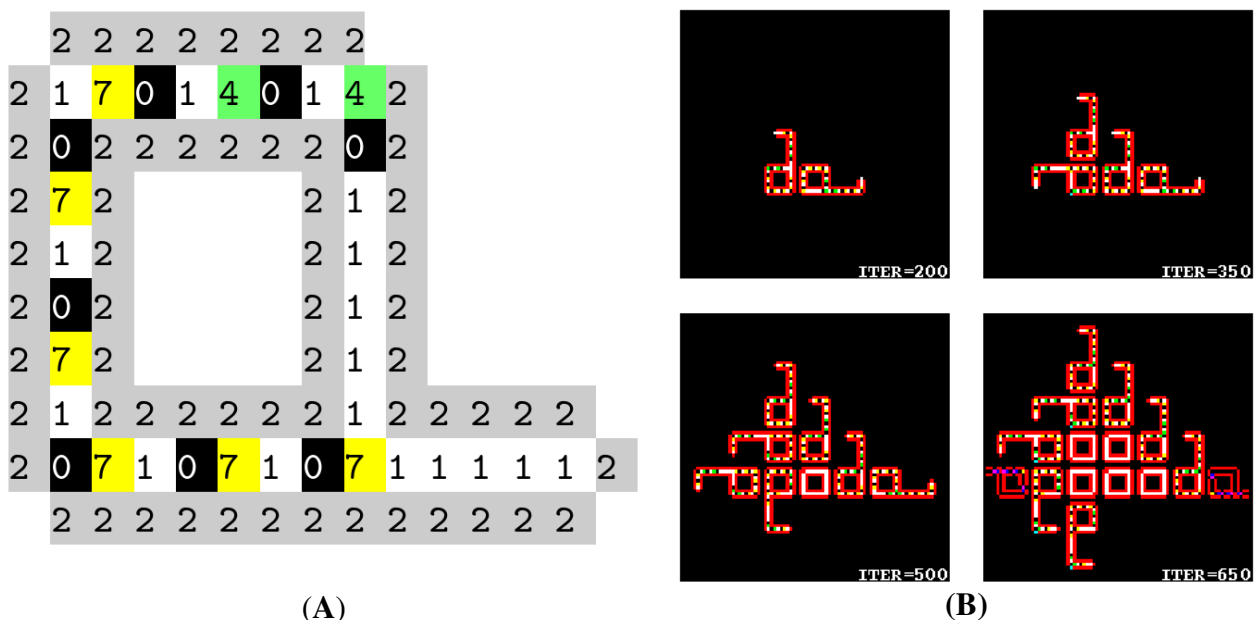


Figure 8.13. Boucle de Langton. (A) Configuration initiale de la boucle de Langton : la boucle initiale comporte 94

cellules. L'automate fonctionne avec 8 états et 29 règles de transition. Un filament fait d'états 1 et de signaux codés par les états 4 à 7, suivis par des 0 (ex : le signal '401' est une instruction de virage ; le signal '107' est un signal d'avancement), est entouré d'une membrane faite d'états 2. Le code génétique est fait du filament. C'est ce même filament qui porte en lui la capacité de réplication, ce parce que de nombreuses règles de transition sont permises entre seulement quelques configurations de cellules. (B) Boucles de Langton se répliquant dans l'environnement de l'automate cellulaire (tiré de [Tempesti 1998]).

Notons cependant quatre points à ce sujet :

i) D'abord, la structure à répliquer est simple, très simple en comparaison de l'être vivant le plus simple qui soit. Parce que cet automate est petit (94 cellules initiales), il est tout à fait envisageable que des particules-états (état 2, état 4 ...) de cet automate cellulaire, dispersées dans l'environnement, et dont on pourrait imaginer qu'elles apparaissent spontanément (par synthèse de type prébiotique), puissent *par hasard* ou par interactions préférentielles (type stéréochimique) se retrouver toutes ensemble dans une configuration qui est celle de la boucle de Langton. C'est peu probable, mais c'est possible. Cette remarque est encore plus vraie si l'on prend d'autres structures d'automates auto-répliquatrices plus compactes comme les boucles de Byl ou celles de Chou-Reggia [Byl 1989, Reggia 1993].

ii) Il faut de plus garder à l'esprit que les automates sont implémentés et instanciés dans des ordinateurs qui ont leur propre complexité de fonctionnement électronique et qui leur donnent vie. Les automates ne sont ainsi pas autonomes. Leurs règles portent en effet elle-mêmes une certaine complexité qui impacte leur mise en œuvre dans un ordinateur. Une sommation d'états, la prise en compte d'une condition, *etc*, sont des opérations qui contiennent de la complexité.

iii) Ensuite, la densité importante d'informations (états et règles) dans ces automates rendent ceux-ci d'une part très fragiles et d'autre part très peu évolutifs. Langton, pour parvenir à ses fins a du réfléchir, en tant qu'ingénieur (chercheur), à une créature numérique très optimisée dans laquelle chaque partie (chaque état cellulaire et chaque interaction possible), porte en elle une très grande complexité. La densité de comportements riches est très grande dans une boucle de Langton : dans un ensemble ordonné (en boucle) comprenant une centaine de cellules dont les états vont de 0 à 7, il y a 29 règles de transition ! Dans ces automates le moindre changement de règle (erreur d'exécution locale) ou de configuration (voisinage mal configuré ou mal lu) provoque une erreur fatale à la créature virtuelle. De plus, envisager l'évolution ou ne serait-ce que des variations accidentelles du *code* d'une telle créature virtuelle, le filament d'ADN-machine de la boucle de Langton, est très complexe sans réduire son fonctionnement inopérant.

Le vivant n'est pas fragile. Dans le vivant, beaucoup d'interactions sont stériles et une très petite partie de chaque molécule a une activité catalytique, ceci notamment parce que chaque acide aminé, nucléotide, sucre, ..., n'a qu'une faible capacité d'interaction et de réaction avec d'autres molécules (peu d'atomes ou de groupements d'atomes exposés et disposant de liaisons suffisamment labiles et d'électrons à mettre en commun avec un éventuel groupement extérieur). Une mutation (par exemple le changement d'un acide aminé pour un autre à cause d'une mauvaise transcription ou traduction) est ainsi souvent sans effet car d'abord beaucoup d'entre eux n'interviennent que dans la

structure d'une protéine et pas dans le site actif, ensuite parce que beaucoup de mutations sont faites avec des acides aminés physico-chimiquement équivalents, que beaucoup de ces mutations ne modifient que peu le comportement de la protéine, n'affectant pas la viabilité de l'organisme et qu'enfin et surtout, une erreur ponctuelle dans le processus de transcription/traduction ou, plus tard, une activité erronée ou une interaction malvenue entre plusieurs éléments cellulaires, n'aura que pas ou peu d'effet car l'activité des quelques molécules affectées par cette mutation restera négligeable dans une cellule dans laquelle la plupart des molécules et interactions fonctionnent bien. Ce raisonnement est d'autant plus vrai dans les organismes pluricellulaires. Dans le vivant, l'activité enzymatique émerge de longues chaînes protéiques ou ribozymiques et de leur structure complexe ; la vie, elle émerge du très grand nombre d'interactions synchronisées entre ces molécules complexes la génétique, par le pouvoir qu'elle a de permettre la génération rapide mais modulée de ces molécules complexes, jouant en plus du rôle de mémoire celui de point central où se fait le contrôle, où naît la synchronie. Il est difficilement imaginable de se passer de la combinaison plan-machine (ADN-machinerie cellulaire) dans ces conditions.

iv) Le vivant est évolutif et cela lui permet de créer une diversité d'organismes (d'espèces). Cette diversité est garante de l'adaptation de la vie à une variation des milieux. Ne pas permettre l'évolution d'un système vivant, c'est le condamner tôt ou tard à périr face à un milieu qui aurait changé pour des conditions trop différentes.

Pour résumer cette digression, on comprend dès lors que, même dans un contexte de biochimie prébiotique extrêmement simple par rapport au vivant actuel, le moindre biais allant dans le sens de l'apparition d'un codage génétique aura toutes les chances d'être 'retenu' d'abord parce que fabriquer les éléments complexes (protéines) de façon fiable est rendu possible, ensuite parce que les systèmes qui disposent d'un plan sont plus viables. S'ils n'incluent pas de plan (génomique), ils sont uniques car trop de leurs éléments sont générés au hasard. A long terme, une collection de systèmes semblables survivra. Ajoutons à cela le caractère évolutif et donc adaptable de ces systèmes rendu possible par une possible complexification du génome d'une part (par duplications, délétions, mutations des chaînes codantes), du code génétique (par expansion notamment) d'autre part. L'encodage génétique de l'information biochimique constitue donc, dès le stade prébiotique, un avantage sélectif par rapport à toute forme autonome émergente auto-organisée qui sera peu extensible et peu adaptable.

8.2.4. Membrane et microcosme

Il n'est pas envisageable que la mise en place de systèmes fonctionnels primordiaux (proto-réseaux métaboliques et proto-génétique) puisse se faire dans un espace libre de diffusion, ne procurant aucun confinement chimique. La dilution par diffusion et autres phénomènes de transport de matière vient vite à bout des réactions (bio-)chimiques qui pourraient timidement s'établir. La concentration, la condensation sur, ou proche des surfaces est nécessaire pour qu'elles puissent se faire correctement, souvent s'initier, et surtout perdurer. De plus, délimiter l'espace ainsi crée l'unité du vivant ou du proto-vivant [Hanczyc, 2008]. La délimitation introduit une différence entre un

intérieur et un extérieur, entre le vivant et l'inerte. Cette différence introduit des gradients dont les systèmes vivants peuvent tirer un profit énergétique ou signalétique (ex : le gradient de protons dans les membranes mitochondriales permet le fonctionnement de l'ATP synthase ; les gradients ioniques de part et d'autre de l'axone assurant la conduction de l'influx nerveux). De plus, si cet espace est divisible, alors de l'unité on a la multiplicité : la reproduction devient possible. De l'imperfection qui accompagne cette reproduction naît l'évolution, la diversification des formes de vie et son adaptation aux conditions du milieu par sélection naturelle [Lopez, 2019].

C'est ce que de von Neumann à Langton, les chercheurs en vie artificielle ont voulu obtenir. C'est le cas aussi des expérimentateurs qui ont cherché à reproduire cet ensemble de propriétés du vivant (contenant de réactions, semi-perméabilité avec le milieu extérieur, reproduction, évolution) [Deamer, 2008], en particulier via des cycles de formation et de division de gouttelettes eau dans l'huile et réactifs dans l'huile, ou de micelles, dans des dispositifs microfluidiques [Jeancolas, 2020].

La membrane, surtout dans ce contexte chimique balbutiant, c'est donc l'élément structurant nécessaire à la vie. L'obtention de membranes lipidique est cependant complexe. La formation des lipides n'est pas triviale. De nombreuses propositions ont été faites (que je ne vais pas passer en revue ici) pour passer outre le problème de la formation de lipides : substrats métalliques, silices, argiles ... D'autres pensent que la formation des lipides fut possible dans des quantités acceptables. Je n'entrerai pas dans cette question ouverte et également très débattue. Je souhaiterais cependant apporter une suggestion pour la structuration de l'espace aux origines du vivant : l'auto-organisation spatiale.

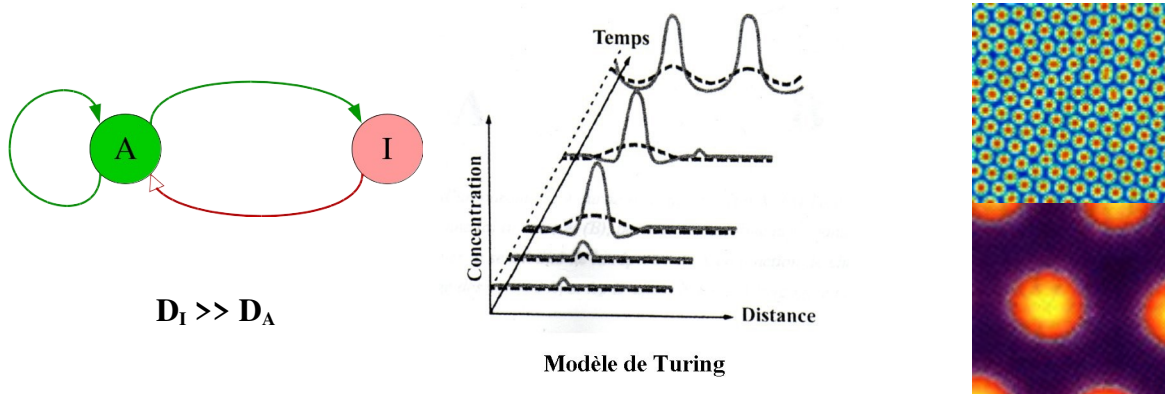


Figure 8.14. Auto-organisation de motifs de Turing. (Gauche) Un activateur A active sa propre formation et celle d'un inhibiteur I. L'inhibiteur inhibe la formation de l'activateur. De plus, la diffusion de l'inhibiteur est très supérieure à celle de l'activateur. (Centre) A partir d'une fluctuation spatiale initiale dans la composition en A et/ou I (ici A), l'hétérogénéité spatiale s'amplifie et se réplique : A se forme mais forme également I. I en diffusant davantage que A inhibe latéralement sa formation. Cela crée latéralement deux nouvelles hétérogénéités qui s'amplifient à leur tour. (Droite) Vues de dessus dans un environnement en 2 dimensions, en fausses couleurs, ces simulations montrent l'obtention d'hétérogénéités régulièrement réparties à l'état stationnaire.

Alan Turing en 1952 [Turing, 1952], après les travaux précurseurs de Kolmogorov et de Rashevsky [Kolmogorov, 1937 ; Rashevsky, 1940 et 1948], théorisa le premier l'auto-organisation spatio-temporelle de motifs chimiques stationnaires, proposant ainsi que des réactions chimiques particulières (en l'occurrence un activateur qui active sa propre formation et celle d'un inhibiteur qui, lui, inhibe la formation de l'activateur) couplées à une diffusion différentielle des espèces chimiques en question, puissent générer des motifs auto-organisés de concentration des espèces chimiques présentes. Plus largement que les motifs de Turing, de nombreux phénomènes de réactions/interactions couplés à des phénomènes de transport de la matière agissent dans la nature, causant des structurations et compartimentations : sédimentation couplée à des réactions [Glade, 2017], milieux excitables comme la réaction de Belousov-Zhabotinski ou de Bray [Bray, 1921 ; Belousov 1951 et 1958 ; Zhabotinski, 1967], trail systems (voir [Glade, 2009]), *etc.*

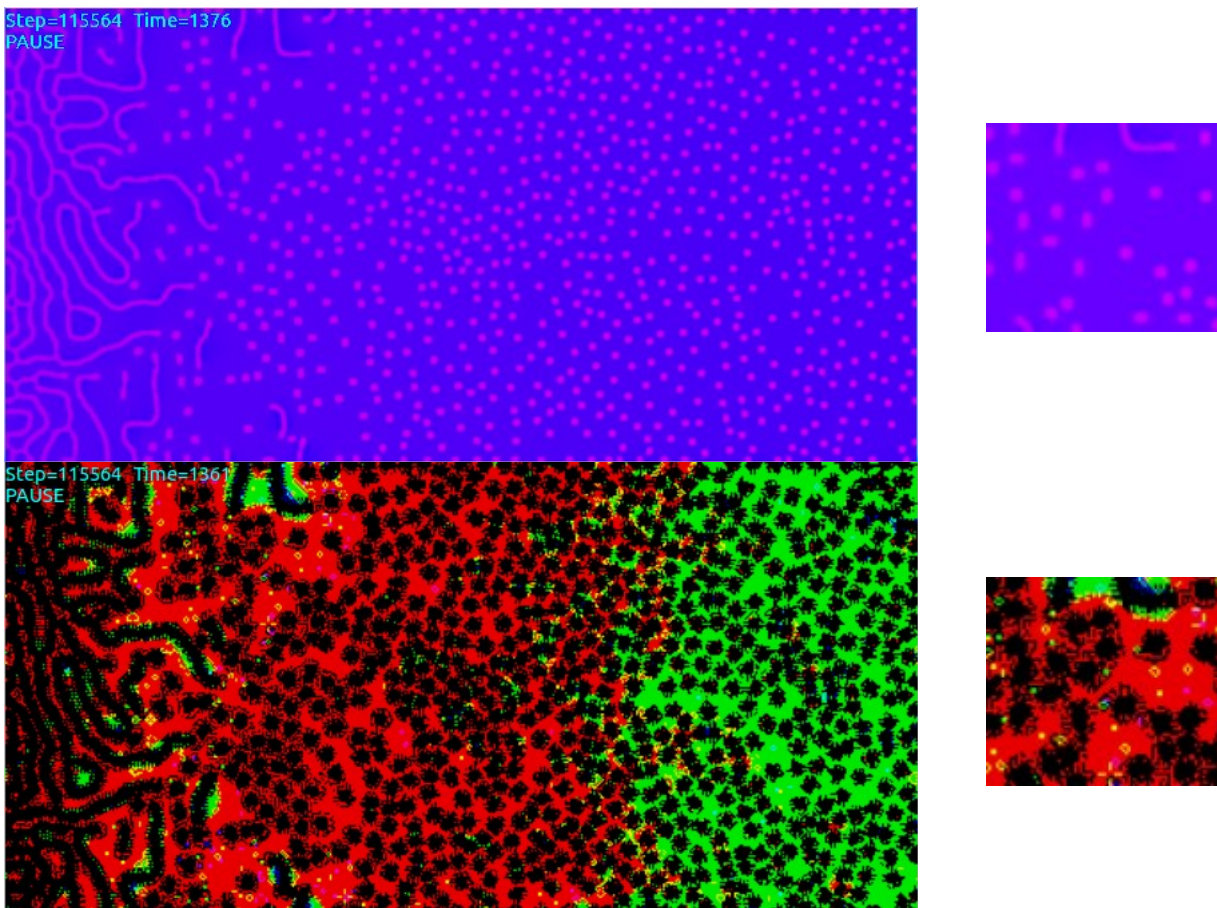


Figure 8.15. Couplage d'un système auto-organisé de Turing avec un jeu de la Vie. Dans cette expérience numérique que je me suis amusée à écrire en OpenCL sous MatrixStudio [MatrixStudio] durant un voyage en avion en direction de Brest en 2011, dans l'idée de m'entraîner à développer en OpenCL, 3 jeux de la vie de Conway [Gardner, 1970] différents et indépendants (visible sur les canaux couleurs Rouge, Vert, Bleu) sont couplés à un système de réaction-diffusion de Turing. Le couplage entre les deux systèmes se fait comme suit : lorsque la concentration en inhibiteur se situe dans une certaine fourchette de valeurs (différente pour les 3 jeux de la vie), il y a naissance de cellules ; quand la concentration en activateur dépasse un certain seuil (variable selon la position en x), les cellules meurent. La réaction du système de Turing ne se fait que dans les endroits où il n'y a pas de cellules de jeu de la vie vivantes.

Au delà de l'amusement et de l'expérience d'OpenCL que cela m'a apporté, cette simulation a le mérite d'illustrer le résultat d'un couplage entre un système auto-organisé et un système de vie artificielle. On observe en haut un motif auto-organisé de type Turing (affecté par le couplage avec les 2 jeux de la vie) et en bas les 3 jeux de la vie visibles sur les 3 canaux couleur R, V, B. C'est peu lisible, certes, mais on voit tout de même des créatures du jeu de la vie (n jaune notamment) se développer dans certaines régions seulement. Elles sont contenues dans des régions particulières du système de réaction-diffusion (voir le zoom à droite).

De l'ensemble de ces théories de l'auto-organisation par des réactions couplées à des phénomènes de transport de la matière s'est développé un gigantesque champ de recherches en biologie théorique [Nicolis, 1977 ; Bard Jonathan, 1981 ; Meinhardt, 1982 ; Harrison, 1987 ; Murray, 1990 ; Helbing, 1997 ; Asai, 1999 ; Sept, 1999 ; Glade, 2002 ; Maini, 2003 ; Suzuki, 2003] : ces phénomènes permettent d'expliquer à moindre coût, c'est-à-dire sans besoin d'un contrôle complexe, la formation de motifs naturels comme les taches et les zébrures des animaux, mais aussi certaines structuration de populations comme les insectes sociaux, ou encore l'auto-organisation de microtubules *in vitro* [Cohen, 1971 ; Tyson, 1989 ; Goss, 1989 ; Mandelkow, 1989 ; Beckers, 1992 ; Kondo, 1995 ; Bonnabeau, 1999].

Revenons au processus d'auto-organisation décrit par Turing, puisque c'est celui qui nous intéresse ici. Ce processus décrit dans la figure 8.14 génère des régions (des spots) concentrées en activateur séparées par des régions pauvres en activateur. De même, on observe un motif d'inhibiteur couplé au motif d'activateur. Dans les spots, l'activateur est concentré car celui-ci active sa propre formation et que la concentration en inhibiteur ne permet pas son élimination locale. En revanche, dès qu'on s'éloigne du centre du spot, l'inhibiteur prend vite le dessus et sa concentration suffit à empêcher l'activateur de se concentrer (par auto-catalyse). On sait obtenir chimiquement ce phénomène [Castets, 1990 ; Vanag, 2003] en utilisant des réactifs chimiques simples dont certains jouent le rôle d'activateurs et d'autres d'inhibiteurs. Il est tout à fait envisageable que les réactifs de la phase prébiotique puissent jouer le même rôle. Ainsi, en plus de permettre l'apparition de boucles proto-métaboliques (donc de nature temporelle), ces réactifs forment, par des processus de réaction couplés à des phénomènes de transport, des motifs spatiaux auto-organisés stationnaires (motifs de Turing) ou spatio-temporels (des front d'ondes chimiques). Dans ce contexte, les réactifs protagonistes se verraient auto-amplifiés et maintenus spatialement ; autrement dit, un tel processus d'auto-organisation pourrait bien jouer le rôle de protocellule, tout du moins de *microcosme biochimique*, hébergeant et auto-maintenant un ou plusieurs proto-métabolismes (voir figure 8.15). Remarquons tout de suite que dans ce cas, la question de la reproduction est assurée car, dans ces processus, le motif auto-organisé se répète naturellement (voir figure 8.14).

Mieux : dans l'espace du milieu réactif, il se pourrait que des motifs auto-organisés entrent en compétition ou se côtoient dans plusieurs microcosmes voisins, formant ainsi un écosystème de proto-systèmes vivants. Par la suite, certains, plus efficaces que d'autres pourraient prendre le dessus, ou encore certains proto-systèmes pourraient fusionner avec d'autres. Ainsi, un tel processus disposerait de capacités évolutives.

D'autres processus (ex : la sédimentation gravitationnelle) pourraient aussi, dans certaines conditions (dans le cas où les variations de densité du milieu seraient suffisamment importantes

pour subir l'action de la gravité, par exemple si les réactifs forment des amas denses comme les *composomes* proposés dans [Hunding, 2006 ; Norris, 2007]), provoquer une brisure de symétrie spatiale et contribuer à séparer l'espace en régions différenciées d'activité et de diversité, créant ainsi des écosystèmes biochimiques un peu comme O. Bastien, P. Ballet et moi-même l'avions montré dans [Glade, 2017].

Nous verrons qu'une partie du projet consiste justement à exploiter la possibilité de tels processus de séparation du milieu dans des simulations de vie artificielle, en particulier de génétique artificielle, pour espérer renforcer l'émergence et le maintien de systèmes auto-entretenus.

8.2.5. Objectifs du projet de recherche

Dans ce projet de recherche, je souhaite réaliser une étude exploratoire dans le contexte des origines du vivant. Ce projet comporte deux volets : expérimental d'une part, théorique d'autre part. L'objectif est de proposer un scénario plausible sur la mise en place d'une ébauche de système génétique maintenu dans l'espace et dans le temps.

Le projet est composé de deux sous-projets :

- une partie expérimentale dont l'objectif est de tester l'hypothèse stéréochimique du code génétique. Nous avons vu cette hypothèse, soit dans sa forme brute (hypothèse de Pelc, Welton et Dunnill [Pelc, 1965 ; Pelc & Welton, 1966 ; Dunnill, 1966 ; Hendry, 1981]), soit sous la forme d'une contribution stéréochimique primordiale permettant l'association entre acides aminés et proto-ARNt (hypothèses de Koonin ; voir [Koonin, 2017 a et b ; Yarus, 2009]). En pratique, pour cela, nous allons mesurer l'affinité spécifique entre des acides aminés et des petites séquences de nucléotides.
- une partie théorique qui s'intéresse à l'auto-organisation et à l'effet de la sédimentation dans des proto-systèmes génétiques et métaboliques.

Ces deux sous-projets se succèdent dans l'ordre fonctionnel du schéma que l'on propose ici pour les origines du vivant : l'apparition de proto-systèmes génétiques, donc d'un couplage fonctionnel entre molécules de stockage de l'information génétique et molécules opérationnelles (peptides et protéines) requiert la possibilité d'un transcodage du système génétique vers le système protéique, potentiellement soutenu par l'hypothèse stéréochimique du code génétique. Valider l'hypothèse stéréochimique est essentiel pour accepter plus facilement les modèles autopoïétiques de génétique prébiotique (en particulier les modèles typogénétiques) [Varela, 1974 ; Hofstadter, 1979 ; Varetto, 1993 et 1998] car elle résoudrait en grande partie le problème de la traduction ancestrale de séquences polynucléotidiques en peptides.

8.2.6. Réalisation

La partie expérimentale du projet consiste en la mesure, par des techniques de calorimétrie (Isotherm Titration Calorimetry), de RMN et de spectrofluorimétrie, des forces d'interaction entre des acides-aminés et leurs anticodons respectifs dans le code génétique, et à mettre éventuellement en évidence la spécificité de cette reconnaissance amino-acide \Leftrightarrow anticodon correspondants dans le code génétique. Ces mesures donneront accès à certains paramètres cinétiques, notamment les constantes d'association et dissociation des acides aminés sur divers trinuécléotides.

8.2.6.1. Types de mesures

Les mesures que nous allons réaliser sont extrêmement sensibles car les interactions attendues sont très faibles et parce que les signaux (de fluorescence notamment) le seront aussi. La mise au point de telles expériences constitue en soit une avancée scientifique.

Nous utilisons 2 types de mesures indépendantes :

- des mesures de micro-calorimétrie : l'interaction spécifique de codons et d'acides aminés en solution concentrée doit provoquer une variation d'enthalpie mesurable par un micro-calorimètre (ou ITC pour *Isotherm Titration Calorimetry*). L'ITC donne accès, dans le cadre d'un modèle d'interaction prédéterminé et d'une stœchiométrie donnée, à tous les paramètres cinétiques, notamment aux constantes d'association et de dissociation. En pratique, l'expérience consiste à doser (titrage) une solution d'oligonucléotides par injections successives d'acides aminés (voir ci-dessous pour la description de la titrage isotherme).
- des mesures de quenching de fluorescence : on peut s'attendre à ce que la proximité des molécules en interaction (ici codons et acides aminés dans une interaction spécifique) provoque un quenching de fluorescence de l'ADN (anticodons) mesurable par un spectrofluorimètre. L'objectif de cette mesure par une méthode indépendante est de confirmer la présence d'interactions spécifiques acides-aminés – anticodons (voir ci-dessous pour la description de la fluorimétrie).

A ces deux types de mesures s'ajoutera peut-être un 3ème si les résultats sont probants :

- des mesures de quenching RMN. Le principe est le même que celui du spectrofluorimètre : la proximité prolongée d'acides aminés avec des séquences nucléotidiques doit provoquer un élargissement des raies spectrales d'atomes de ces nucléotides. Cette technique repose sur l'effet NOE (Nuclear Overhauser Effect) qui décrit une interaction entre deux spins à travers l'espace et non pas à travers les liaisons chimiques comme le couplage scalaire. Cette interaction est limitée à environ 5-6Å [source *Wikipédia*]. L'intérêt de cette technique dont la mise en œuvre est plus lourde est de donner accès à la structure dans l'espace des atomes des molécules en interaction.

8.2.6.2. Description des dispositifs expérimentaux

Dans ce paragraphe, je prends le temps de décrire les 3 dispositifs expérimentaux dont nous disposons pour réaliser nos mesures car ces dispositifs ne sont pas communs. Cela me permet de faire ressortir la quantité et la qualité du travail expérimental que je souhaite réaliser dans les années à venir.

8.2.6.2.a. Microcalorimétrie par titrage isotherme – mesure avec un ITC

Le Titrage Calorimétrique Isotherme est une technique essentiellement utilisée dans les études quantitatives des interactions biomoléculaires. Elle consiste à mesurer la chaleur libérée ou absorbée, durant une liaison entre des biomolécules.

Cet appareil est composé de deux cellules calorimétriques³⁰ identiques. Une cellule de référence remplie avec de l'eau et une autre contenant l'échantillon. Les principales caractéristiques de ces cellules sont les suivantes :

- Elles sont thermiquement conductrices et chimiquement inertes,
- Elles sont entourées d'une enveloppe adiabatique.

Une seringue ajoute progressivement le ligand (dans notre cas des acides aminés) dans la cellule échantillon contenant la solution cible à titrer (dans notre cas des oligonucléotides). Mise en rotation, la pointe de la seringue assure également le mélange de l'injectât dans le solution cible. Les mesures sont réalisées à température constante.

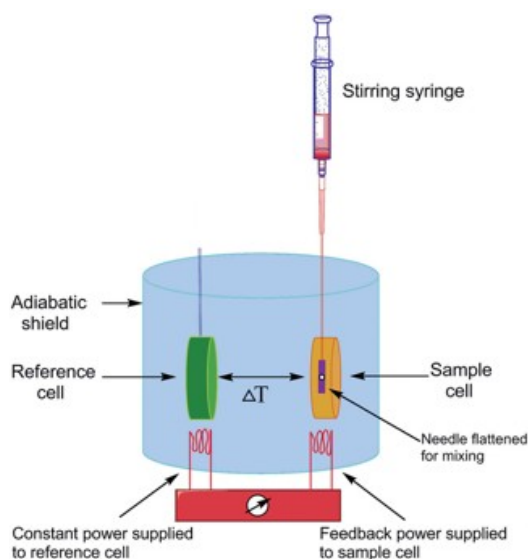


Figure 8.16. Schéma d'un ITC. L'ITC est constitué de 2 cellules : une de référence, en vert, l'autre de mesure, en orange. Dans cette dernière se trouve la solution cible et plonge une seringue permettant d'injection du ligand. L'appareil mesure (en rouge) la puissance nécessaire au ré-équilibre de température des deux cellules.

³⁰ On désigne par « cellule » une petite cuve contenant quelques centaines de microlitres, dans laquelle est effectuée la mesure calorimétrique.

Lors d'une expérience, on place une solution de molécules cibles (oligonucléotides) dans la cellule de mesure. Celle-ci est progressivement saturée par les injections automatiques et régulières de ligand via la seringue. A chaque injection, un échange thermique caractérise l'interaction entre la molécule et le ligand. Ces échanges se caractérisent par une variation d'enthalpie et un dégagement ou une absorption de chaleur. Le calorimètre mesure la différence de température ΔT entre la cellule de référence et la cellule échantillon. Pour faire cette mesure, le calorimètre fournit la puissance de rétroaction nécessaire pour revenir à la condition isotherme en augmentant ou en diminuant la température de la cellule échantillon. C'est cette puissance qui est mesurée (voir figure 8.16).

Les données brutes sont donc des valeurs de puissance exprimées en micro-calories par seconde. Une fois ces données brutes collectées par l'appareil, il reste à intégrer les pics de variation d'énergie permettant ainsi d'obtenir une représentation graphique de la variation d'enthalpie en fonction du rapport molaire ligand/molécule cible. A partir de cette courbe, on peut déterminer de façon graphique la stœchiométrie de la réaction n , la constante de dissociation K_D et la variation d'enthalpie moyenne ΔH (figure 8.17).

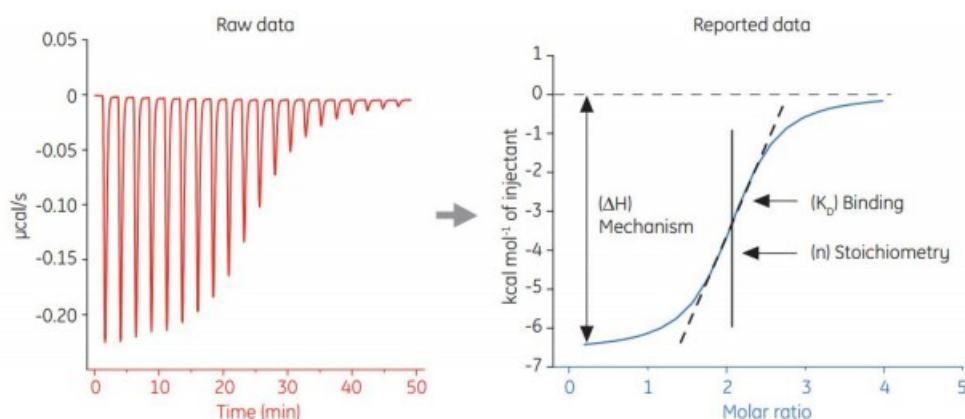


Figure 8.17. Données obtenues par ITC. (Gauche) Données brutes : chaque pic dirigé vers le bas et se relaxant vers le haut représente la puissance nécessaire à l'ITC pour équilibrer les températures des 2 cellules après injection de ligand dans la solution cible, chaque pic démarrant juste après l'injection. L'amplitude des pics diminue au fur et à mesure que les molécules cibles sont saturées de ligands et n'interagissent plus avec les nouvelles molécules de ligand ajoutées. (Droite) Courbe de la variation d'enthalpie en fonction du ratio molaire entre ligand et cible. Cette courbe est obtenue par intégration de chaque pic de la courbe de puissance. De cette courbe sont extraites l'enthalpie moyenne, la constante de dissociation et la stœchiométrie de la réaction.

Le principal avantage de cette méthode est qu'elle est directe. Elle repose sur une mesure directe de la chaleur libérée ou absorbée pendant une réaction entre la molécule et le ligand. De plus, une fois la mesure réalisée, on peut directement déterminer le graphique tous les paramètres de liaison en une seule et unique mesure. Elle ne nécessite aucune modification de l'échantillon ou du ligand, que

ce soit par marquage fluorescent ou par immobilisation... L'ITC mesure directement l'affinité des participants à la liaison dans leur état natif.

Un des inconvénients est que cette méthode est longue à réaliser. En particulier, une étape de calibrage pouvant prendre des dizaines de séances (chaque séance durant une demi-journée à une journée à cause de l'équilibrage des températures des cellules) est nécessaire. De plus, nous verrons que de nombreux tests et blancs doivent être faits, en particulier pour s'assurer de la neutralité des tampons.

8.2.6.2.b. Fluorimétrie

Dans ce paragraphe je décris le fonctionnement d'un fluorimètre en commençant par rappeler le principe de fluorescence, celui de quenching et la question particulière de la fluorescence des acides nucléiques.

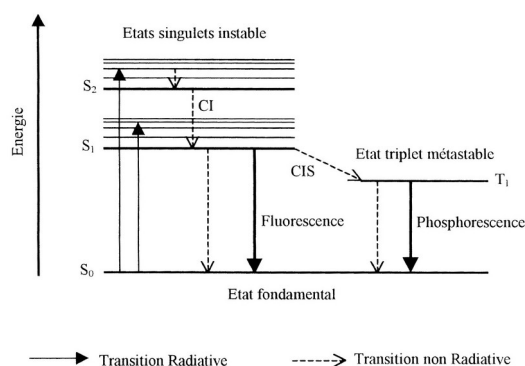


Figure 8.18. Diagramme de Jablonsky. Cette représentation explique le principe de fluorescence suite à l'excitation d'une molécule en montrant ses changements d'états énergétiques.

Fluorescence. La fluorescence est la propriété que possèdent certaines molécules à absorber de la lumière à une longueur d'onde pour ensuite l'émettre à une longueur d'onde égale (on parle de fluorescence de résonance) ou avec une longueur d'onde plus grande. L'émission est émise après un temps qui va correspondre à la durée de vie de fluorescence. Les électrons d'une molécule peuvent passer d'un niveau d'énergie fondamental à un niveau d'énergie supérieur par absorption d'énergie (sous forme de photon). Dans ce cas, la molécule est dite excitée. Cette absorption d'énergie se traduit par le déplacement d'un électron sur une couche supérieure. Pour que la molécule passe d'un niveau fondamental S_0 à un niveau excité S_1 il faut que l'énergie du photon soit supérieure ou égale à l'énergie de S_1 . Cette molécule peut revenir à son état fondamental en émettant de la lumière. Ces photons sont ré-émis avec une énergie inférieure aux photons d'excitation. Cette perte est due à une perte d'énergie liée aux transitions non-radiatives vers l'état vibratoire le plus bas de l'état S_1 . Elle entraîne une augmentation de longueur d'onde (donc d'énergie plus faible car l'énergie est liée à la longueur d'onde par la relation $\nu = hc/\lambda$) : la longueur d'onde de la lumière émise est supérieure

à la longueur d'onde absorbée. Le déplacement de Stokes est le décalage entre la longueur d'onde de la lumière émise et la longueur d'onde de la lumière absorbée (Figure 8.18).

Fluorescence des acides nucléiques. Toutes les molécules ne sont pas capables d'absorber de la lumière et plus précisément les photons. Pour faire de la spectrofluorimétrie, il faut des molécules capables d'absorber et d'émettre des photons (des fluorophore). Les acides nucléiques présentent une fluorescence entre 300 et 400 nm pour une excitation à 267 nm (Figure 8.19). Cette fluorescence est cependant très faible et de courte durée (relaxation rapide) [Vaya, 2010 ; Markovitsi, 2012] que nous souhaitons cependant exploiter.

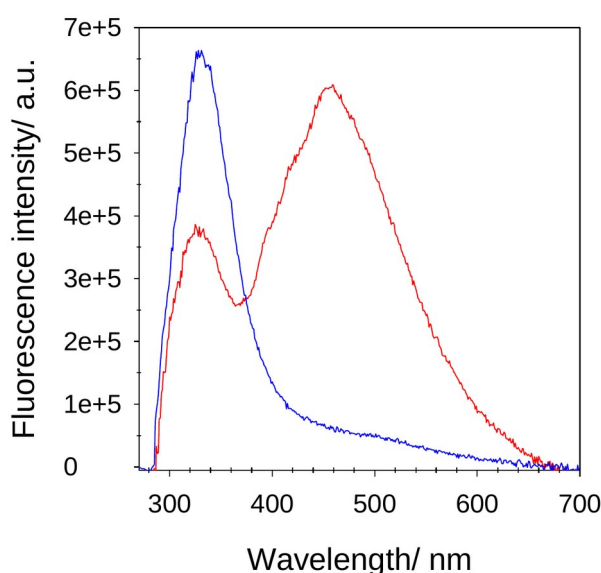


Figure 8.19. Fluorescence des acides nucléiques. La courbe bleue montre le spectre de fluorescence d'ADN de thymus de veau dans du tampon phosphate après excitation à 267 nm. (la courbe rouge est à ignorer). La figure est extraite de [Vaya, 2010].

Fluorimètre. Un spectrofluorimètre est utilisé pour mesurer l'intensité de fluorescence. Une source continue de lumière excitatrice de type lampe au Xénon vient exciter les électrons de certaines molécules présentes dans la cuve en quartz. En réponse, ces molécules émettent de la lumière de plus basse énergie. L'utilisation d'une lampe au Xénon à large bande spectrale permet d'avoir *une gamme d'excitation importante* entre 200 et 800nm. L'échantillon reçoit une lumière avec une longueur d'onde monochromatique, sur une des faces de la cuve de section carrée. Cette lumière excite les fluorophores présents qui réémettent de la lumière de longueur d'onde plus faible. Cette lumière est analysée en sortie sur une face de la cuve orthogonale à celle d'excitation. Seule la lumière issue de la fluorescence peut provenir de cette face (voir figure 8.20).

L'analyseur (monochromateur d'émission), comme le dispositif d'excitation (monochromateur d'excitation) sont mobiles et peuvent sélectionner une longueur d'onde assez précise (avec une fenêtre d'environ 1 nm de large en général). Il est possible alors de balayer la longueur d'onde

d'excitation et, pour chaque longueur d'onde d'excitation, balayer le spectre d'émission. Cela permet l'obtention de spectres complets d'excitation-émission.

Figure 8.20. Schéma de fonctionnement d'un spectrofluorimètre. La lumière produite par la source d'excitation (au xénon) passe par un monochromateur d'excitation, qui filtre la lumière pour une longueur d'onde déterminée, puis atteint la cuve sur une de ses faces. La mesure de la lumière réémise se fait sur une face orthogonale ; celle-ci ne peut provenir ainsi que d'un phénomène de fluorescence. Un monochromateur d'émission couplé à un photomultiplicateur permet d'analyser le spectre d'émission. Le balayage du spectre d'excitation et celui du spectre d'émission pour chaque longueur d'onde d'excitation permet d'obtenir un spectre 2D d'excitation-émission.

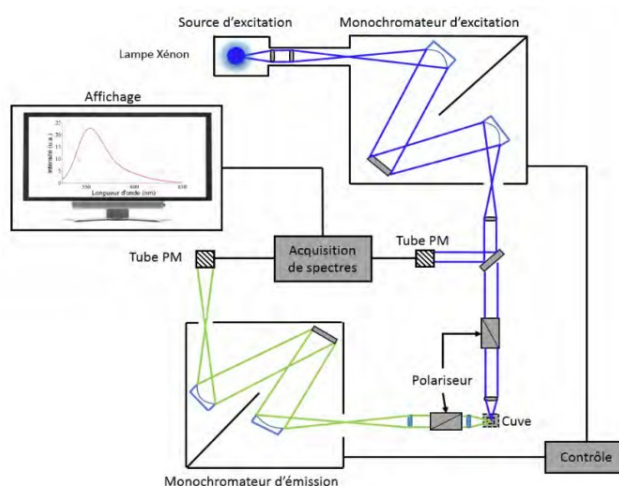
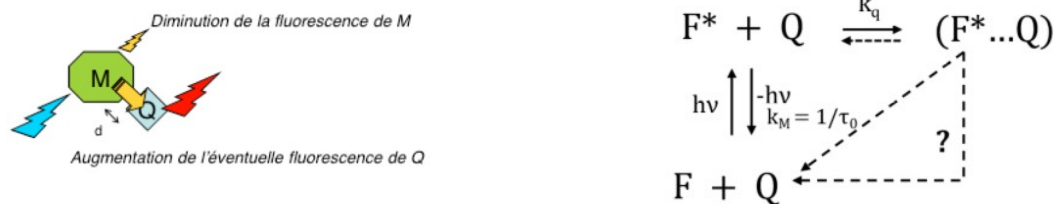


Figure 1.5 – Principe de fonctionnement d'un spectrofluorimètre standard

Quenching de fluorescence. Nous souhaitons exploiter le quenching de fluorescence dans cette expérience. On appelle quenching de fluorescence le processus de diminution de l'intensité de fluorescence d'une solution due à une interaction moléculaire. Le quenching met en jeu l'interaction d'une molécule excitée qu'on appelle F^* (le Fluorochrome) avec une molécule Q (le Quencheur). Lors de l'excitation par une source lumineuse, les molécules Q vont affecter la fluorescence des molécules excitées F^* . Les molécules F^* et Q se comportent comme un complexe ($F^*...Q$) qui aura son propre chemin de désactivation (relaxation de fluorescence) et sa propre durée de vie de fluorescence τ comme le montre le schéma de Valeur and Berberan-Santos (Figure 8.21). C'est un processus qui implique une diminution du rendement quantique Φ d'un fluorochrome ce qui se traduit par un affaiblissement de l'émission de fluorescence. Le rendement quantique de fluorescence rend compte de la compétition entre les phénomènes de désactivation radiatifs et non radiatifs. Il correspond au rapport du nombre de photons émis sur le nombre de photons absorbés par le fluorophore, mais aussi au rapport du nombre de photons émis sur le nombre de molécules excitées. C'est aussi la probabilité pour qu'une molécule excitée émette un photon. Ainsi, la présence d'un quencheur, dans notre cas des acides aminés (Q) entrant en interaction (donc formant un complexe) avec des acides nucléiques (F), doit affecter le spectre de fluorescence.

Le transfert d'énergie entre la molécule F^* et le quencheur provoque la baisse de l'intensité de fluorescence et, en cas de transfert radiatif (transfert d'énergie avec perte d'un photon), on peut assister à un décalage de la longueur d'onde d'émission vers les plus grandes longueurs d'onde (moins grande énergie).

Un des avantages de cette technique est qu'elle est simple à mettre en œuvre et non destructive, mais elle est très longue à mettre en œuvre car l'acquisition de spectres est très lente avec l'appareil dont nous disposons (une acquisition dure plusieurs heures).



8.2.6.2.c. Mesure RMN et Nuclear Overhauser Effect (NOE)

Je décris ici plus brièvement le dispositif de mesure par RMN-NOE car les mesures par résonance magnétique nucléaire sont classiquement utilisées et parce que le NOE s'apparente beaucoup au quenching de fluorescence décrit ci-dessus.

Le principe de la RMN est le suivant : à l'aide d'un aimant puissant, on oriente les spins (moments cinétiques) des noyaux des atomes présents dans le milieu étudié selon l'axe du champ magnétique. Les spins nucléaires sont alors uniformément orientés dans une direction ou dans la direction opposée selon l'axe du champ magnétique. L'application d'une radiofréquence (une onde électromagnétique donc) permet de faire basculer une certaine proportion de spins avec une orientation différente de celle du champ magnétique. La relaxation vers l'orientation première (axe du champ) se fait avec réémission de radiofréquences. C'est cette réémission qui est mesurée ainsi que les temps de relaxation. Ces informations donnent accès au voisinage atomique des atomes ayant subi cette excitation-relaxation.

Lorsque le voisinage des atomes ciblés change, cela se traduit par des changements sur le spectre RMN. Parmi les changements de voisinage atomique possibles, il y a le rapprochement de deux molécules dans le cas de la formation de complexes non-covalents comme celui décrit précédemment dans la partie fluorescence et que nous espérons voir se former entre trinuécléotides et acides aminés correspondants. Lorsqu'un rapprochement moléculaire se fait à une distance suffisamment faible (4-6 angström) et pendant une durée comparable au temps de relaxation, un transfert d'énergie peut se faire entre les espèces moléculaires en présence. Cela se traduit par un élargissement des raies spectrales. C'est l'effet nucléaire overhauser (Nuclear Overhauser Effect).

8.2.6.3. Moyens matériels

Laboratoire et petit matériel. Avec le financement IDEX, j'ai pu acheter tout le petit matériel nécessaire pour mon projet de recherche (pipettes, contenants en plastique et en verre, ..., matériel de protection, mais aussi les produits comme les tampons, les séquences nucléotidiques, les acides aminés ...). Je dispose également d'un spectrophotomètre et j'ai fait l'acquisition d'un pH-mètre.

ITC. Nous avons à Grenoble 2 ITC à notre disposition : un ITC200 au laboratoire IBS (Institut de Biologie Structurale à Grenoble), et un CSC 5300 Nano-ITC au laboratoire des très petites interactions au CNRS, institut Néel à Grenoble. Avec mon étudiante en M2, J. Garcia, nous avons suivi en 2020 une formation détaillée sur les deux instruments et réalisé bon nombre de tests avec les tampons (voir plus bas). J. Garcia, en s'inspirant d'un programme développé par H. Guillou, a également développé un programme en Python pour calculer la ligne de base et intégrer les pics d'enthalpie.

Spectrofluorimétrie. Un Perkin Elmer Luminescence Spectrometer Model LS-50B est mis à ma disposition au laboratoire national des champs magnétiques intenses (LNCMI) du CNRS, à Grenoble, par mon collaborateur Eric Beaugnon.

Ce spectrofluorimètre génère des fichiers au format propriétaire SP3 (binaires). Pour les convertir en fichiers HP5 (ASCII) exploitables, nous utilisons un convertisseur écrit en C, disponible à [SP3, 2013]. Pour lire les fichiers H5, Justine Garcia a développé un programme en Python permettant d'extraire les données d'intérêt.

RMN-NOE. Au laboratoire national des champs magnétiques intenses (LNCMI) du CNRS, à Grenoble, fin 2019, mon collaborateur Eric Beaugnon et le directeur de l'équipe Steffen Krämer ont obtenu du Max Planck Institute un aimant destiné à la RMN. Nous souhaitons utiliser cet instrument qui est cours de montage pour ce projet et pour une thèse (*Nuclear magnetic resonance investigations on paramagnetic relaxation enhancement of ultra-high spin clusters*) que nous avons déposée et qui est financée par l'IDEX.

8.2.6.4. Plan expérimental

Choix des acides aminés et des séquences nucléotidiques. Il n'est pas possible dans un premier temps de mesurer les 1280 combinaisons d'interactions entre les 64 codons et les 20 acides aminés. De plus, il faut que ces séquences obéissent à un certain nombre de critères. J'ai donc choisi de n'en mesurer qu'un certain nombre, répertoriés dans le tableau ci-dessous (figure 8.22).

D'abord, il faut que les codons entre eux aient un nombre d'interactions limitées et tout du moins qu'ils ne s'hybrident pas entre eux (par exemple CGG peut s'hybrider avec lui-même). Dans le tableau, on élimine ainsi tous les couples codons - anti-codons en blanc ; les verts ne peuvent pas s'auto-hybrider trop spécifiquement. Ensuite, j'ai sélectionné un certain nombre de codons-anticodons dont les 2 premières bases (supercodons et super-anticodon) et l'acide aminé correspondant se trouvent dans le code génétique ancestral supposé à 2 lettres (voir section 8.2.2 la figure 8.8). Ces paires acide aminé – anticodon sont repérées dans le tableau avec un ○ rouge. Enfin, j'ai choisi de retenir des codons-anticodons pour acides aminés de types assez différents (la proline, la leucine et l'isoleucine neutres, la sérine alcoolisée, l'histidine, la lysine, et l'arginine basiques, le glutamate acide, la cystéine soufrée) et dont les interactions sont suspectées fortes avec leurs anti-

codons (comme l'arginine).



Figure 8.22. Choix des anticodons.
 Ce tableau montre la correspondance anticodons-codons substitués par un acide aminé, prévue par calcul par Hendry et al [Hendry, 1981] et les anticodons sélectionnés pour les mesures d'affinité anticodons-acides aminés. La correspondance anticodons - acides aminés du code génétique est indiquée en noir. Les zones grisées et noires représentent les meilleures affinités obtenues par calcul par Hendry et al. Les lignes pointillées rouges — correspondent aux codons-anticodons stop. La ligne vert pointillée — correspond au codon initiateur (méthionine). Les triangles oranges Δ représentent les couples anticodons - acides aminés que je souhaite tester. Les ronds rouges O représentent les couples anticodons - acides aminés existants dans le code génétique à 2 lettres.

J'ai ainsi sélectionné 12 anti-codons repérés dans le tableau par une ligne remplie de symboles triangulaires oranges Δ : 5'GGG3' codant pour la proline, 5'AGA3' codant pour la sérine, 5'GAG3', 5'GAA3' et 5'AAC3' codant pour la leucine, 5'AAG3' codant pour la phénylalanine, 5'GTG3'

codant pour l'histidine, 5'TTT3' et 5'TTC3' codant pour la lysine, 5'CTC3' codant pour le glutamate, 5'ACA3' codant pour la cystéine, 5'TCT3' codant pour l'arginine. Tous ces trinucleotides ont été synthétisés par la plateforme Nanobio de l'UFR de chimie de l'Université Grenoble Alpes.

Choix des tampons et préparation des solutions. Toutes les solutions sont préparées à pH 7. Le choix des tampons [Ferreira, 2015] est critique dans l'expérience de microcalorimétrie car il a été montré que certains tampons comme le TRIS classiquement utilisé en biologie et particulièrement adapté aux solutions d'oligonucleotides, peuvent se montrer très problématiques car réagissant entre eux et potentiellement avec les cuves métalliques des microcalorimètres. Avec de tels tampons, il est presque impossible d'obtenir une ligne de base stable et surtout, l'ajout de ligand (injections) provoque des réactions avec le tampon. Ces réactions sont faibles, mais une ITC mesure des variations de température de l'ordre du microkelvin ; les variations de chaleur liées au tampon sont donc catastrophiques. Des tests ont été réalisés en février et début mars 2020 avec un tampon TRIS qui ont confirmé une forte dérive. Nous utiliserons donc du tampon phosphate qui ne provoque pas ce type de dérive.

Les volumes utilisés dans l'ITC et dans les cuves spectro sont de 350 μ l. Les injectas (acides aminés) de l'ITC sont d'environ 40 μ l au total. Les solutions d'oligonucleotides sont concentrées à 30 μ M et les solutions d'acides aminés à 300 μ M. Toutes les solutions sont préparées à partir de lyophilisats (oligos et AAs) resuspendues en solution concentrées dans le tampon et diluées par la suite. Les solutions d'oligonucleotides sont ensuite dosées par spectrométrie d'absorption (loi de Beer-Lambert $A = E \cdot c \cdot l$ avec E le coefficient d'absorption, c la concentration de l'entité à doser, l la longueur du chemin optique) pour connaître leur concentration exacte (essentielle pour l'ITC). La longueur d'onde d'absorption λ_{max} est corrigée par rapport à la fraction de nucleotides A, T, G, C dans chaque oligonucleotide ($\lambda_{max} = (256.7 \times fA) + (265.7 \times fC) + (247.3 \times fG) + (265.1 \times fT)$), de même que le coefficient d'absorption E qui est corrigé par la méthode du plus proche voisin, comme indiqué dans la méthode [ATDBio]. La pureté des solutions d'oligonucleotides est également vérifiée en analysant les rapports A_{260nm}/A_{280nm} .

Expérience en ITC. Pour titrer dans l'ITC les oligonucleotides par des acides aminés et obtenir le niveau d'affinité entre ces espèces chimiques, nous utilisons les oligonucleotides comme cibles et des solutions d'acides aminés comme ligands à injecter. Pour chaque trinucleotide, on souhaite tester 10 interactions avec 10 acides aminés différents.

En pratique, pour optimiser l'expérience, nous avons choisi dans un premier temps de vérifier qu'une interaction est mesurable entre de l'ADN de sperme de saumon et une solution de plusieurs acides aminés. Dans un second temps, nous utiliserons des mélanges de 2 ou 3 cibles nucleotidiques différentes face à un acide aminé unique. Ces triplets de trinucleotides sont choisis tels qu'ils ne peuvent pas s'auto-hybrider (ex : 5'GGG3', 5'AGA3' et 5'GAG3'). En utilisant plusieurs triplets,

mettant en commun un trinuécléotide, nous espérons ainsi optimiser le nombre d'expériences à réaliser. Si une forte interaction est trouvée entre un triplet de trinuécléotides et un acide aminé, alors il est toujours possible de refaire 3 expériences confrontant chaque trinuécléotide du triplet et l'acide aminé en question. J. Garcia a ainsi établi un protocole expérimental optimisant ces triplets de trinuécléotides.

Expérience en spectrofluoroscopie (et RMN). Il s'agit de mettre en évidence le quenching de fluorescence (ou RMN) d'oligonuécléotides du à un acide aminé affine. Pour chaque oligonuécléotide, il faut d'une part établir le spectre de fluorescence de l'oligonuécléotide seul, ensuite en présence d'acide aminé. La différence entre les spectres peut montrer un effet de quenching.

8.2.7. Simulation de proto-métabolismes

Dans la partie théorique de ce projet, on souhaite étudier par simulation la formation de proto-métabolismes, c'est-à-dire de boucles auto-entretenues d'éléments capables de réagir entre eux, et leur auto-organisation spatiale. Notons que plutôt qu'une modélisation au sens de la reproduction à partir d'observations d'un système biochimique, il s'agit davantage de la conception, à partir de concepts et d'hypothèses, d'un système de biochimie artificielle pour de l'expérimentation numérique. L'objectif souhaité est de valider un scénario plausible du passage de l'expérience de Miller à l'obtention des premiers métabolismes auto-entretenus doublée d'une compartimentation de l'espace propice à leur pérennité. Au fil des raffinements du modèle et des simulations, ce scénario prendra en compte des connaissances issues de la biologie et de biochimie.

Je présenterai cette partie sous la forme de défis dans sa mise en œuvre, soulignant deux problèmes majeurs d'ordre scientifique et une difficulté relativement mineure d'ordre technique. Je décrirai dans la foulée le travail préliminaire que nous avons fait avec H. Bouvier, étudiant en M1 informatique pour éliminer ou contourner un certain nombre de ces difficultés.

8.2.7.1. 1^{er} problème : saisir l'essence du système proto-métabolique auto-amplifié

C'est le problème le plus fondamental faisant obstacle dans la réalisation de cette partie du projet : concevoir un système artificiel capable à la fois de représenter ce que l'on souhaite simuler, ici l'émergence d'un proto-métabolisme, sans tomber dans l'excès de réalisme impossible à prendre en compte (ex : une chimie simulée dont ni la simulation, à cause de limites techniques, ni la complétude nécessaire pour représenter un système réaliste, ni l'analyse, à cause de l'immense diversité des composés et de leurs configurations, ne sont possibles), ni dans le travers inverse qui est une symbolisation excessive d'un système qui ne s'apparentera alors plus (plus assez du moins) à l'objet biologique que l'on souhaite étudier, qui ne contiendra de complexité que celle que l'on aura déjà prévue en construisant le système artificiel. Dans le cadre de proto-métabolismes, rappelons que l'on parle de l'émergence de boucles auto-entretenues entre d'une part des effecteurs (les

peptides) et d'autre part des molécules support d'information (acides nucléiques). Dans un système réel, ces molécules adoptent possiblement des conformations particulières présentant certains groupements chimiques, configurations desquelles émergent des activités chimiques incluant de simples appariements avec d'autres molécules, des hydrolyses et un très grand nombre de réactions de combinaison (condensations ...) de molécules entre elles. Le nombre de configurations (de parties de molécules) chimiquement actives est proprement énorme ; le nombre de configurations qui ne le sont pas l'est tout autant. De cet immense diversité moléculaire et diversité de configurations chimiques actives ou non, de nouvelles molécules vont se former et participer avec l'ensemble à la chimie du mélange. Ce que l'on attend, c'est que de ce mélange émerge un cycle qui amplifie tout ou partie de ce mélange lui même.

Réaliser un cycle auto-entretenu avec un modèle numérique est simple. Des systèmes de réécriture (ex : P-Systems) peuvent facilement simuler une auto-amplification d'un milieu chimique. Pour autant, en procédant ainsi, on perd l'essentiel de ce que l'on veut montrer : que des acides nucléiques et leur propriétés (configurations, réactions, ...), couplés avec des peptides et leurs propriétés (configurations, réactions, ...), constituent un système qui au fur et à mesure des réactions fasse émerger une boucle d'amplification de certains acides nucléiques et certains peptides. Comme il n'est pas possible d'intégrer une chimie réaliste, il faut clairement définir la modélisation de chacun des éléments qui joueront un rôle dans ce système artificiel. Les systèmes typogénétiques de Hofstadter et Varetto sont une solution acceptable [Hofstadter, 1979 ; Varetto, 1993 & 1998]. En effet, ces auteurs incluent le fait que les acides nucléiques sont des molécules informatives codant pour des peptides, la traduction étant implicite dans ce modèle, et que les peptides sont les espèces réactives. Les peptides sont formés d'acides aminés vus comme des micro-fonctions (copie, coupe, ...) et ont une conformation codée dans la succession d'acides aminés elle-même. La conformation replie la séquence d'acides aminés/micro-fonctions d'une telle manière que ce peptide constitue un programme qui est exécuté sur la molécule ciblée (celle qu'on lui présente), un acide nucléique. D'un mélange complexe de peptides et d'acides nucléiques, il est effectivement impossible de prévoir à l'avance (de tête) quel devenir ce mélange aura ; on espère une convergence vers une boucle de type métabolique. Presque tous les ingrédients nécessaires à la recherche de l'émergence de proto-métabolismes sont présents.

Varetto n'explorait cependant pas l'émergence de cycle à partir d'une grande combinatoire moléculaire, mais mettait 'seulement' en œuvre des espèces particulières dont il avait identifié qu'elles permettraient l'obtention de petits cycles. Les ordinateurs de l'époque ne lui permettaient en aucun cas de tester cela. Mais plus que ça, l'étude de Varetto se limitait à montrer qu'en utilisant ces 2 types de molécules et une complexité intrinsèque des peptides créée par leur structuration spatiale, il pouvait trouver des peptides-programmes (ma première frappe était « peptides-machines ») qui, associés aux bonnes chaînes nucléotidiques, généraient un cycle. Je trouve ce travail théorique pensé par Hofstadter et mis en pratique par Varetto admirable mais, selon moi, en procédant ainsi par ingénierie, on perd malheureusement une partie de ce que l'on veut montrer : un mélange de composés possédant les propriétés précitées doit, s'il réunit les composés adéquats, se diriger

inexorablement vers la formation de cycles. Le potentiel se trouve dans les propriétés des composés en présence (AN et peptides) mais pour s'exprimer (générer des cycles métaboliques) il faut tout de même réunir une composition-état du cycle ou une composition n'appartenant pas au cycle mais qui permet de le rejoindre. En m'exprimant ainsi, je dévoile plus clairement le fond de ma pensée : les cycles métaboliques à atteindre sont des états asymptotiques d'un système dynamique de grande dimension (chaque dimension étant un composé, l'axe sa concentration) ; ce système dynamique incluant 'tous les possibles' est peuplé d'attracteurs qui sont soit des points fixes, soit des cycles. Une première question est de savoir s'il est probable de partir de conditions initiales menant à un cycle. La deuxième est la robustesse des trajectoires, y compris asymptotiques, d'un tel système dynamique très stochastique (toutes les espèces moléculaires ne se rencontrent pas ; toutes les interactions ne sont pas nécessairement couronnées de succès). Dit autrement : quelle proportion de simulations d'un système comme celui de Varetto mèneront spontanément et de façon robuste à des cycles auto-entretenus.

Il manque deux points importants dans ce paragraphe. D'abord, si ce n'est pas par l'ingénierie que bon nombre des poly-nucléotides et des peptides-programmes associés seront générés dans les simulations que je souhaite réaliser, c'est par formation spontanée, par auto-assemblage. Nous avons discuté de ce point dans la section 8.2.3 : la formation spontanée de petits peptides et petits acides nucléiques est possible avec une dynamique assez lente ; elle peut être accélérée en présence de molécules *templates* ou de matériaux (métaux notamment) favorisant leur condensation. Dans mes simulations, des briques élémentaires (acides aminés ou nucléotides) ont donc la possibilité de s'assembler soit entre eux, soit avec des polymères déjà formés, dans tous les cas avec une dynamique bien plus lente que celle catalysée par les peptides-programmes.

L'autre ajout par rapport au modèle de Varetto est une définition moléculaire du processus permettant la traduction des polynucléotides en peptides. En l'occurrence, l'idée est maintenant de remplacer la traduction implicite que proposait Varetto en une traduction fondée soit sur l'hypothèse stéréochimique, soit sur la présence d'adaptateurs comme des ARN de transfert ou leurs précurseurs reconnaissant les acides aminés par des petites séquences de 3 à 5 nucléotides (voir [Hopfield, 1978 ; Koonin, 2017]). Les polynucléotides favoriseront ainsi la formation de peptides particuliers plutôt que générés au hasard des auto-assemblages, mais leur formation sera aussi conditionnée par la disponibilité des acides aminés présents, ceci pouvant introduire des biais importants dans la formation de certaines espèces plutôt que d'autres.

Notons qu'il est attendu que la composition du milieu à ses débuts, pendant la phase peu catalytique, les peptides étant encore rares et de fonctionnalités limitées, fixe rapidement des conditions initiales particulières guidant le devenir du système. Dans un environnement spatiale large, des conditions initiales très variées peuvent donc *a priori* émerger et générer des devenirs très différents, potentiellement, des cycles proto-métaboliques différents en compétition lorsqu'ils sont dans un voisinage proche. Ceci nous amène au deuxième problème.

8.2.7.2. 2ème problème : réunir les conditions nécessaires pour obtenir une brisure de symétrie spatiale

L'objectif associé avec celui d'obtenir des cycles auto-entretenus est donc d'abord d'obtenir, dans un environnement spatial suffisamment large, tout un bestiaire de comportements asymptotiques locaux, incluant notamment des cycles proto-métaboliques. Ces comportements locaux ne sont pas isolés et sont en compétition pour la ressource : les molécules présentes dans le milieu, produites et consommées par ces micro-systèmes, diffusent dans le voisinage ; elles peuvent faciliter une réaction dans un cycle ou au contraire annihiler une autre ; elles servent donc de moyen de communication entre ces micro-systèmes locaux et peuvent jouer le rôle de l'activateur ou de l'inhibiteur de la réaction spatialement auto-organisée de Turing.

Pour obtenir une instabilité de Turing par exemple, il faut réunir un certain nombre de conditions, en particulier que l'inhibiteur diffuse plus vite que l'activateur. D'autres types d'instabilités (ex : bifurcation de Hopf) peuvent amener à d'autres types de comportements auto-organisés (ex : propagation de front d'onde dans un milieu excitable), surtout dans un système formé de nombreux composants, une riche variété de comportement pouvant alors être obtenue [Liehr, 2013].

Ce système typogénétique présente l'avantage de rassembler l'essentiel des éléments souhaités pour représenter des systèmes proto-métaboliques fondés sur le dialogue entre molécules informatives et molécules effectrices. Cependant, le fait qu'ils soient composés d'une très grande variété de réactifs les rends analytiquement insolubles. Il n'est même pas possible d'explorer l'espace paramétrique de tels modèles. Cette limitation pose clairement la question de la possibilité d'obtention de tels motifs auto-organisés à partir de ce vaste ensemble de réactions chimiques.

Le dernier point (la difficulté technique) que je vais à présent décrire fournit toutefois une alternative numérique à la compréhension analytique des comportements dynamiques d'un tel système : la spatialisation en petits sous-éléments discrets d'espace va *de facto* favoriser l'exploration d'autant de mélanges réactifs qu'il y a d'éléments d'espace. Si au moins un cycle proto-métabolique démarre et s'amplifie dans un élément d'espace, ce cycle devrait se propager par diffusion des composés nécessaires, si ceux-ci dominent les autres composés présents dans le voisinage. Il est probable alors que, s'éloignant de la région dans laquelle le cycle est le plus actif et, rencontrant des conditions (composés) différentes, des cycles différents apparaissent, le cycle initial agissant comme un forçage périodique sur les réactions en cours dans le voisinage.

8.2.7.3. Difficulté technique : gérer l'explosion combinatoire et les difficultés rencontrées pour l'analyse.

Le programme, pour simuler la primo-formation de petits peptides, certains fonctionnels, et de petits polynucléotides, va calculer à partir de concentrations de briques élémentaires et éventuellement de petites molécules déjà formées (dans les deux cas des quantités discrètes représentées dans la

simulation par des nombres entiers, les quantités attendues étant limitées) les réactions d'assemblage possibles et en réaliser certaines. On comprend qu'avant même de considérer les possibles réactions des peptides-programmes sur les polynucléotides ou les possibles traductions de polynucléotides en peptides, on rencontre déjà un problème d'explosion combinatoire (nombre de réactions et nombre de composés très vite gigantesque) et un problème de gestion de la temporalité des réactions dès que les premières réactions catalysées par les peptides ou même les réactions de traduction commenceront, les réactions d'auto-assemblage étant bien plus lentes dans ce modèle (comme dans la réalité). Un stockage bien pensé des données de simulation est également nécessaire puisqu'il va s'agir de stocker une très grande diversité de molécules (de tailles et compositions différentes) de façon à d'une part y accéder rapidement par type et taille (pour énumérer et réaliser les réactions possibles), mais aussi pour les cataloguer à des fins d'analyse (catalogage par masse et par composition), le but étant la détection de motifs chimiques spatiaux et/ou temporels.

La difficulté est très grande mais peut-être résolue par des approximations acceptables que nous avons testées H. Bouvier et moi en utilisant un toy-model de chimie artificielle bien plus simple que les typogénétiques de Hofstadter et Varetto. D'abord, on simulera des petits volumes contenant un nombre limité de molécules. Il faut toutefois que ce nombre soit suffisamment élevé pour permettre un mélange riche susceptible de générer des cycles auto-entretenus. La quantité nécessaire n'est évidemment pas connue, mais il est tout à fait envisageable d'entretenir plusieurs milliers à dizaines de milliers de molécules par élément de volume.

Le problème de la diversité et de la temporalité des réactions est plus complexe. Certaines réactions comme l'assemblage sont peu probables pour chaque réaction mais nombreuses tandis que les réactions catalysées sont probables (au sens de leur efficacité) mais peu nombreuses tant que le nombre de molécules complexes capables de catalyse est faible. Cela signifie qu'une simulation risque de simuler un très grand nombre de réactions d'assemblage au début, mais rapidement, alors que les molécules complexes se formeront, la grande majorité des réactions seront des réactions catalytiques, les auto-assemblages devenant rares en proportion. Il faut pour permettre ce régime à deux vitesses choisir un algorithme de dynamique réactionnelle adapté. Le meilleur compromis est l'utilisation des algorithmes de Gillespie en temps et espace discret (tau-leaping) [Gillespie, 1976 & 1977 ; Hattne, 2005]. Le principe est d'évaluer, à partir de l'équation maître du système réactionnel défini par les constantes de vitesse des différentes réactions (la diffusion étant comprise comme une réaction de disparition d'un élément d'espace et de synthèse dans un élément voisin [Hattne, 2005]), combien de réactions de chaque type vont se produire dans un laps de temps discret. Cependant, étant donné la possible diversité des composés présents, il ne sera pas envisageable d'évaluer la totalité des réactions possibles. La solution sera donc une approximation réalisant un tirage hypergéométrique dépendant des proportions naturelles des composés de chaque espèce, ou uniforme sur l'ensemble des molécules, à chaque pas de temps parmi les espèces moléculaires présentes dans le volume.

Du point de vue technique, le programme sera pensé pour être facilement porté sous CUDA et bénéficier de l'accélération GPU. Cela permettra notamment la parallélisation des tâches sur

l'espace réactionnel.

Le problème de la détection est également très complexe. On souhaite détecter l'émergence de cycles réactionnels, mais on ne sait pas à l'avance quels composés seront impliqués dans de tels cycles. Le nombre et surtout la diversité des espèces qui seront créées dans chaque élément de volume rend la tâche difficile. Enregistrer toutes les molécules ayant existé au cours de la simulation avec leur instant d'apparition et leur instant de disparition, de façon à permettre une analyse *a posteriori*, ne paraît pas une solution viable à cause de l'énorme volume de données requis pour cela. Dans un premier temps en tout cas, il faut pouvoir détecter des indices de ces cycles et, en cas de détection, commencer l'enregistrement des instances de molécules pour connaître la composition du cycle.

Pour cela, nous avons imaginé une solution qui consiste à ne s'intéresser au début qu'à l'histogramme de la masse (nombre d'éléments) des molécules. Il est attendu qu'au début de la réaction, de petites molécules se forment, puis, au fur et à mesure que la simulation avance, que ces petites molécules se rallongent progressivement par ajouts d'éléments (auto-assemblage). Un histogramme du processus d'auto-assemblage au cours du temps prendrait la forme d'une exponentielle décroissante (loi de probabilité exponentielle). Les processus catalytiques, eux, vont générer des composés particuliers ; à la différence de l'auto-assemblage, ils ne fonctionnent pas au hasard. Cependant, si aucun cycle entretenu n'apparaît, ces processus ne contribueront sur le moyen terme qu'à former un bruit dans l'histogramme. En effet, si un composé A est en quantité faible (car formé au hasard) et génère un composé B, alors il est probable qu'une certaine accumulation de ce dernier se fasse. Cependant, cette accumulation ne devrait d'une part pas perdurer car la source de B, le composé A, si elle n'est pas auto-amplifiée (dans un cycle), risque d'être détruite ou modifiée par un autre processus (auto-assemblage ou processus catalytique). D'autre part, il n'y a pas de raison pour que la formation de B surpasse en quantités discernables dans le temps celle des autres composés formés aussi de façon catalytique. Un histogramme des masses intégrant ces composés formés de façon catalytique mais n'appartenant pas à un cycle devrait donc suivre une loi exponentielle bruitée. En revanche, des composés amplifiés dans un cycle devraient apparaître sur-représentés dans un histogramme des masses, d'autant qu'il est attendu qu'un certain nombre de ces composés seront vraisemblablement complexes, apparaissant ainsi sous la forme de pics bien distincts au-delà des masses faibles.

Une autre analyse gros grain à mener en parallèle est celle de la composition par tranches de masses moléculaires : pour une masse moléculaire donnée, on détaille la composition en acides aminés et nucléotides. Autrement dit, on résout en composition chaque classe de l'histogramme de masses moléculaires. En laboratoire, cette double résolution masse-composition correspond par exemple à réaliser un MALDI-TOF/TOF³¹ (en se fondant sur la masse moléculaire et leur fragmentation sub-moléculaire uniquement) ou à une électrophorèse bidimensionnelle (dont la séparation est fondée sur le point isoélectrique pHi et la masse). Cette technique exploite des caractéristiques physico-chimiques des éléments composants les molécules comme la masse et la charge. Dans un modèle de

31 MALDI-TOF : Matrix Assisted Laser Desorption/Ionization-Time Of Flight.

chimie artificielle, les composés n'ont pas nécessairement de propriétés physico-chimiques comme le pHi ou une charge. Une possibilité simple est de leur en attribuer une à des fins d'analyse.

8.2.7.4. État des travaux de simulation numérique.

Les tests préliminaires que nous avons réalisés avec H. Bouvier sur un système de chimie artificielle simple (pas d'encodage génétique, juste des réactions chimiques catalysées par des combinaisons particulières d'assemblages moléculaires) permettaient la formation de composés auto-amplifiés. Nous avons pu tester un système de stockage assez efficace d'un écosystème chimique dans des éléments de volume cubiques, un moteur réactionnel stochastique, un système d'analyse fondé sur l'histogramme des masses moléculaires. Le programme est écrit en C++11 et compatible NVidia CUDA.

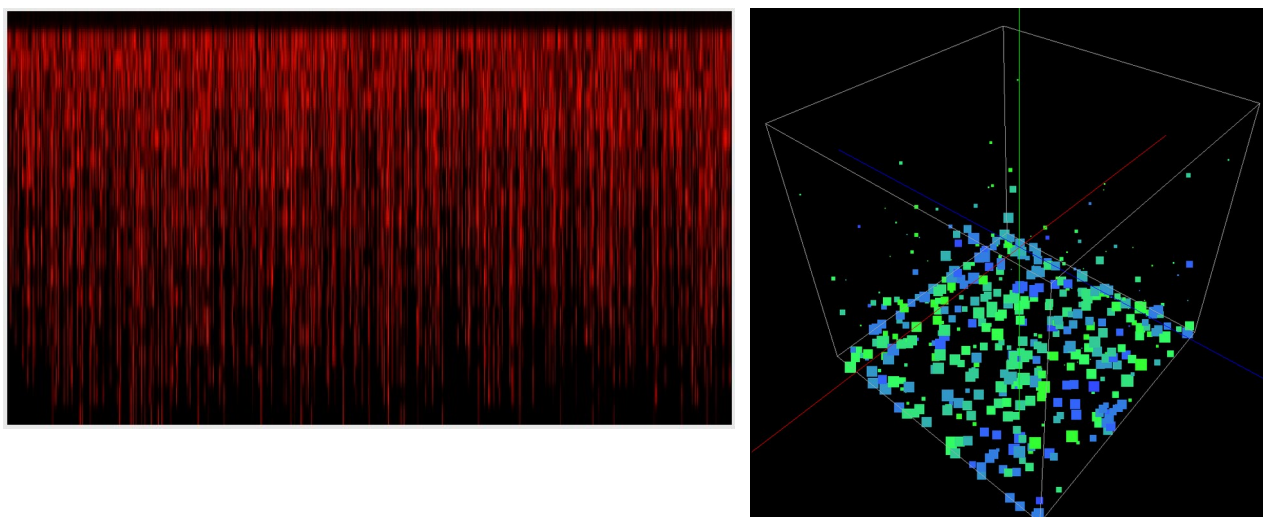


Figure 8.23. Simulation d'écosystème de vie artificielle. A gauche, chaque ligne verticale représente le génome d'une créature, chaque "barre" correspondant à un gène, la valeur de son octet codant pour un niveau de rouge. A droite, une simulation en cours dans laquelle on a encodé la masse des créatures avec la taille des points et le nombre de gènes avec un code couleur. Beaucoup de créatures ont sédimenté, leurs éventuels organes de déplacement (nageoires) ne suffisant pas à les propulser hors du fond de l'environnement. D'autres au contraire, plus légères et dotées de moyens de propulsion, nagent librement dans ce milieu.

De plus, un travail de modélisation antérieur d'un écosystème de créatures artificielles évolutives dotées d'une génétique et de fonctions d'interaction avec leur environnement a été réalisé en 2017 par un étudiant en M2, C. Hégé (voir figure 3.3, chapitre 3 et figure 8.23). Les créatures possèdent 0 à plusieurs gènes (nombre indéterminé) codant sur un octet (8 bits) un organe parmi plusieurs possibles (codage sur 2 bits : 0b00 aucun organe, 0b01 nageoire, 0b10 bouche, 0b11 senseur de créatures) et leurs positions respectives sur une des 6 faces de la créature, un cube tridimensionnel (codage sur 3 bits) ; les 3 bits restants sont considérés comme un potentiel évolutif. Les créatures possèdent en outre un gène codant pour leurs propriétés de reproduction par division, à savoir une masse minimale et une masse maximale. A chaque division fondée sur une probabilité de division augmentant à mesure que la masse de la créature dépasse de moitié la masse moyenne de la

créature, ces créatures se répliquent et le génome des créatures filles peut subir des mutations (mutation ponctuelles sur des bits de chaque gène : délétion, addition, décalage de bits ; mutations globales de duplication d'un gène ou de délétion d'un gène), mutations dont l'occurrence est également déterminée par un tirage probabiliste (les mutation de duplication ou de délétion étant 8 fois moins probables que les mutations ponctuelles, car impliquant 8 bits). Ce travail (non clos) visait à faire suite à celui que nous avons réalisé P. Ballet, O. Bastien et moi, travail publié en 2017 [Glade, 2017] : l'étude de la biodiversité et de la survie différentielle d'espèces dans un environnement stratifié par l'action d'un facteur brisant la symétrie spatiale, à savoir la sédimentation des créatures.

Ces simulations numériques de vie artificielle sont en réalité assez proches dans leur forme de celles fondées sur des typogénétiques que je souhaite mettre en œuvre, d'abord parce qu'elles doivent gérer une possible explosion combinatoire, en tout cas une très grande diversité d'espèces et d'individus (agents), mais également elles posent le problème de l'analyse de telles simulations : comment rendre intelligibles ces mondes et créatures virtuels de façon à comprendre leur évolution et les facteurs qui la contrôlent ? Dans le cas de ces simulations, nous utilisons des caractéristiques comme la masse des créatures ou leurs vitesses de déplacement, ou encore la régularité de leurs trajectoires, comme variables à suivre. Cependant, cela ne permettait pas d'étudier la biodiversité dans un tel système, problème donc identique à celui posé par l'explosion de diversité attendue dans les systèmes de typogénétique. L'autre point commun concerne l'idée d'une émergence d'écosystème (écosystème de vie artificielle ou écosystème de métabolisme proto-biotique) et sa stratification dans un espace soumis à un facteur comme la gravité ou l'émergence de microcosmes auto-organisés, d'écosystèmes locaux.

Je souhaite maintenant capitaliser ces travaux réalisés par des étudiants en M1 et M2 en les reprenant, en les développant, et commençant leur étude systématique : effet des paramètres de mutation, de la sédimentation, de l'agitation thermique, ..., en commençant par le système de vie artificielle, plus contrôlable et plus avancé. Les développements nécessaires en vie artificielle, en particulier le développement de méthodes d'analyse inspirées des méthodes existantes en biologie peuvent s'ajouter à un tel sujet de recherche sur les origines du vivant pour justifier une thèse.

8.2.7.5. Impact - perspectives

Ces travaux sont théoriques, clairement non appliqués, et poser la question du fameux « impact sociétal » ou plus simplement du « à quoi ça sert » paraît incongru. A cette dernière question, je répondrais volontiers ce que qu'a répondu à ma place Jean-Pierre Mazat alors que quelqu'un me posait cette même question après un séminaire que j'avais donné il y a une quinzaine d'années lors d'un des ateliers Epigénétique organisés par F. Képès à Evry : « parce que c'est intéressant ! ». Car oui, je crois que c'est le rôle principal de la recherche que de produire de la connaissance, dans ce cas-ci sur les systèmes biologiques et la biologie à ses origines. Je crois en effet aussi, comme j'en ai déjà discuté dans la partie introductive, que la biologie théorique, y compris les approches

d'expérimentation numériques en vie artificielle, servent à composer des images mentales des systèmes naturels et aident à les appréhender.

Je vais cependant faire l'effort de proposer certaines perspectives à ce travail.

En premier lieu, il y a le challenge actuel de la recherche de la vie ailleurs, dans le système solaire évidemment, mais également ailleurs dans l'univers. S'il paraît présomptueux d'espérer trouver la vie ailleurs dans l'univers à part peut-être dans le système solaire, plus accessible, toute démarche scientifique sur les origines du vivant et sur la vie actuelle est utile pour d'une part identifier ou du moins suspecter quelles conditions peuvent favoriser l'émergence de la vie, ou au contraire l'empêcher, mais aussi pour suggérer des marqueurs de la vie ou de ses prémisses, comme les molécules prébiotiques que l'on cherche sur les planètes et lunes du système solaire, dans les météorites, dans les comètes, mais également dans l'espace interplanétaire, les disques protoplanétaires nouvellement observés ou dans les nuages moléculaires observables dans la voie lactée. Il s'agit de proposer des possibilités de signatures du vivant.

La présence d'oxygène suffit-elle ? Certainement pas ! Le développement de la vie est-il inéluctable dès lors qu'on dispose d'une soupe prébiotique et de chaleur comme on peut imaginer la trouver dans les océans sub-glaciaires d'Encelade ou d'Europe ? C'est peu probable ! Récemment, il semble qu'on ait trouvé de la phosphine sur Vénus [Greaves, 2020] ; des sulfures (molécules absorbant les UVs) sur Vénus également [Francès-Monerris, 2022] ; des dégagements ponctuels de méthane depuis le sol ont été détectés sur Mars [Webster, 2015]. Ces annonces font systématiquement sensation tant l'attente est forte et l'espoir présent de trouver de la vie ailleurs que sur Terre. Pourtant, y a-t-il une alternative abiotique à la formation de ces molécules assez emblématiques de la présence de vie ? Oui ! Jusque à présent, bien que souvent encore débattues, des explications fondées sur la géologie, l'hydrologie ou la chimie atmosphérique ont toujours été trouvées pour expliquer la présence de telles molécules (voir par exemple [Francès-Monerris, 2022]).

Mais rechercher des signatures du vivant, c'est peut être aussi trouver des brisures de symétrie. La plus évidente : l'énantiométrie que le vivant a fait balancer tantôt d'un côté pour les acides aminés (forme L dans le vivant), tantôt d'un autre pour les sucres (formes D dans le vivant), une brisure de symétrie aisée à mettre en évidence avec une mesure de déviation de la lumière polarisée ... sur Terre ! Doit-on reproduire un équivalent de l'énantiométrie dans les simulations numériques, en permettant aux molécules numériques d'adopter une conformation chirale plutôt qu'une autre, conformation asymétrique qui pourrait être mise en évidence par une mesure globale dans la simulation comme on ferait une mesure de polarité sous un polarimètre ? Ce serait une solution élégante et probablement efficace. Mais la recherche de brisure de symétrie est-elle applicable aussi dans la pratique ailleurs que sur Terre ? Avec des sondes prélevant des échantillons sur Mars ou tout autre corps du système solaire, c'est facile à mettre en œuvre. J'avais cependant imaginé qu'il serait possible de faire de même en se contentant de regarder le spectre des atmosphères des corps célestes en lumière polarisée. La réponse que m'a donnée Xavier Delfosse (IPAG, UGA / CNRS) lors d'un séminaire du CDP COOL (voir plus bas) est sans appel : « ce n'est pas possible actuellement ! On ne sait même pas détecter la vie sur Terre en observant la lumière cendrée de la

Lune (illumination faible de la Lune produite par l'éclairage de la Terre) en lumière polarisée ! » Cela m'a fait prendre conscience d'une limitation importante des mesures possibles pouvant être faites à distance : les signaux (lumineux) s'atténuant très vite, il est très complexe de faire ressortir des informations sur des molécules complexes dans des spectres de corps célestes distants. Comme le disait un astrochimiste plaisantant à moitié seulement lors du même séminaire : « l'hydrogène moléculaire, le cyanure³², ce sont déjà des molécules complexes pour l'astrochimie ! »

Les rythmes sont aussi une éventualité du vivant à surveiller. Les organismes, les écosystèmes, forcés par les saisons souvent, mais aussi par d'autres facteurs, répondent de façon rythmique en augmentant leur biomasse, en changeant de composition (les arbres changent de couleur en automne-hiver), ou en affectant (consommation ou production) différemment leur environnement. C'est cela qui a provoqué un certain émoi concernant la découverte d'émission cycliques de méthane par le sol martien [Chang, 2018] (émissions expliquées maintenant par des processus géologiques). C'est également une question de rythmes, précisément de changements saisonniers de turbidité de l'atmosphère vénusienne par des aérosols dont l'origine potentielle fut attribuée à des micro-organismes dans cet article somme toute sérieux publié dans *Astrobiology* [Limaye, 2018]. La démarche scientifique standard qui suit ce type d'annonces, est évidemment de chercher une explication abiotique à ces phénomènes singuliers. Comprendre que des processus géologiques, hydrologiques ou atmosphériques puissent les générer, et en quelle proportion, est indispensable ! La mise en évidence de rythmes inattendus est toutefois indice de présence de vie.

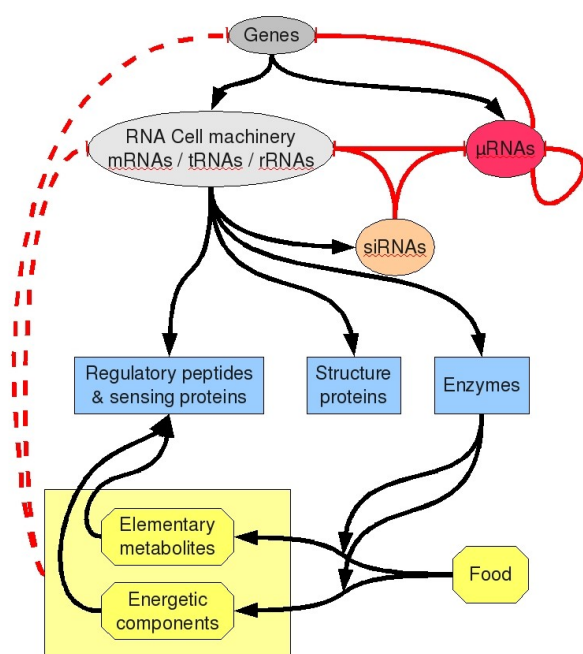


Figure 8.24. Schéma des processus de régulation cellulaire. Les microARNs (en rouge) exercent une régulation négative en interagissant (hybridation partielle) avec des ARN messagers. De plus, une régulation contextuelle potentielle pourrait être exercée par les petits métabolites produits par le métabolisme (acides aminés, pyruvate ...) ou les nucléotides, énergétiques notamment (ATP, GTP), sur l'ADN ou la machinerie cellulaire incluant ARNm, ARNr, ARNt.

Dans les deux cas, la chiralité ou les rythmes, c'est une forme d'ordre qui est identifiée. C'est par cet ordre que nous suspectons le vivant. C'est aussi parce que nous savons mesurer l'ordre que nous

32 De forme chimique HCN (avec une triple liaison entre le carbone et l'azote).

nous y attachons. Y a t'il d'autres choses qui puissent être identifiées ? Je pense qu'un travail théorique et expérimental fait sur la chimie prébiotique, sur les métabolismes prébiotiques et la systémique prébiotique, mais aussi sur le fonctionnement du vivant actuel comme nous en discuterons dans le chapitre suivant, parce qu'il force à réfléchir à la nature de la matière vivante et à ce qui la distingue de la matière inerte, ne peut que contribuer à suggérer de possibles signatures du vivant ... si tant est que cet objectif reste bien présent en tête de celui qui réfléchi à l'aide de tels systèmes artificiels. Une vue selon un prisme différent et possiblement une meilleure compréhension du vivant actuel ! C'est l'autre perspective qui vient naturellement. Les fondations du vivant actuel sont celles d'un organisme primordial vraisemblablement très rudimentaire, très modulaire, doté de mécanismes internes de réplication, de structuration interne, de production d'énergie, bien plus simples que ceux que l'on associe actuellement à ces tâches dans les cellules modernes, eucaryotes comme procaryotes. Je développerai particulièrement ce point dans le chapitre suivant, mais on peut déjà poser la question de l'influence, voire du rôle non trivial que jouent les petits métabolites, en particulier les nucléotides et toutes petites séquences d'ARN comme les micro ARNs (miRNAs) dans le fonctionnement normal ou pathologique des cellules. En 2008 déjà, dans un projet (IXXI, Projet *DynamiRs*, 2008, 5000€) et dans un article [Demongeot, 2009] j'avais souligné ce possible rôle fondamental des miRNAs et des petits métabolites (voir figure 8.24).

L'idée est proche de celle de l'interférence à ARN mais doublée d'une logique dans la régulation héritée des âges prébiotiques, protobiotiques et des origines du vivant. Cette idée que je vais illustrer d'abord avec un métabolite énergétique, le GTP, est la suivante : le GTP, Guanosine Tri-Phosphate, est une petite molécule qui fournit l'énergie d'assemblage-désassemblage des microtubules, éléments fibrillaires essentiels du cytosquelette. Le GTP se forme dans la cellule par régénération du GDP (Guanosine Di-Phosphate) à partir de l'ATP (Adénosine Tri-Phosphate) formé principalement dans la mitochondrie grâce au fonctionnement de l'ATP synthase. Mais le GTP est aussi un nucléotide ! c'est un des 4 constituants élémentaires (ATP, TTP, GTP, CTP) à partir desquels l'ADN est fabriqué par les ADN polymérase dans les cellules. La cellule contient une quantité importante de GTP estimée à $468 \pm 224 \mu\text{M}$ (soit environ 100 à 150 fois plus que pour l'ATP) [Traut, 1994]³³. Lors d'un usage intensif ou au contraire de l'arrêt d'une fonction consommatrice de GTP, par exemple lors d'une restructuration du cytosquelette (chaque micromètre de microtubule contient 1625 molécules de tubuline, chacune consommant un GTP pour s'assembler à la vitesse d'environ $1 \mu\text{m}/\text{min}$) ou d'une duplication de l'ADN, quelle influence aura cette consommation ou au contraire absence de consommation de la part d'une fonction cellulaire sur les autres ? Cette idée est quelque peu supportée par des travaux s'intéressant aux variations des quantités de métabolites intra-cellulaires, mesures difficiles s'il en est (par exemple, pour le GTP, voir [Bianchi-Smiraglia, 2017]). Ainsi, il a été montré que les quantités de GTP intra-cellulaire influent sur la voie que vont suivre les cellules entre l'apoptose et une différenciation [Meshkini, 2011].

33 [Traut, 1994] donne ces concentrations \pm écart type (en μM): ATP, $3,152 \pm 1,698$; GTP, 468 ± 224 ; UTP, 567 ± 460 et CTP, 278 ± 242 . Pour les désoxyribonucléotides triphosphates (dNTP), les concentrations dans des cellules en division sont : dATP, 24 ± 22 ; dGTP, 5.2 ± 4.5 ; dCTP, 29 ± 19 , dTTP 37 ± 30 et dUTP environ $0.2 \mu\text{M}$.

De façon générale, ce qu'on questionne ici, c'est donc le fait qu'une grande fonction cellulaire modifiant les quantités d'un nucléotide ou d'un petit métabolite, puisse affecter d'autres grandes fonctions cellulaires par son intermédiaire. La question jointe, c'est dans quelle mesure ceci peut être vu comme une forme de régulation de contexte : une fonction majeure en régule une ou plusieurs autres via des petits métabolites.

Introduire cette notion de compétition pour la ressource entre les fonctions cellulaires doit être pensé dans un contexte d'évolution car certains systèmes, exploitant des ressources utilisées par d'autres systèmes déjà en place, ont été retenus par l'évolution. Les microtubules dont les constituants élémentaires, les molécules de tubuline (ou leurs variantes bactériennes FtZ), sont des GTPases, sont vraisemblablement très récents dans l'évolution par rapport à l'ADN. Ils se sont mis en place en tant qu'agents structurants dans les cellules, dans un contexte où le GTP existait déjà et subissait des variations de quantités liées aux rythmes des changements d'état des cellules comme lors de la division cellulaire. Mais en pensant le partage de ressources intracellulaires comme une balance régulatrice, on peut alors aussi penser que c'est cette balance, cette régulation et ses rythmes, ses phases de la vie cellulaire, associés, qui ont été sélectionnés. Cette question, je souhaite l'approfondir dans les années à venir.

En plus du partage de ressources entre fonctions, les petits métabolites et petites séquences d'ARN ont un autre effet potentiel : s'ils interagissent avec d'autres éléments cellulaires comme les ARN messagers, les ARN de transfert, ..., alors il est possible aussi qu'ils les régulent de façon assez globale ; une régulation de contexte en quelle que sorte. Par exemple, une quantité importante de GTP (ou d'ATP) dans une cellule la place dans un état énergétique important (au sens où beaucoup d'énergie chimique est disponible et peut être immédiatement mise à la disposition de nombreux processus actifs), mais cette présence forte d'un nucléotide dans le milieu cellulaire pourrait provoquer une forme d'inhibition : c'est un nucléotide qui peut s'apparier de façon très transitoire avec son complémentaire (CTP pour le GTP, TTP pour l'ATP). L'appariement entre paires de nucléotides est le mécanisme par lequel les brins d'acides nucléiques s'hybrident de façon complémentaire ; c'est ce même mécanisme passif qui permet la reconnaissance des triplets de nucléotides codant pour un acide aminé, les codons des ARN messagers, par leurs complémentaires, les anti-codons, des triplets de ribonucléotides présentés par les ARN de transfert. L'affinité de cette hybridation de 3 contre 3 ribonucléotides (codon / anti-codon) est renforcée dans les cellules actuelles par le complexe ribosomal qui accueille à la fois l'ARN messager en cours de traduction ; ce complexe favorise la présentation de la boucle anti-codon de l'ARNt face à l'ARNm et augmente la stabilité de leur appariement. La durée de vie moyenne d'un appariement de deux séquences complémentaires isolées d'acides nucléiques sera (en première approximation, d'autres facteurs pouvant intervenir également, comme la structure de ces séquences) d'autant plus courte que l'une au moins des séquences sera courte. Néanmoins, même furtif, un appariement aura lieu, sollicitant les deux protagonistes, gênant par là même les actions dans lesquelles ils pouvaient être impliqués. Rappelons qu'il est aussi évident que la fréquence naturelle à laquelle pourraient se produire de tels appariements est directement liée à la concentration respective de ces protagonistes ; les équilibres biochimiques sont déplacés par la simple loi d'action des masses. Dès lors, une petite séquence

d'acides nucléiques présente en quantité importante à proximité d'un ARN messenger en cours de traduction pourrait bien gêner la dynamique de sa traduction, introduisant des pauses dans l'extension peptidique en cours et éventuellement son arrêt prématuré. C'est en partie cette idée de compétition qu'introduisait la possibilité d'une interférence à ARN. C'est aussi ce type de processus que pourraient bien causer les micro-ARNs (miRNAs). Ces séquences d'une vingtaine de paires de bases s'apparient par l'intermédiaire de moins d'une dizaine de bases (typiquement 8) avec les ARNm présents dans le milieu cellulaire. Aux miRNAs est associé un complexe protéique nommé RISC (pour RNA Induced Silencing Complex) qui hydrolyse les séquences ARNm reconnues. Mais avant même de considérer ce système évolué, il reste que les miRNAs, simplement en s'appariant avec les ARNm peuvent jouer ce rôle de freins d'une dynamique, ici la traduction. De plus, les miRNAs étant donné la taille très réduite de leur séquence d'appariement, ont finalement des cibles assez peu spécifiques. C'est le fait qu'un miRNA donné pouvait potentiellement reconnaître des centaines, voire des milliers de cibles ARNm différentes qui faisait dire à [Bartel, 2004] que les miRNAs se comportaient davantage comme des micro-managers de l'expression génique en exerçant une régulation de contexte, une régulation touchant des centaines de cibles à la fois, des cibles différentes, codant pour des protéines différentes, mais des cibles potentiellement associées dans un état fonctionnel particulier de la cellule (par exemple, lors de la division cellulaire, c'est un très grand nombre de processus différents qui se mettent à opérer ; c'est ceci que je désigne par « état fonctionnel de la cellule ») ou dans des grandes fonctions. L'idée là, c'est que même si ces petites séquences semblent avoir une spécificité de ciblage limitée, la régulation large qu'elles provoquent reste globalement spécifique.

Il est tentant d'étendre une telle idée aux nucléotides isolés, leur niveau (concentration) dans la cellule agirait ainsi, par l'intermédiaire d'une pression compétitive, comme un curseur affectant potentiellement des pans entiers du fonctionnement cellulaire (ce que semblerait soutenir les observations de [Meshkini, 2011]). Dans ce chapitre, l'hypothèse stéréochimique directe (reconnaissance acides aminés – anti-codons) ou indirecte (reconnaissance acides aminés – petites séquences de nucléotides correspondant aux sites de fixation des acides aminés sur les ARNt primordiaux) est celle qui est mise en avant pour soutenir à la fois l'encodage et la traduction primordiale. Mais si cela s'est passé il y a des milliards d'années dans des proto-cellules, il est fort possible que ceci persiste dans le vivant actuel. Que les acides aminés puissent interagir avec des séquences nucléotidiques, il n'y a aucun doute là dessus. En revanche, si ce que soutient l'hypothèse stéréochimique est correct, à savoir un lien stéréochimique entre acides aminés et code génétique, alors de la même manière que je l'ai dit précédemment pour les nucléotides libres, les niveaux (concentrations) d'acides aminés dans les cellules pourraient bien, de la même manière, exercer une régulation de contexte. Mieux encore, le différentiel de niveaux d'acides aminés pourrait de cette manière très directe moduler finement l'activité de la cellule. Ces niveaux d'acides aminés sont liés directement aux quantités et types de nutriments entrant dans les cellules. La cellule vivante disposerait alors d'un système de régulation gros grain mis en place très tôt, dès l'apport de nutriments, avant qu'une régulation plus fine, plus spécifique, reposant sur la régulation génétique classique (s'il en est) que l'on connaît, ne se mette en place.

J'ai présenté là, dans la conclusion de ce chapitre, un ensemble de propositions fortes sur des mécanismes actuels éventuellement hérités d'autres, fondamentaux, remontant aux premiers âges du vivant. Ce qui demande à être établi, c'est la nature spécifique de leur action de régulation. Je vais à présent, changer légèrement d'échelle et passer dans le chapitre 8.3 à certains aspects des voies de communication intracellulaire et des réseaux d'interaction intracellulaires. Au cours de ce nouveau chapitre, je discuterai à nouveau des reliques du passé que pourraient constituer certains modes de communication intracellulaire, mais je traiterai surtout de la question de l'héritage des fonctions biologiques et de la trajectoire évolutive des réseaux d'interaction.

8.3. Réseaux, voies de communication biologiques et traces du passé

Nous venons de passer un moment à imaginer ce qu'ont pu être les prémisses de la vie sur Terre : l'apparition d'un code génétique fonctionnel reliant la mémoire à la fonction, et la co-émergence de flux métaboliques et de boucles autopoïétiques. Ça, c'était il y a 4 milliards d'années environ ! Depuis, ces boucles fragiles, ces flux mal définis se sont renforcés, ont évolué, se sont dupliqués et différenciés, *etc.* Mais reste-t-il des traces de ces organisations primordiales dans les cellules actuelles ? Peut-on retracer, au moins de façon théorique, par exemple en utilisant des abstractions de la réalité biologique, comme des réseaux Booléens, les trajectoires évolutives que ces organisations ont suivies ? Certaines évolutions sont-elles interdites, leurs trajectoires conduites par des contraintes inhérentes au tissu complexe et fragile des relations qui se sont mises en place ? Et pour revenir au sujet de la machine, l'évolution des organismes va-t-elle dans le sens d'une plus grande structuration, plus forte dans ses relations, au point qu'ils puissent s'apparenter peu à peu à des machines complexes ? Le rare, le faible, l'amorphe qui prédominait aux origines disparaît-il à mesure que les organismes évoluent ou bien est-il maintenu comme une garantie pour les systèmes vivants de continuer à s'adapter aux conditions environnementales rencontrées et à évoluer ? En fin de compte, qu'est-ce qu'un organisme vivant au juste ? Voilà encore des questions qui m'intéressent et qu'il me semble possible d'explorer à l'aide d'outils et concepts de la biologie théorique.

Dans ce chapitre, qui constitue la suite logique du paragraphe précédent et qui présente un certain déroulement du vivant que j'ai choisi d'explorer, je vais adopter un positionnement davantage holistique que mécaniste pour proposer une vue particulière du fonctionnement des organismes et de leur évolution. J'utiliserai d'abord comme support, comme abstraction de cette réalité biologique qui n'est pas passée par magie de l'inerte à la vie actuelle qui nous apparaît déchiffrable (en grande partie à tort probablement), les réseaux Booléens. Les réseaux Booléens, ces outils théoriques, parfaits représentants de la machine tant ils sont ordonnés, structurés, paramétrés, pour obtenir un comportement déterministe, peuvent être *assouplis* pour peut-être alors ressembler aux machines molles de von Neumann. Cependant, pour cet exercice, après avoir dressé un tableau général de la suite de mes réflexions sur les organismes et leur évolution, telle que je la propose, et dès lors que je commencerai à rentrer dans le vif du sujet, présentant tour à tour les objets biologiques et théoriques servant de support à cette réflexion, à savoir les réseaux Booléens et le cytosquelette de

microtubules, je ferai un travelling inverse de celui que j'ai fait dans le chapitre sur les origines du vivant. Pour les origines du vivant, j'allais de l'inerte, du rare, du faible et de l'amorphe vers la structuration et le fonctionnement. Cette fois je vais partir de la machine établie, le réseau structuré, le cytosquelette ordonné et au fonctionnement attendu. De là, je glisserai progressivement vers une vision différente de ces systèmes, proposant d'utiliser les réseaux par larges ensembles plutôt qu'individuellement, de voir les structures ordonnées de la cellule comme des biais de communication intracellulaire, introduisant ainsi une vue différente, plus incertaine de la structure et du fonctionnement cellulaire.

8.3.1. Les organismes vivants, héritages complexes

Qu'est ce qu'un organisme, donc ? Un organisme est la complexe combinaison d'un plan de formation (génétique et partiellement métabolique et structural, la cellule initiale d'un organisme ayant elle-même sa propre configuration initiale - un ensemble de molécules, de niveaux métaboliques, *etc* - héritée de l'organisme dont elle provient), d'un développement embryologique (pour les organismes multi-cellulaires) et d'un maintien homéostatique particulier, contraint par ce plan et par la configuration courante de la cellule (les quantités de composés qu'elle contient à un moment donné, sa forme, l'influence de son environnement local, *etc.*). Si on prend un peu plus de recul, l'organisme c'est finalement aussi l'aboutissement d'une longue trajectoire au cours des âges qui a vu l'évolution de ses composés, de leur chimie, de leur forme, de leur organisation mutuelle, des relations qui les unissent. Les organismes se sont construits sur un héritage qui, de génération en génération, s'est accumulé, modifié, élagué. Que reste-t-il de cet héritage ? Quelle forme prend-t-il ?

La vie, ses comportements, en particulier leur complexité, dépendent d'une part d'éléments réactifs (molécules et groupes de molécules), d'autre part de l'ensemble des interactions que ces éléments peuvent avoir entre eux. Toutes les interactions ne sont pas fortes et le plus souvent couronnées de succès, c'est-à-dire probables avec leurs protagonistes proches et nombreux dans l'espace. Ces interactions fortes, qu'on identifie en biologie comme finalement pratiquement déterministes, et dont on tire une grande partie de nos raisonnements sur le fonctionnement de la cellule, cohabitent avec une foule d'autres interactions, plus incertaines, plus diffuses, qui peuvent se limiter au simple contact provoquant une petite pause dans la dynamique des molécules en cause, un délai dans leurs possibles réactions. Il faut par moments se détacher un peu de l'image du fonctionnement cellulaire que la biologie construit depuis plus de 70 ans, c'est-à-dire ne pas s'imposer cette vision un peu trop mécaniste pour voir davantage la cellule comme un sac biochimique, pas particulièrement ordonné, surtout quand on le regarde de près, fourmillant d'un nombre infini de très petites interactions entre les très nombreux éléments moléculaires de tous types : petites molécules comme les ions, l'eau, les sucres, les acides aminés, ... beaucoup de molécules de taille intermédiaire comme de petites séquences nucléotidiques, de petits ARN, certains issus de voies de synthèse, d'autres n'étant que des produits de dégradation, beaucoup de grosses molécules aussi comme les protéines, les acides nucléiques, ... et enfin de gros éléments supramoléculaires incluant les assemblages hétérogènes de protéines, acides nucléiques, mais aussi les membranes. La cellule et

son environnement, vus à l'échelle mésoscopique, sont un fouillis infernal ! Et pourtant, tout cela fonctionne ! Pourquoi ? ou plutôt comment ? comment est-ce possible ?

Ces questions sont d'autant plus importantes si on les considère dans le contexte des origines du vivant. Le biologiste qui étudie la vie actuelle est rassuré de dire que ce bon fonctionnement vient de l'harmonieuse (*fort ingénieuse* devrait on presque dire) façon qu'ont les éléments cellulaires d'interagir entre eux, un tel produisant un effet sur un autre, et ainsi de suite. Les recherches que l'on mène vont nécessairement dans ce sens car nos représentations mentales et les instruments qui servent à analyser la matière vivante sont construits dans le paradigme fort du vivant machine. Ceci ne peut pourtant être valable aux origines du vivant, durant la longue période (longue de combien ?, aucune idée...) qui sépare l'émergence des proto-métabolismes, proto-système génétique, proto-réseaux d'interaction, ..., à un organisme, une cellule, dont chaque partie agit avec un niveau de fiabilité tel qu'il n'a pas vraiment à rougir devant un organisme actuel. Entre temps ? eh bien, du modulaire, des interactions pas très spécifiques, des branchements de ce qu'on pourrait éventuellement qualifier de réseaux, des réseaux fragiles mais aussi dynamiques, temporaires, ... Des vestiges évidents de ce passé ? Vraisemblablement les plasmides bactériens ou les particules virales qui maintiennent encore une flexibilité importante du fonctionnement cellulaire, par l'écosystème moléculaire que constituent les interactions croisées entre génomes viraux et petits ARNs (notamment miARNs) cellulaires [Berezikov, 2005] et par leur rôle dans l'édition génétique des cellules [Witzany, 2006 ; Forterre, 2006 ; Koonin, 2006], ou la coexistence de métabolismes très efficaces mais complexes et exigeants (ex : cycle de Krebs) et de métabolismes simples mais fiables (ex : (resp.) fermentation lactique) qui ont été maintenus au cours de l'évolution et la bascule progressive ou rapide, mais souple, continue, permise entre eux. Il y en a certainement d'autres bien sûr ; dans cette partie, nous y reviendrons et en discuterons de façon plus approfondie, en particulier quand nous aborderons la théorie de l'atavisme.

Ces vestiges, s'ils sont toujours présents, le sont parce qu'ils ont été sélectionnés et maintenus sous pression évolutive depuis des milliards d'années, car finalement indispensables aux organismes, aux écosystèmes, à la Vie en général. Ils sont vraisemblablement indispensables pour assurer au moins la robustesse, la résilience et l'évolutivité des organismes vivants. Ces vestiges ont co-évolué avec le reste des systèmes des organismes et sont donc tout autant modernes que des systèmes jugés *a priori* plus complexes, plus régulés. C'est leur caractère *jugé primitif* qui surprend. Car oui, je pense que si l'on identifie davantage de *systèmes* au sens de machines, d'assemblages, de formes complexes, c'est parce que c'est bien ceci que l'on cherche dans les cellules (pensant que c'est bien là que se joue le fonctionnement du vivant) et non des *systèmes* au sens d'ensembles fragiles et informes d'interactions faibles et éventuellement rares. Pourtant le rôle de ces derniers est vraisemblablement crucial pour la Vie, puisque l'évolution en a décidé ainsi.

Je trouve dès lors intéressant de regarder non plus les organismes comme on les regarde habituellement, comme des machines molles homéostatiques, mais comme *d'une part des tranches temporelles dans une trajectoire évolutive à travers un espace morphogénétique complexe, ces tranches définissant leur fonctionnement interne, d'autre part un ensemble de senseurs et effecteurs avec l'environnement extérieur, interface permettant à la créature-tranche morphogénétique d'être*

sujette à la pression évolutive. Cette phrase est compliquée, plus que les précédentes. Je m'attellerai à mieux la définir dans ce chapitre. Pour le moment, parlons réseaux !

8.3.2. Réseaux Booléens – représentations d'une réalité biologique

Un réseau d'interactions (comme tout modèle d'ailleurs) est une projection d'une image mentale simplifiée que nous produisons, image mentale qui résulte de l'expérience et de la formation qu'a reçu chacun, de son expérience personnelle, ainsi que d'un consensus scientifique auquel nous adhérons en tant que scientifiques.

Les réseaux et leurs cousins, les systèmes différentiels permettent de donner vie à des comportements complexes que nous assimilons à ceux que nous pouvons observer dans le vivant. Ils constituent en cela de puissants outils pour nous aider à réfléchir sur son fonctionnement. Les réseaux Booléens en sont probablement l'expression la plus simple.

Je ne suis personnellement pas très convaincu de l'approche « représentative » de la modélisation dans de nombreuses situations et surtout pour beaucoup de formalismes. Malheureusement loin de mes certitudes de jeune chercheur, je pense maintenant (et certains de mes collègues plus optimistes que moi trouvent que c'est à tort) que, quoi qu'on fasse, lorsqu'on cherche à représenter « le plus fidèlement possible » un système naturel au moyen d'un modèle mathématique ou informatique, on dénature tellement l'objet biologique étudié qu'il paraît trop présomptueux d'en tirer un réel enseignement sur sa réalité biologique. Fort heureusement, j'ai aussi appris à rester un minimum pragmatique et je sais aussi que quelle que soit la façon dont on s'y prend, expérimentale ou théorique, bien que la réalité du monde naturel soit insaisissable, l'objet des sciences n'est pas d'en saisir la substance mais bien d'aider notre esprit à s'en forger une idée communicable et pouvant faire consensus sous la forme d'une construction collective. En cela, je pense que tous les enseignements que nous pouvons tirer par l'observation des systèmes réels comme par les projections que nous en faisons sur des abstractions, tels les réseaux Booléens, participent à la construction d'une compréhension du vivant.

Un réseau, donc, est une représentation simplifiée d'entités, que l'on représente par des *nœuds*, et des relations qui les unissent, que l'on représente par des *arcs*. Dans le contexte de la modélisation des réseaux de régulation génétique (qui vont nous intéresser ici), les nœuds représentent des gènes ou des produits de gènes, les arcs les relations de régulation (inhibition, activation ...) que des nœuds imposent à d'autres. Selon les formalismes utilisés, des paramètres sont affectés aux nœuds et/ou aux arcs pour spécifier la dynamique de transition des états sur les nœuds. Des formalismes pour les réseaux de régulation, il y en a beaucoup, les principaux étant les réseaux de Thomas [Thomas, 1980], les réseaux de Petri [Petri, 2008], les réseaux de Kauffman [Glass, 1973 ; Kauffman, 1969] et leur sous-classe, les réseaux de Hopfield [Hopfield, 1982]. Les premiers sont très fréquemment utilisés car proches de la description naturelle de la régulation génétique et surtout, parce qu'ils sont le pendant discret des systèmes d'équations différentielles, ils peuvent être facilement obtenus à partir de ces derniers. Les réseaux de Petri sont particulièrement adaptés à décrire les processus chimiques finement, en tenant compte de la stoechiométrie

notamment. Les réseaux de Kauffman sont eux des automates Booléens dont les interactions sont dictées par des relations logiques. Leur sous-classe, les réseaux Booléens à seuil, dits « de Hopfield », ont été initialement proposés par McCulloch et Pitts pour la modélisation des réseaux de neurones formels [McCulloch, 1943]. Ils le sont toujours. Ils ont ensuite été popularisés par J. Hopfield [Hopfield, 1982].

Sur les réseaux de régulation et leurs diverses représentations, il y a beaucoup à dire. Une bonne dizaine de thèses probablement ! Disons le tout de suite : je vais botter en touche sur ce point. D'abord parce qu'on n'en finirait jamais de présenter cela correctement et que cela diluerait le message que je souhaite porter ; pour en savoir un peu plus, je renvoie plutôt le lecteur vers ces articles assez récents, synthétiques et explicatifs [Bornhold, 2008 ; Tran, 2013]. Ensuite parce que je ne suis certainement pas le plus compétent pour parler de la modélisation des réseaux de régulation génétiques « réels » et certainement pas par les réseaux de Thomas auxquels je me suis très peu intéressé. Enfin parce que j'ai fait un choix particulier d'un formalisme particulièrement utile et efficace pour ce que je souhaite réaliser et pour appuyer mes réflexions, formalisme que je vais maintenant justifier par leur historique à TIMC-IMAG, mon propre historique avec ces réseaux et l'utilisation que je compte en faire dans mes projets.

8.3.3. Réseaux Booléens à Seuil – Réseaux Booléens à Signe

8.3.3.1. Historique de leur utilisation dans mes travaux

Les réseaux que McCulloch et Pitts, puis Hopfield ont décrits et utilisés sont des Réseaux Booléens à Seuil, en anglais *Threshold Boolean Networks*, que j'abrègerai par TBN. Ces réseaux ont été développés à l'époque (1943) pour modéliser le neurone biologique. Ils sont la source de ce qui deviendra le réseau de neurone qui a permis l'émergence d'une immense quantité de travaux sur les systèmes neuronaux biologiques, sur l'IA et les automatismes ; ce sont les mêmes (et leurs descendants) dont on parle toujours actuellement lorsqu'on évoque le Deep Learning en IA.

Lorsque j'ai commencé à travailler à TIMC-IMAG sur les réseaux Booléens, en 2008, j'ai utilisé ce formalisme qui présentait plusieurs avantages à mes yeux, en premier parce qu'il s'agissait d'automates (déterministes) et que l'on pouvait concevoir des systèmes oscillants finalement proches de ceux que l'on pouvait obtenir avec des systèmes d'équations différentielles comme l'oscillateur anharmonique, l'oscillateur de van der Pol ou l'oscillateur de Wilson-Cowan. Avec Hedi Ben Amor, le premier étudiant en thèse que j'ai co-encadré, nous avons exploité ces systèmes dans le cadre du stockage dynamique d'information et de l'évocation mnésique [Elena, 2008 ; Ben Amor, 2008, 2009, 2010 a, 2010b, 2013a, 2011, 2013 ; Glade, 2009, 2011]. C'est aussi avec ce système qu'Hedi Ben Amor, Fabien Corblin et moi, avons exploré la question de l'inférence de réseaux Booléens appliquée à la modélisation de systèmes de régulation biologiques [Glade, 2011 ; Ben Amor, 2013]. De tels travaux sur l'inférence de réseaux Booléens avaient déjà été faits sur d'autres types de réseaux Booléens, les réseaux de Thomas, en particulier par Fabien Corblin en thèse avec Laurent Trilling et Eric Fanchon [Fanchon, 2005 ; Corblin, 2009, 2010, 2011] en utilisant la même technique

que nous avons appliquée aux TBNs, la programmation logique (non-monotone) par contraintes, en ASP (le langage de programmation *Answer Set Programming*).

Ce qui est peut-être nouveau c'est d'abord le substrat sur lequel nous avons travaillé, des TBNs. Pour replacer les choses dans leur contexte, nous travaillions à ce moment sur des systèmes computationnels et mémoriels fondés sur des TBNs, systèmes pouvant potentiellement être implémentés sous la forme de systèmes électroniques, et nous nous sommes posé la question de la possibilité d'inférer toutes les solutions possibles, pour un nombre de nœuds donnés, pour réaliser un tel système computationnel ou mémoriel de façon à ce qu'il effectue ce qui est demandé : en clair, il s'agissait de déterminer la structure interne (au formalisme de TBN) de tous les systèmes réalisant une certaine fonction ou un ensemble de fonctions, comme si on demandait à un ingénieur toutes les manières d'implémenter le système électronique d'une télévision. C'est en réalisant ce travail que nous nous sommes aperçus que transposer les raisonnements que nous avions aux réseaux de régulation biologiques avait du sens.

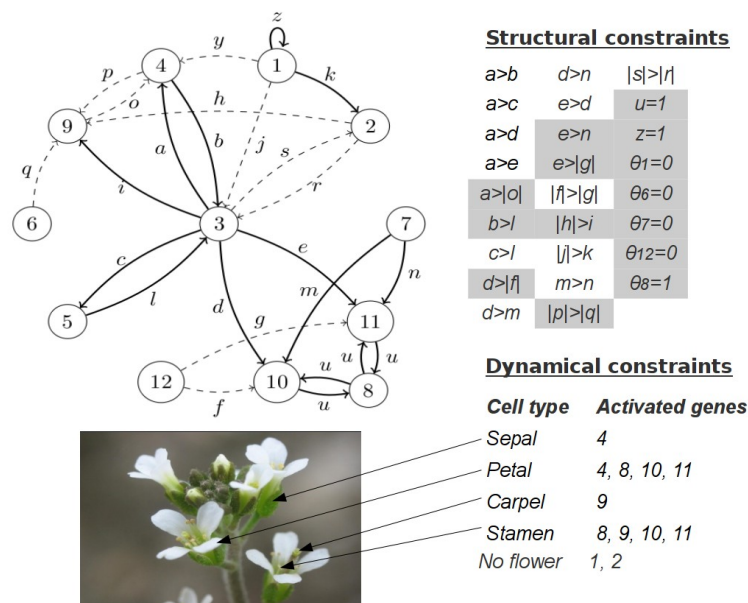


Figure 8.25. Réseau de régulation de la morphogenèse florale d'*Arabidopsis thaliana*. L'activation différentielle de gènes de ce réseau induit la différenciation des cellules végétales en 4 types cellulaires, sépales, pétales, pistil et étamines (+ 1 mutant *no flower*). C'est ce qui est requis sous la forme de contraintes sur la dynamique du réseau de régulation proposé par Mendoza [Mendoza, 1998] : les 5 points fixes correspondants doivent exister dans son graphe de transition. De plus, le réseau, modélisé par un TBN, comprend des contraintes sur la structure de son graphe d'interactions ; ces contraintes décrivent les relations entre les poids des arcs du réseau sous la forme d'inégalités. Mendoza & Alvarez-Buylla exprimaient déjà ces relations, mais certaines s'avéraient être des non-sens : ils comparaient en effet des valeurs de poids sur des arcs n'entrant pas sur la même fonction. Nous avons donc éliminé ces contraintes.

Ce qui est également nouveau, donc, c'est la proposition que j'ai faite à ce moment là et que nous avons couchée sur papier dans l'article de 2013, à savoir la possibilité de considérer des ensembles entiers de réseaux, plutôt qu'un modèle unique, pour modéliser un réseau de régulation génétique : ce ne serait plus un modèle (une "machine" représentée dans notre ordinateur par un réseau Booléen) qui fonctionne et génère un comportement mais toute une population de TBNs. A ce propos, pour le présenter très rapidement, le point de vue qui occupera votre lecture dans la section

8.3.6 et que je défends comme cadre conceptuel pour réfléchir sur les réseaux biologiques est que tant qu'aucun modèle unique n'a pu être déterminé, par élimination des autres grâce à une connaissance approfondie du système biologique d'intérêt, tous les modèles qui satisfont la connaissance sur la structure du système biologique (ex : nombre de gènes, interactions connues, type d'interaction, *etc.*) et sur sa dynamique ont la même légitimité ; dès lors le système 'réel' peut être interprété comme une combinaison de l'expression d'un ensemble de systèmes modèles.

Un autre point de vue possible, respectant l'idée de « à comportement biologique donné, modèle unique donné », consiste à considérer qu'un modèle est un résultat possible parmi l'ensemble des explorations qu'a permise ou qu'aurait permise l'évolution. Je parlerai à cet égard de « stratégies retenues par l'évolution », une *stratégie* n'exprimant ici aucun dessein mais une configuration possible d'un système telle que celui-ci assure (au moins) une certaine fonction (ou plusieurs) nécessaire(s) au système biologique qui en dépend. C'est ainsi que nous avons repris la modélisation faite par Mendoza et Alvarez-Buylla [Mendoza, 1998] du réseau de régulation de la morphogenèse florale d'*Arabidopsis thaliana*, le végétal cobaye des phyto-physiologistes. Nous avons traduit le graphe de transition et la dynamique requise du TBN proposé par ces auteurs en inégalités puis en contraintes logiques sur les valeurs des paramètres du réseau et de sa dynamique (voir Figure 8.25). Du point de vue de sa dynamique, ce réseau devait au moins permettre l'obtention de 4 points fixes correspondant aux 4 tissus floraux (sépalés, pétales, pistil, étamines) et un mutant (dit *no flower*). Nous avons développé un programme en Prolog pour implémenter ces contraintes de structure et de dynamique et obtenu, pour *Arabidopsis thaliana*, 532 modèles satisfiables, c'est-à-dire respectant les connaissances (partielles) sur sa structure et sur sa dynamique. Ces modèles sont tous différents. Outre le fait que tous possèdent au moins la contrainte de structure imposée (donnée Figure 8.25), certains se ressemblent, ne différant que sur quelques paramètres, d'autres au contraire sont très différents d'aspect. D'où cette idée de « stratégies » qu'aurait pu explorer l'évolution. C'est là que j'ai vraiment commencé à m'intéresser à la diversité de ces stratégies, à comment les classer, sur quels critères, comment mesurer des distances entre réseaux ou parties de réseaux, *etc.* La première classification que j'ai tentée, car j'avais seulement ces données et peu de temps³⁴, c'est celle des réseaux par rapport à la taille des 5 bassins d'attraction correspondant aux attracteurs « tissus floraux » et *no flower* [Glade, 2011]. En classant par ACP³⁵ ou par la méthode des distances, je pensais alors différencier, sur ce critère très limité certes, des grands groupes de stratégies. J'ai pu obtenir le graphique montré figure 8.26 : la classification très grossière que j'avais faite à l'époque permettait d'obtenir 4 groupes de tailles de bassins d'attraction, groupes que j'associais – rapidement – à mes fameuses stratégies. S'il y a beaucoup à redire sur la façon de classer et sur le critère de classement lui-même, cette première approche a eu le mérite de susciter en moi un certain nombre de questions d'intérêt. Ces questions, j'ai commencé à les développer en 2013 avec une

34 j'ai fait ce travail – une classification des réseaux par la taille de leurs bassins d'attraction et l'écriture du brouillon d'un article qui devait être soumis le lendemain matin - dans le train de retour entre Brest et Grenoble.

35 L'Analyse en composantes principales, ou ACP, consiste en la recherche de d vecteurs propres (les composantes principales) et leurs valeurs propres associées (« tailles » des vecteurs), décrivant au mieux un nuage de points dans un espace à d dimensions. Ces vecteurs sont obtenus par diagonalisation de la matrice de covariance (ou de corrélation) du nuage de points dans son espace d'origine. La variance du nuage de points est conservée dans les valeurs propres.

excellente étudiante en L3, Roselyne Chauvin³⁶, qui a fortement contribué à faire progresser conceptuellement ce qui allait devenir une de mes thématiques principales vers 2017. A ma demande, R. Chauvin a exploré plusieurs points, sans nécessairement les approfondir, en particulier l'usage du langage ASP (au lieu de Prolog) pour faire de l'inférence sur des TBNs, la recherche par inférence de modules fonctionnels réalisant une fonction particulière dans des réseaux de taille plus importante, ce qui allait devenir une question majeure de mon activité actuelle avec notamment le problème de l'extensibilité des réseaux sur lequel je reviendrai largement, la question du degré de dégénérescence vs minimalité qui fut un véritable casse tête jusqu'à ce que nous laissions les TBNs pour les SBNs, l'établissement d'une distance entre réseaux et son usage pour classer des réseaux ou rechercher des sous-réseaux, ...

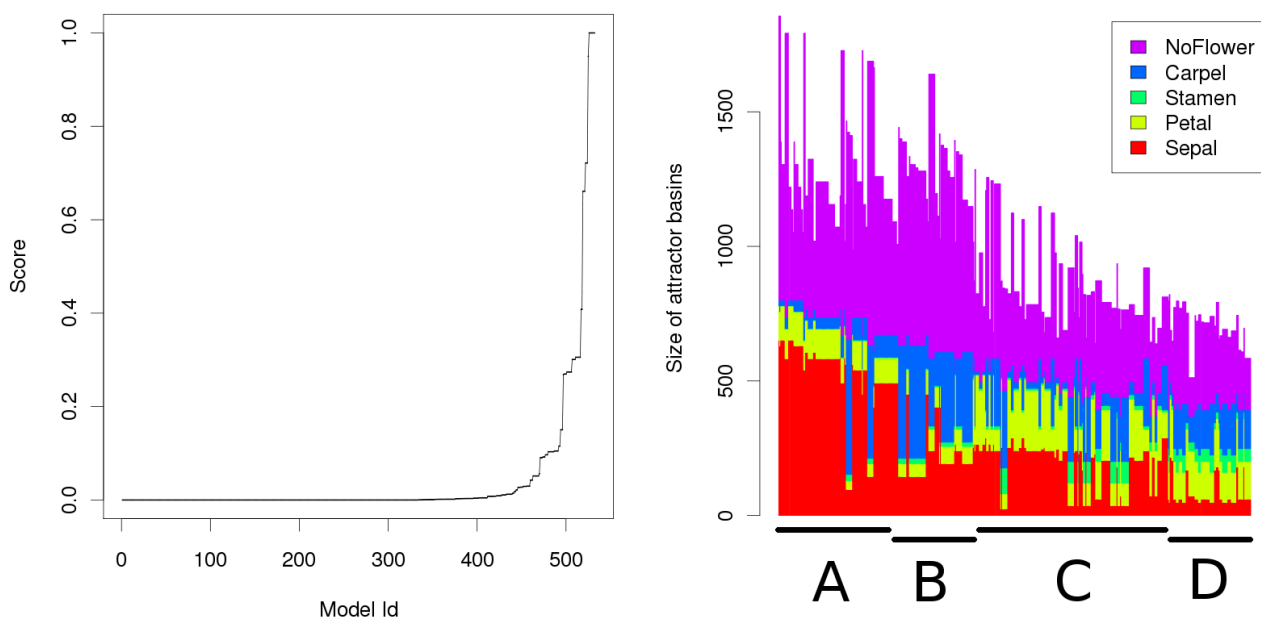


Figure 8.26. Classification par la taille des bassins d'attraction des TBN. Ces graphiques montrent le résultat d'une classification des TBNs consistants avec les connaissances sur la structure et la dynamique de la morphogenèse florale d'*Arabidopsis thaliana*. La classification résulte ici d'un scoring sur la taille des 5 bassins d'attraction requis. Ce score, déterminé empiriquement, est le produit de fonctions gaussiennes centrées sur une taille *optimale* déterminée pour chaque bassin d'attraction (*optimale* sous-entend une notion de robustesse) ; la taille dite optimale est calculée comme la moyenne de la taille des bassins des 4 tissus floraux pris ensemble. Avec un tel score, on peut ordonner les modèles dans un barplot montrant la taille cumulée des 5 bassins d'attraction. Cette toute première approche reste très approximative mais montrait tout de même qu'il semblait y avoir des groupes, grossièrement : (A) gros bassin d'attraction pour les sépales, (B) gros bassin d'attraction pour le pistil, (C) gros bassin d'attraction pour les pétales, (D) bassins d'attraction également équilibrés.

Avant de poursuivre sur l'actualité de ce travail (de 2017 à maintenant), nous allons maintenant prendre le temps de comprendre le formalisme de ces réseaux, leur fonctionnement, mais aussi d'autres manières moins classiques de les présenter et qui nous seront utiles par la suite.

³⁶ sur ma recommandation, Roselyne a ensuite été admise à l'ENS Paris pour un Master de neurosciences et elle a poursuivi en thèse de neurosciences (*The efficient brain: on how connectivity modulations underpin cognitive tasks*) en Hollande, à Donders Univ.

8.3.3.2. Formalisme

8.3.3.2.a. Fonctions et Réseaux Booléens

Un réseau Booléen, comme son nom l'indique, est formé de nœuds aux états Booléens (0 ou 1), ces nœuds, reliés par des arcs orientés, réalisant des fonctions $f : \{0, 1\}^d \rightarrow \{0, 1\}$ (la fonction a d entrées). Un réseau Booléen est un tuple (V, F) dans lequel V est l'ensemble des variables du réseau, variables incluent les états des nœuds et les paramètres du réseau, et F une mise en correspondance (*mapping*) associant une fonction Booléenne à chaque variable dans V . Pour chaque valeur $x \in V$, la fonction Booléenne $F(x)$ associée à x calcule la nouvelle valeur de x (son état successeur) à partir des valeurs des variables dans V . F définit la façon dont x sera mise à jour. Dans ce document, nous nous concentrerons sur les fonctions Booléennes propositionnelles (PBF), les fonctions Booléennes à seuils (TBF), et particulièrement celles à signe (SBF). Les paramètres des fonctions TBF et SBF seront à valeurs entières dans \mathbb{Z} .

En réalité, utiliser des réseaux Booléens avec des paramètres entiers n'est pas une nécessité et nous pourrions tout aussi bien travailler avec des paramètres à valeurs réelles et des états multivalués. Cependant, un tel réseau pourrait être ramené sans trop de difficultés à un réseau Booléen équivalent de taille égale ou supérieure. Utiliser de tels réseaux à valeurs réelles, ou multivalués, n'est qu'une question d'expressivité : cela permet à un modélisateur d'exprimer le plus directement possible son modèle, c'est-à-dire dans le langage qui lui semble le plus proche de ce qu'il perçoit du système biologique. L'usage des réseaux Booléens avec des paramètres entiers facilite cependant leur étude. Je reviendrai néanmoins sur la question de l'expressivité que je trouve très intéressante et qui soulève plusieurs problèmes dont celui de l'unicité des solutions et des modèles des mécanismes sous-jacents des systèmes modélisés.

8.3.3.2.b. Fonctions et Réseaux Booléens Propositionnels (PBF et PBN)

Ce sont les réseaux Booléens « classiques » dans lesquels les fonctions de mise à jour F sont des propositions logiques pouvant prendre toute forme possible pour effectuer des opérations avec les entrées V de $F(x)$. Par exemple, une fonction ayant 2 entrées x et y peut combiner ces entrées sous la forme $x \oplus \neg y$ (ou $x \text{ XOR } \neg y$ ³⁷, équivalente à $\neg x \oplus y$ qui revient finalement à $\neg(x \oplus y)$) pour mettre à jour son état, comme illustré dans le réseau figure 8.27.

La définition de tels réseaux est facilitée par le pouvoir expressif conféré par les propositions logiques. Au même titre que pour des langages de programmation, une forte expressivité est utile au concepteur/modélisateur pour formuler dans le langage qui est le plus adapté à modéliser un système naturel. Néanmoins, cela signifie que les fonctions Booléennes ne sont pas nécessairement (sauf à avoir la même expression logique) de même nature. Dans l'exemple donné ci-dessus, les fonctions A et B ont des expressions différentes : elles sont donc de nature différente. Pour bien réaliser ce que

³⁷ La fonction XOR est le « ou logique exclusif ». Déterminer $x \text{ XOR } y$ signifie que si $x=1$ et $y=0$, ou bien si $x=0$ et $y=1$, alors $x \text{ XOR } y=1$, mais si $x=1$ et $y=1$, alors le « ou » n'est pas exclusif et $x \text{ XOR } y=0$.

signifie l'expression « même nature » que j'emploie ici, le mieux est de considérer des fonctions informatiques (par exemple *mov* ou *add* ou *sub* en langage Assembleur³⁸), fonctions dont on comprend bien qu'elles ne font pas la même chose parce qu'elles ont un « code interne » (sous-jacent) différent. Ces réseaux ont donc une expression hétérogène qui ne facilite pas l'analyse que nous souhaitons en faire. Les réseaux Booléens à seuil, eux, sont davantage homogènes dans leur écriture : toutes les fonctions ont la même formulation à une dimension donnée.

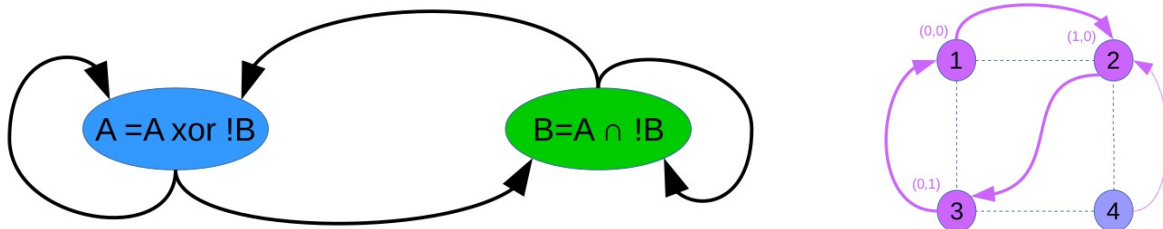


Figure 8.27. Exemple de réseau Booléen. Ce réseau Booléen (à droite) est constitué des fonctions A et B. Si l'actualisation des états de A et B est parallèle, alors le graphe de transition (voir sous-section 8.3.3.2.f ci-dessous) correspondant (A,B) est donné à droite. Ce graphe contient un seul bassin d'attraction composé d'un attracteur formé de 3 états successifs 1→2→3. L'état 4 est transitoire et rejoint l'attracteur en 2.

8.3.3.2.c. Fonctions et Réseaux Booléens à Seuil (TBF et TBN)

Une fonction Booléenne à seuil (TBF pour Threshold Boolean Function) est définie d'une part par d entrées, d étant le nombre de nœuds du réseau (sa *dimension*), dont les arcs correspondants sont pondérés par des poids $w_1, \dots, w_d \in \mathbb{R}$, d'autre part par un seuil d'activation θ que la somme des entrées pondérées doit dépasser strictement³⁹ pour permettre l'activation du nœud mis à jour. Autrement dit, une TBF est intégralement définie par ses paramètres $(w_1, \dots, w_d, \theta)$ et est mise à jour comme suit :

$$f(x_1, \dots, x_d) = \begin{cases} 1 & \text{si } \sum_{i=1}^d w_i x_i > \theta \\ 0 & \text{sinon} \end{cases} \quad (\text{Equation 1})$$

Par exemple (voir figure 8.28), la fonction $x \vee y$ (x OU y) peut être écrite sous la forme d'une fonction TBF à deux entrées et un seuil nul, de paramètres $(1, 1, 0)$ ce qui, nous le verrons, revient à une fonction de type SBF à deux entrées, de paramètres $(1, 1)$. En revanche, la fonction $x \wedge y$ (x ET y) ne peut être écrite que sous la forme d'une fonction TBF à deux entrées et un seuil non nul, de paramètres $(1, 1, 1)$. Autre exemple, la fonction $\neg(x \vee y)$ peut aussi être écrite comme une unique TBF à deux entrées, de paramètres $(-1, -1, -1)$. Certaines propositions logiques ne peuvent en revanche pas être restituées sous la forme de TBFs seules, par exemple la fonction $\neg(x \oplus y)$ qui ne peut être exprimée que sous la forme d'une combinaison de TBFs (voir figure 8.29) dont la durée de

38 Ces fonctions *mov* v,m , *add* v,m et *sub* v,m , réalisent respectivement le déplacement d'une valeur v dans une case mémoire m , l'addition ou la soustraction d'une valeur v avec celle contenue à l'adresse mémoire m .

39 Notons qu'il existe dans la littérature plusieurs façons de mettre en relation le seuil et la somme des états pondérés, par exemple avec une activation lorsque la somme des états pondérée est supérieure ou égale au seuil. Cela ne change rien aux conclusions que l'on pourrait tirer de travaux utilisant ces variations diverses, ni même aux méthodes de calcul employés pour les analyser ou en faire l'inférence.

traitement des entrées est plus longue (nous verrons un peu plus loin que ceci a son importance).

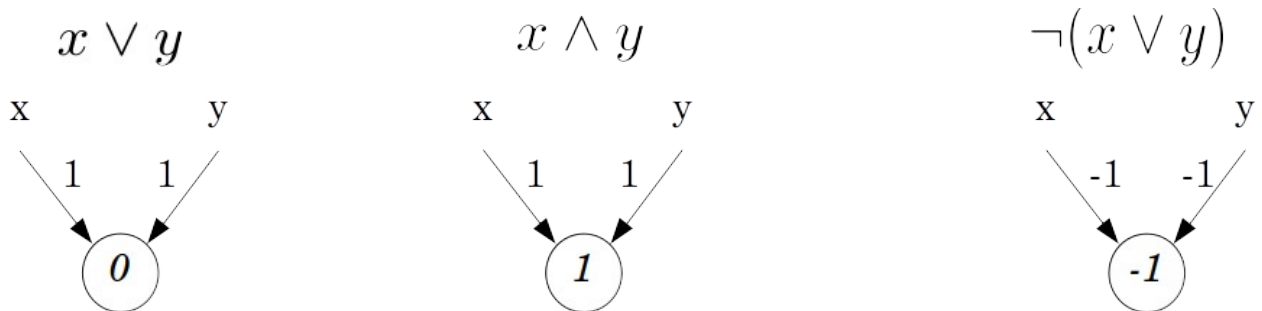


Figure 8.28. Exemples de fonctions Booléennes à Seuil (TBF). Ces 3 fonctions Booléennes à seuil sont la retranscription directe de fonctions propositionnelles correspondantes (écrites au dessus).

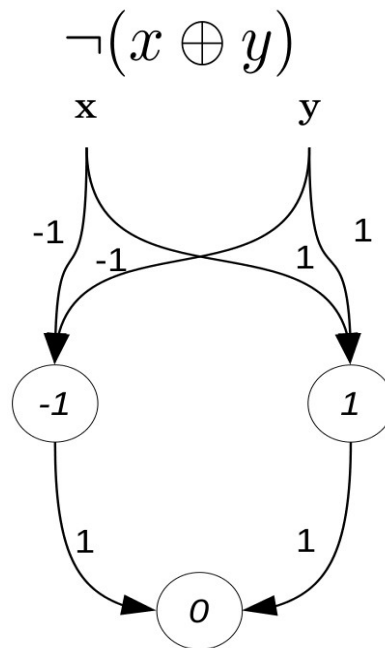


Figure 8.29. Exemples de composition de fonctions Booléennes à Seuil (TBF). La fonction propositionnelle $\neg(x \oplus y)$ ne peut s'exprimer sous la forme d'une seule fonction Booléenne à seuil. Pour la « simuler », il faut combiner l'action de 3 TBFs.

Comme nous le verrons quand nous parlerons de classe d'équivalence, le fait que les paramètres (les poids et les seuils) soient de nature différente pose beaucoup de problèmes pour certains calculs à réaliser, en particulier la détermination des réseaux minimaux et les calculs de complexité structurelle des réseaux.

En réalité, toute fonction à seuil peut être exprimée sous la forme d'une (si le seuil est nul) à deux fonctions à signe couplées (voir figure 8.30), dont l'une est une fonction auto-activée et initialisée à 1 (donc en permanence à l'état 1) et l'autre, celle qui représente la TBF d'origine que l'on traduit en SBF, reçoit un arc de la première avec un poids égal à $-\theta$. Le seuil peut donc être traduit sous la

forme d'une structure de réseau particulière : le seuil est assimilable à un mécanisme sous-jacent. De plus, le fait que la SBF simulant l'effet du seuil doive être initialisée à 1 indique que le paramètre seuil contient aussi un élément de dynamique. En cela, le paramètre seuil diffère fondamentalement des paramètres poids. Nous verrons, lorsque nous parlerons de classe d'équivalence que cela nous a posé problème.

8.3.3.2.d. Fonctions et Réseaux Booléens à Signe (SBF et SBN)

Descendons encore d'un cran dans les niveaux d'expressivité, mais gagnons en homogénéité avec les SBNs. Les SBNs utilisent le même formalisme que les TBNs, à ceci près que les seuils des fonctions (SBFs) sont tous de valeur 0 : une fonction est donc activée lorsque la somme des états pondérés entrants est supérieure à 0. On n'a donc plus besoin du paramètre seuil. C'est ce type de formalisme que nous utiliserons principalement par la suite, mais de temps en temps, nous ferons appel aux PBNs ou aux TBNs, en particulier lorsque nous aborderons le problème de la décomposition des PBNs en SBNs.

Notons comme conséquence du passage des TBNs aux SBNs fait au cours de nos travaux, qu'une grande partie de TBNs de dimension d (ceux qui n'ont pas de TBFs de seuil non nul) n'existe en réalité que sous la forme de SBNs de dimension $d+1$.

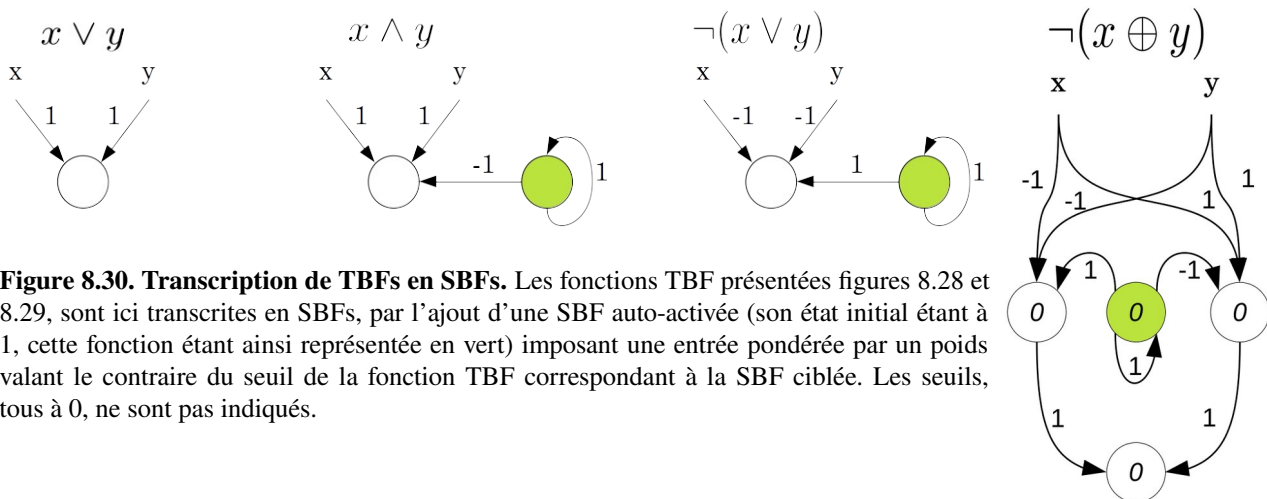


Figure 8.30. Transcription de TBFs en SBFs. Les fonctions TBF présentées figures 8.28 et 8.29, sont ici transcrites en SBFs, par l'ajout d'une SBF auto-activée (son état initial étant à 1, cette fonction étant ainsi représentée en vert) imposant une entrée pondérée par un poids valant le contraire du seuil de la fonction TBF correspondant à la SBF ciblée. Les seuils, tous à 0, ne sont pas indiqués.

8.3.3.2.e. Partitionnement de l'espace des paramètres, Classe d'équivalence, Réseau minimal

Toute fonction de type TBF ou SBF, entièrement définie par ses paramètres, peut être représentée par l'ensemble des fonctions qui lui sont équivalentes, à savoir sa classe d'équivalence. Je vais commencer par en donner une définition insuffisante et inadéquate par rapport à notre conception des réseaux d'interaction. Je discuterai ensuite, justement, de cette conception particulière que nous défendons, des réseaux d'interaction biologiques. Enfin, je reviendrai sur la définition que nous retiendrons des classes d'équivalence.

De façon générale, et ceci est toujours valable quelle que soit la définition qu'on lui donne, la classe

d'équivalence d'une fonction regroupe les fonctions similaires selon un ensemble de critères que le chercheur aura décidé. Elles peuvent ainsi regrouper les fonctions ayant le même comportement, la même structure, ... Ce sont les classes d'équivalence ainsi définies (et nous verrons que cette définition est laissée à l'appréciation de celui qui décide de la façon de regrouper les fonctions, donc de la façon d'apprécier leur similarité ; nous en choisirons une) dont nous nous servons pour caractériser les fonctions Booléennes (notamment pour calculer leur complexité). L'ensemble des classes d'équivalence des fonctions de dimension d (d entrées, y compris les entrées de poids nul) peut être représenté dans un espace de dimension d (pour les SBFs ayant d paramètres, leurs poids respectifs), et de dimension $d+1$ (pour les TBFs qui ont comme paramètre supplémentaire le seuil). Je ne parlerai ici que des classes d'équivalence des SBFs ; une discussion sur les soucis posés par le seuil des TBFs se trouve dans les articles [Christen, 2019 ; Segretain, 2020b].

Cela nous permet également de définir efficacement la notion de fonction minimale, unique à la permutation des entrées près, introduite par A. Elena [Elena, 2009]. Deux SBFs (ou TBFs) peuvent s'avérer équivalentes, même si leurs paramètres diffèrent, tant que leur table de vérité est identique. Par exemple, les SBFs de paramètres (3,2), (7,3) et (1,1) sont équivalentes ; elles produisent les mêmes sorties dans la table de vérité. La fonction Booléenne minimale d'une classe d'équivalence est celle dont les valeurs des poids (entiers) sont minimales (c'est-à-dire les plus proche possible de 0). Là aussi, je préciserai quelque peu ce point sous peu.

Partitionnement de l'espace des paramètres. Tout vecteur dans l'espace des paramètres \mathcal{W} , de base orthogonale (w_1, w_2, \dots, w_d) , décrit une instance de fonction (SBF dans l'espace des poids). Ces instances peuvent être regroupées par similarité en classes d'équivalence. Pour obtenir les classes d'équivalence, on commence par considérer le partitionnement de \mathcal{W} (les poids des SBFs) par l'ensemble des hyperplans (droites en 2D, plans en 3D, *etc.*) dont les équations correspondent à toutes les combinaisons possibles avec les poids, par exemple en dimension 3, avec les poids w_1, w_2, w_3 , on a les 7 plans d'équation $w_1=0, w_2=0, w_3=0, w_1+w_2=0, w_1+w_3=0, w_2+w_3=0$ et $w_1+w_2+w_3=0$ (voir figure 8.31). Ces plans délimitent des régions coniques de l'espace, des partitions, que nous appelons également *layouts* \mathcal{L} (ou arrangements). Les fonctions à l'intérieur d'un layout ont toutes la même table de vérité. Une classe d'équivalence peut être constituée de 1 à plusieurs layouts. Les fonctions sont en effet considérées équivalentes à la permutation de leur entrées près. Par exemple, les fonctions (-1,2) et (2,-1) appartenant respectivement au layouts \mathcal{L}_{21} et \mathcal{L}_{22} dans la figure 8.31 sont équivalentes. Dans cet exemple, les layouts \mathcal{L}_{21} et \mathcal{L}_{22} appartiennent à la même classe d'équivalence \mathcal{A}_2 . Je développerai davantage le partitionnement de \mathcal{W} un peu plus loin dans le texte.

Classe d'équivalence – 1ère définition (obsolète). De prime abord, la classe d'équivalence d'une fonction pourrait être définie comme l'ensemble des fonctions possédant la même table de vérité, permutation des entrées comprises (voir figure 8.31). La table de vérité d'une fonction contient à la fois les informations de structure et de dynamique de la fonction (contrairement à une définition des fonctions fondée uniquement sur les valeurs de leurs paramètres). Chaque classe d'équivalence

comprend ainsi un représentant de fonction minimale. Je vais profiter de ce moment pour discuter un peu de ce que j'entends, tout de même, par fonction minimale, dans le cadre de notre travail.

A priori, si les descriptions de ce que sont les classes d'équivalences et les réseaux minimaux que je viens de donner paraissent évidentes, elle me posent problème : dans certaines classes d'équivalence en effet, réduire une fonction de paramètres quelconques à la fonction minimale correspondante amènera à affecter la valeur 0 à certains poids ! Par exemple, la SBF de dimension 2 et de paramètres $(-2, -1)$, appartenant à la classe d'équivalence $A_4(f)$ dans la figure 8.31 et dont la sortie est toujours 0, pourra en réalité être réduite à la SBF minimale de paramètres $(0, 0)$. Du point de vue de la logique des réseaux, ce n'est pas choquant. Du point de vue de la biologie et de sa modélisation en revanche, je trouve cela très gênant pour les raisons que je vais décrire tout de suite.

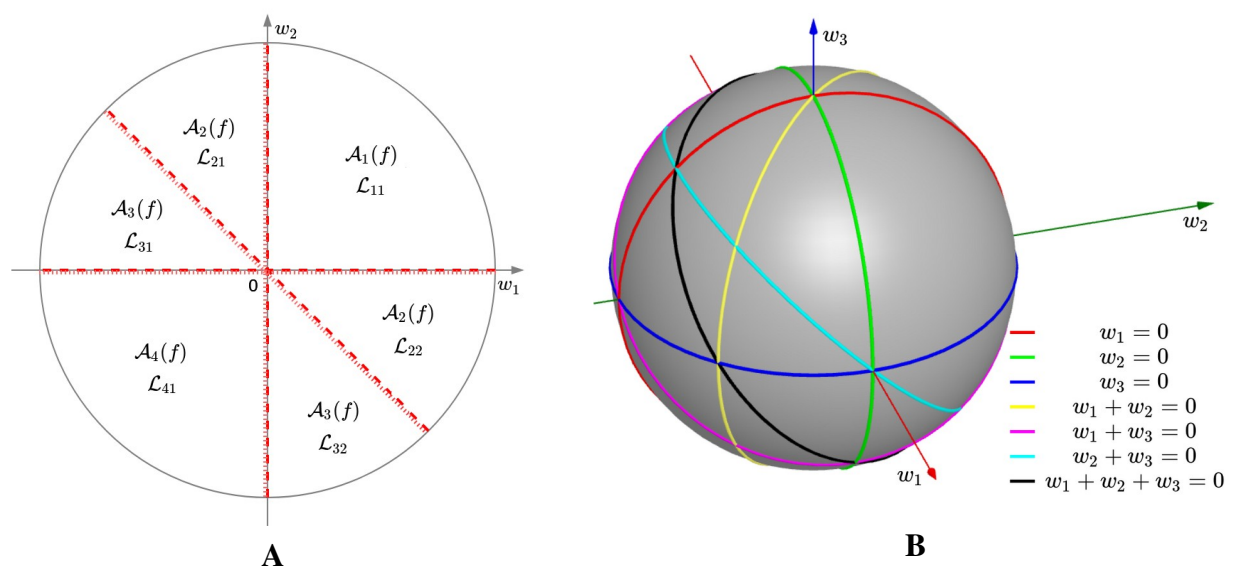


Figure 8.31. Boule des paramètres et classes d'équivalence. (A) Les SBFs en 2 dimensions sont représentées par leurs 2 paramètres w_i et w_j . L'espace des paramètres peut alors être partitionné par les différentes équations des plans dont les poids sont les variables, en partitions coniques L_{ij} (Layouts). Dans chaque layout, toutes les fonctions Booléennes ont le même comportement (même table de vérité). Dans ces 2 représentations, on délimite l'espace dans une d -boule pour des raisons pratiques que nous aborderons lorsque nous parlerons des calculs de complexité. Cette d -boule possède des symétries correspondant aux permutations des entrées dans la table de vérité, ce que nous désignons par L_{ij} . Une classe d'équivalence $A_i(f)$ est formée de l'ensemble des agencements L_{ij} qui la composent. En (B) est montré la 3-boule des SBFs de dimension 3, ainsi que l'ensemble des plans délimitant les différents agencements d'entrées de f formant les classes d'équivalence.

En modélisant les réseaux biologiques par des réseaux (Booléens ici), on souhaite exprimer le fait qu'une voie de communication, même si elle n'a pas d'effet, existe entre deux éléments d'un système. Si par cette voie de communication l'élément source parvient éventuellement à provoquer parfois des changements de l'état de l'élément cible, alors on pourra dire qu'il s'agit d'une interaction effective. Par exemple, un facteur de transcription (émis par une source) se fixant à un gène et provoquant un changement de l'activité de ce gène permettra au biologiste d'identifier une interaction effective et cela pourra être modélisé par un arc de poids non nul. Ce qui pose problème, ce sont les interactions que je qualifierais de « silencieuses ». Supposons une molécule pouvant se fixer sur une région régulatrice inhibitrice en amont d'un gène mais ne provoquant aucun

changement de son activité ; cela serait imaginable dans le cas d'une mutation de la région promotrice découplant celle-ci de sa région sensée être régulatrice. Il existe un canal de communication, mais il est inopérant. Là, il y a trois manières de concevoir la modélisation de cette situations :

- La première, pragmatique, aura pour objectif de représenter ce que l'on peut mesurer d'effectif, ce qui est apparemment à l'œuvre dans la cellule, et ne prendra donc en compte que les interactions effectives, celles qui peuvent éventuellement provoquer un changement d'état d'un nœud. Dans ce cas, les interactions inopérantes ne seront pas prises en compte : leur poids sera nul et on ne représentera pas l'arc correspondant étant donné qu'on n'a de toutes façon pas identifié de canal de communication potentiel.
- Dans la seconde, il faut supposer que le biologiste a pu identifier une molécule provenant d'une source et se fixant sur une cible, mais qu'il n'a pas pu identifier de changement d'activité de la cible lié à la fixation de cette molécule. Dans ce cas, une possibilité est de considérer un arc existant entre le nœud source et le nœud cible mais de lui affecter un poids nul. On introduit ici une interaction potentielle. Cependant, comment alors distinguer dans l'ensemble des interactions potentielles d'un réseau, exprimées dans la matrice d'interaction, les arcs de poids nul des absences d'arcs (*de facto* de poids nul aussi dans la matrice d'interaction) ?
- La troisième possibilité consiste à signifier une interaction et sa nature (*a priori* inhibitrice ou activatrice) par un arc modélisant le canal de communication inopérant et indiquant par un poids non nul la nature de cette interaction. C'est cette façon de procéder que je souhaite retenir dans mon travail sur les réseaux.

Ces trois façons de procéder ne sont pas équivalentes. En particulier, le fait de modéliser un canal de communication inopérant permet de prendre en compte des potentialités, évolutives notamment, d'un réseau. Un arc ne causant aucun changement sur l'activité du réseau pourrait, suite à un processus évolutif modifiant ce réseau, être amené à jouer un rôle important dans le réseau évolué. L'importance de cette interaction pourrait ainsi être amenée à changer, par exemple faisant passer une fonction à signes de paramètres $(-1,2)$ à la fonction de paramètres $(-2,2)$. Dans le premier cas, lorsque les 2 nœuds source sont actifs, la cible passe à 1 ; en revanche, dans le deuxième cas, la double stimulation de la cible fait passer celle-ci à 0. L'ajout d'un nouvel arc entrant lors de cette évolution pourrait aussi donner un nouveau rôle à l'arc encore inutile, révélant ainsi sa potentialité. Par exemple, l'ajout d'un arc de poids négatif (ex : -1) à la fonction de paramètres $(-1,2)$, transformant ainsi celle-ci en une nouvelle fonction $(-1,-1,2)$, change le comportement de cette dernière puisque l'action combinée des 2 arcs inhibiteurs peut maintenant compenser l'action de l'arc activateur, ce qui n'était pas le cas avec un seul arc inhibiteur.

De plus, dans le cadre d'une modélisation non déterministe d'un réseau, comportant une activité au moins partiellement stochastique, la présence d'un arc inactif pouvant éventuellement s'activer de temps à autres change la donne.

Enfin, la modélisation de réseaux biologiques se fait en délimitant le système d'intérêt de façon

assez arbitraire, souvent par manque de connaissances de ce qui l'entoure, mais aussi par choix de ne pas se retrouver de fil en aiguille à modéliser la totalité de la cellule dans un gigantesque réseau. De la même manière que pour la question évolutive, le plongement d'un tel réseau comportant des potentialités exprimées sous la forme d'arcs inopérants dans un réseau de plus grande taille et établissant un dialogue avec celui-ci, n'aura pas les mêmes conséquences que si de tels arcs inopérants mais bien présents n'étaient pas pris en compte.

Nous verrons que cette potentialité ainsi exprimée trouvera tout son intérêt lorsque nous parlerons de l'extensibilité.

Nouveau partitionnement de \mathcal{W} en dimensions pures – Nouvelle définition des classes d'équivalence. Le simple partitionnement de \mathcal{W} en layouts par les différents hyperplans $w_i + \dots = 0$ ne permet pas, nous l'avons vu, de distinguer les fonctions se trouvant sur les axes, donc possédant un à plusieurs poids nuls, des fonctions se situant à un endroit quelconque de \mathcal{W} , hors axes (ex : la fonction $(-1, -1)$, avec un tel partitionnement, se trouve dans la même classe d'équivalence que $(0, 0)$). Il faut donc apporter une distinction plus fine des SBFs dans \mathcal{W} .

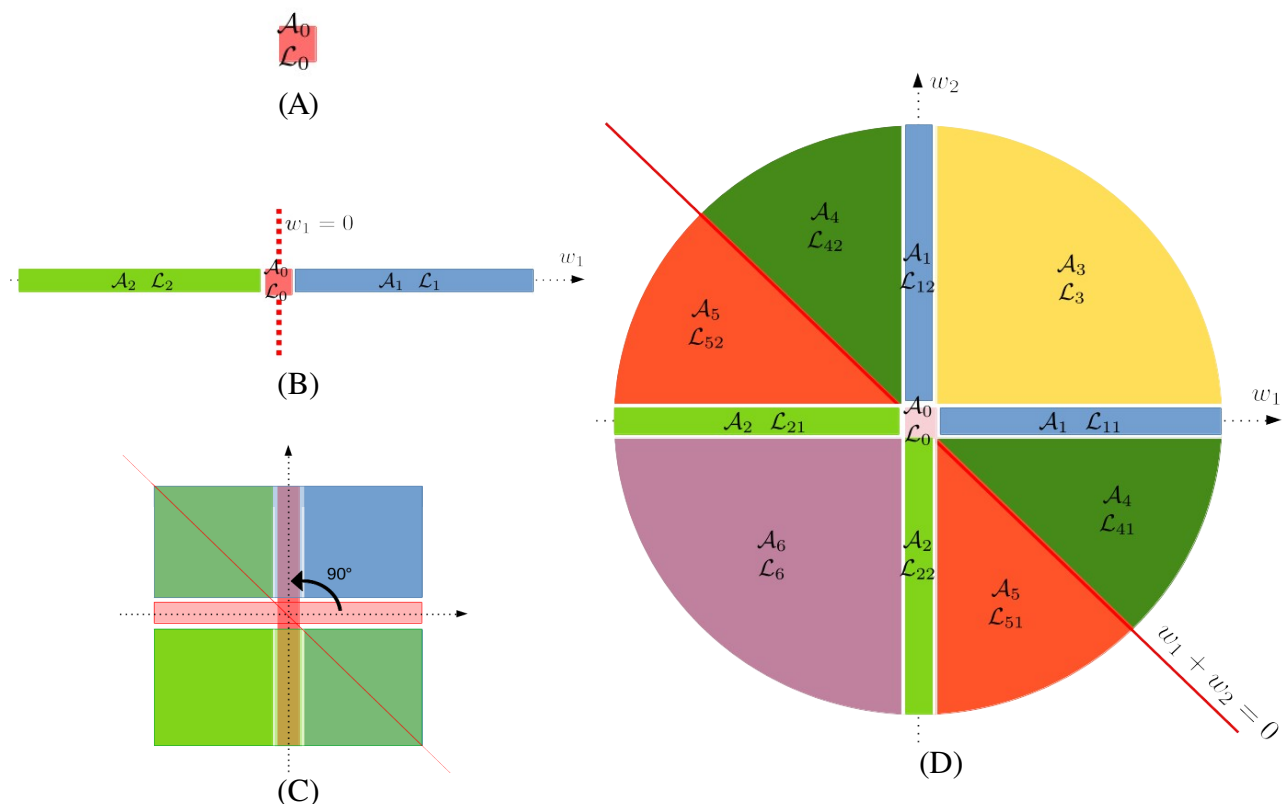


Figure 8.32. Partitionnement des espace de paramètres W_q en layouts et classes d'équivalence. (A) L'espace W_0 de dimension 0 est partitionné en 1 cône de dimension 0. Il ne contient qu'une classe d'équivalence. (B) L'espace de dimension 1 est partitionné en 3 cônes, 1 de dimension 0, 2 de dimension 1. Le partitionnement de W_1 aboutit donc à 3 classes d'équivalence. (C) Le partitionnement de W_2 est obtenu d'une part en prolongeant le partitionnement de W_1 dans une dimension supplémentaire et en le superposant à lui même après rotation de 90° , d'autre part en partitionnant à nouveau avec l'hyperplan d'équation $w_1 + w_2 = 0$. (D) L'espace de dimension 2 est partitionné en 11 cônes, 1 de dimension 0, 4 de dimension 1, 6 de dimension 2. Par symétrie par rapport à $w_1 = w_2$, les layouts sont regroupés dans une seule et même classe d'équivalence. Le partitionnement de W_2 aboutit donc à 7 classes d'équivalence.

L'espace \mathcal{W}_d de dimension d contient des fonctions à d entrées non nulles, mais également des fonctions à $d-1$ entrées non nulles (donc une entrée nulle), $d-2$ entrées non nulles (donc 2 entrées nulles) jusqu'à $d-k=0$ (aucune entrée). Il est donc plus judicieux de décrire comment est formé cet espace \mathcal{W}_d de dimension d , à savoir comme une composition d'espaces de fonctions de dimensions inférieures. Ainsi (voir en parallèle la figure 8.32) :

- L'espace \mathcal{W}_0 de dimension 0 est un point qui ne contient qu'une fonction sans entrées : (\emptyset) . Il est donc formé d'un seul layout et d'une seule classe d'équivalence.
- L'espace $\mathcal{W}_1 \setminus \mathcal{W}_0$ (c'est-à-dire \mathcal{W}_1 privé de \mathcal{W}_0) est constitué de 2 segments de droite $] -\infty, 0[\cup] 0, +\infty[$, séparés par un hyperplan de dimension 0, d'équation $w_1=0$, passant par 0 et orthogonal à la droite. \mathcal{W}_1 contient donc 2 layouts de dimension 1 pure (2 classes d'équivalences) et 1 layout de dimension 0 (1 classe d'équivalence).
- L'espace $\mathcal{W}_2 \setminus \{\mathcal{W}_0, \mathcal{W}_1\}$ (c'est-à-dire \mathcal{W}_2 privé de \mathcal{W}_0 et \mathcal{W}_1) est un espace plan, dont on a retranché les axes et l'origine. De plus, un hyperplan de dimension 1 (une droite) d'équation $w_1+w_2=0$ partitionne de façon supplémentaire cet espace. Le partitionnement de \mathcal{W}_1 s'applique sur chacune de dimensions de \mathcal{W}_2 . \mathcal{W}_2 est donc partitionné en 6 layouts (regroupés en 4 classes d'équivalence) de dimension 2 pure, 4 layouts de dimension 1 pure (regroupés en 2 classes d'équivalence) et 1 layout de dimension 0 (1 classe d'équivalence).
- L'espace $\mathcal{W}_3 \setminus \{\mathcal{W}_0, \mathcal{W}_1, \mathcal{W}_2\}$ (c'est-à-dire \mathcal{W}_3 privé de \mathcal{W}_0 , \mathcal{W}_1 et \mathcal{W}_2) est constitué de la même manière par un espace de dimension 3 privé des plans formés par l'application du partitionnement de \mathcal{W}_2 se traduisant par les plans $w_1=0$, $w_2=0$ et $w_3=0$ en rouge, vert, bleu respectivement dans la 3-boule montrée figure 8.31, ainsi que les plans $w_1+w_2=0$, $w_1+w_3=0$, $w_2+w_3=0$ montrés en jaune, magenta et cyan qui correspondent à l'application de l'hyperplan d'équation $w_1+w_2=0$ partitionnant \mathcal{W}_2 , et d'un partitionnement supplémentaire par un hyperplan de dimension 2 d'équation $w_1+w_2+w_3=0$ en noir dans la figure 8.31. Cet espace \mathcal{W}_3 est donc partitionné en 32 layouts de dimension 3 (regroupés en 10 classes d'équivalence), en 6 layouts (regroupés en 4 classes d'équivalence) de dimension 2 pure, 4 layouts de dimension 1 pure (regroupés en 2 classes d'équivalence) et 1 layout de dimension 0 (1 classe d'équivalence), soit un total de 43 layouts regroupés en 17 classes d'équivalence.
- Et ainsi de suite pour les dimensions supérieures.

Maintenant, l'équivalence des fonctions est définie à la fois grâce à la table de vérité et par le nombre d'entrées non nulles. Cela aura des conséquences sur la complexité des fonctions.

La façon dont sont partitionnés les espaces de paramètres semble clairement définissable par l'algèbre linéaire. Nous sommes tout juste en train de nous y intéresser. L'application de l'algèbre linéaire dans le cadre des espaces de paramètres des SBNs pourrait bien nous aider à résoudre certaines de nos questions comme celle de la composition de fonctions.

8.3.3.2.f. Graphe de transition

Le graphe de transition d'un réseau désigne l'ensemble des états que peuvent prendre les nœuds de ce réseau et décrit les transitions que leur activité provoque. Dans le cas des réseaux déterministes de type PBNs, TBNs ou SBNs, ces graphes sont déterministes. Un graphe de transition pour ces réseaux décrit complètement et de façon unique la structure du réseau puisqu'il impose sa dimension et, en faisant correspondre à chaque état du réseau un nouvel état déterministe, il reproduit la table de vérité de ce réseau (figure 8.33).

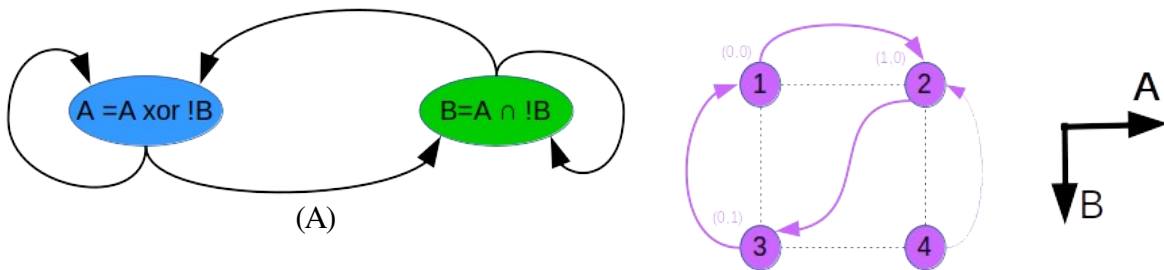


Figure 8.33. Exemple de réseau et de son graphe de transition associé (A) Exemple de réseau propositionnel composé de 2 fonctions *complexes*. (B) Le graphe de transition est composé de 4 états (labels dans les nœuds du graphe dont les valeurs des nœuds du réseau sont indiquées entre parenthèses). La table de vérité correspondante est indiquée en dessous avec les valeurs d'entrées A_i et B_i , et les valeurs en sortie A_s et B_s . Les états sont numérotés ainsi que leurs successeurs.

A_i	B_i	A_s	B_s	stateId	successorId
0	0	0	1	1	2
0	1	0	1	2	3
1	0	0	1	3	1
1	1	0	1	4	2

(B)

8.3.3.2.g. Modes de mise à jour

Les modes (ou schémas) de mise à jour correspondent à l'ordre dans lequel les nœuds du réseau sont mis à jour. Un schéma de mise à jour peut ainsi être entièrement parallèle sur l'ensemble du réseau, ou entièrement séquentiel (aucun des nœuds n'est mis à jour en même temps qu'un autre ; tous ont un numéro de mise à jour différent). Il peut aussi être mixte et comporter des blocs d'exécution parallèle, eux-mêmes exécutés séquentiellement [Elena, 2009].

Dans l'ensemble des réflexions qui suivront (et en pratique dans nos travaux) je ne me placerai que dans le cadre de schémas d'exécution parallèle. Cependant, ces réflexions sont transposables aux autres schémas de mise à jour, ceux-ci apportant leur dose de complexité. En effet, tout schéma de mise à jour asynchrone dans un réseau peut être traduit sous la forme d'une mise à jour synchrone dans un autre réseau de plus grande dimension. L'asynchronie est une question de temps de communication entre les parties du réseau, délimitant ainsi des blocs d'exécution parallèle et un ordre d'exécution de ces blocs. J'y reviendrai en abordant la question de la décomposition des réseaux (section 8.3.5.2.a).

8.3.4. Inférence et exploration des TBNs et SBNs

Je vais maintenant aborder d'une part la façon dont nous cherchons des réseaux présentant certaines caractéristiques, mais surtout je vais doucement introduire l'idée de considérer des populations de

réseaux satisfiables plutôt que des instances particulières de réseaux pour répondre à un problème. Lorsqu'on souhaite modéliser les relations qui harmonisent le fonctionnement d'un système biologique par un réseau d'interaction, on doit identifier les protagonistes qui constituent les nœuds du réseau (telle ou telle espèce animale, un ensemble de gènes, *etc.*) et idéalement leurs caractéristiques comportementales propres, ainsi que l'ensemble des interactions que chacun d'entre eux reçoit directement ou indirectement des autres. Ce faisant, on cherche ainsi à identifier la structure de ce réseau d'interaction (le graphe d'interactions). On peut aussi s'intéresser à l'activité globale (ou seulement une activité partielle) de ce réseau (son graphe de transition), ce qu'il produit d'observable, comme des signes de présence de telle ou telle espèce animale dans un milieu, un dénombrement d'individus, ou un phénotype lorsqu'on étudie l'activité d'un ensemble de gènes.

Souvent, les informations souhaitées sont parcellaires. Par exemple, on peut savoir qu'un réseau de régulation génétique comporte certains gènes non identifiés mais requis pour produire une activité particulière ; même leurs interactions peuvent ne pas être totalement connues. Ou bien, on ne connaît que certaines parties du graphe de transition, celles qui ont été révélées par les recherches patientes et minutieuses des expérimentateurs et observateurs du système naturel. Il est pourtant nécessaire, notamment pour simuler son fonctionnement et éventuellement s'en servir dans un cadre élargi, par exemple la simulation de cellules virtuelles dotées de capacités de régulation, mais aussi pour aider à la réflexion sur ce système (voir la section 3.1.3), de disposer d'une version fonctionnelle d'un tel réseau d'interactions, donc de compléter ce qui manque. En pratique, pour les raisons évoquées ci-dessus (section 8.3.3.2.f) quels que soient les efforts déployés pour affiner la connaissance, un réseau ne sera jamais qu'une projection incomplète (par son formalisme mais aussi par le choix de la délimitation du système biologique à modéliser) du système naturel.

En pratique, le processus de construction d'un tel réseau fonctionnel, passe par des suppositions faites par le chercheur sur les parties manquantes et les approximations qui sont soit présentées comme des hypothèses à tester, soit retenues en l'état comme acceptables car complétant de façon efficace les connaissances manquantes ; dans tous les cas ces compléments proviennent d'inférences produites par l'esprit de celui qui s'intéresse au système naturel depuis un temps suffisant pour élaborer des projections possibles. Cependant, au terme de cette construction, un seul modèle est retenu si tant est qu'il satisfasse (acceptablement) aux observations connues sur la structure et la dynamique du système étudié. C'est une approche pragmatique car il est alors aisé, pour l'étude biologique, de travailler sur un modèle bien identifié et éventuellement de le rejeter au terme de nouvelles observations entrant en contradiction avec sa structure ou son activité. Dans ce cas, le modèle est révisé ou on recherche un autre modèle.

C'est là que ma façon de voir les choses diffère. Avant d'expliquer en quoi mon point de vue diffère, je vais parler un peu de la détermination des réseaux satisfaisant les observations.

Étant donné un ensemble de connaissances (partielles) sur la structure et la comportement d'un réseau, il peut exister entre 0 (aucun modèle) et un nombre proprement gigantesque de modèles (des projections possibles du système naturel dans un formalisme donné). L'ensemble des modèles qui satisfont (au moins) l'ensemble des contraintes que constituent les connaissances dont on dispose est appelé *ensemble satisfiable*. Le travail du théoricien consiste notamment à élaborer des méthodes

efficaces pour déterminer de tels ensembles satisfiables. Dans ce contexte, le travail du biologiste consiste alors à réduire la taille de tels ensembles en apportant de nouvelles connaissances, donc en contraignant davantage la recherche des réseaux satisfiables.

Une des façons de procéder consiste à énumérer les variations de réseaux et à retenir celles possédant au moins les caractéristiques observées. Il va sans dire qu'une telle recherche devient vite explosive quand la taille des réseaux augmente : s'il n'y a que 60 SBNs de dimension 2 (donc composés de 2 fonctions à 2 entrées), il y en a environ 10^{20} dès la dimension 5 ! Le tableau 8.34 ci-dessous récapitule ce dénombrement des SBNs (dénombrement approximatif car ne prenant pas en compte le rejet des réseaux non connectés) obtenu par la formule

$$N_{SBNs} = \frac{(f + d - 1)!}{(f - 1)!} \times d!^{d-1} = \left(\prod_{i=f}^{f+d-1} i \right) \times d!^{d-1} \text{ avec } d \text{ la dimension et } f \text{ le nombre de}$$

classes d'équivalence (c'est-à-dire le nombre de fonctions minimales, aux layouts près).

Dimension d	Nombre de Layouts (partitions de l'espace des paramètres des SBFs) de dimension d pure	Nombre de SBFs minimales f de dimensions d pures (classes d'équivalence pures)	Nombre de SBFs minimales f de dimensions d cumulées (classes d'équivalence cumulées)	Nombre de SBNs de dimension d - y compris non connectés et redondants par permutation - non redondants
0	1	1, sans paramètres (pas d'arc entrant) : (\emptyset)	1	1
1	2	2, de paramètres (1) ou (-1)	3, de paramètres (0), (1), (-1)	3
2	6	4, de paramètres (1,1), (2,-1) ou (-1,2), (1,-1) ou (-1,1), et (-1,-1)	7, de paramètres (0,0), (0,1) ou (1,0), (0,-1) ou (-1,0), (1,1), (2,-1) ou (-1,2), (1,-1) ou (-1,1), et (-1,-1)	<i>112</i> 60
3	32	10	17	<i>209 304</i> 31 199
4		36	53	<i>1.22 10¹¹</i> 3 623 481 898
5		224	277	<i>3.5 10²⁰</i> ?
6		3024	3301	<i>2.5 10³⁵</i> ?

Tableau 8.34. Énumération des layouts, classes d'équivalence en dimension pure ou cumulées, et des SBNs, en fonction de la dimension du réseau. Le nombre de réseau indiqué en italique est celui obtenu par le calcul donné plus haut ; celui donné en caractères droits est le nombre exact de réseaux non redondants obtenu par calcul.

Une autre façon de procéder consiste à utiliser la programmation logique par contrainte : au lieu de

parcourir l'espace de tous les réseaux possibles et de ne conserver que ceux qui sont satisfiables, on réduit d'emblée l'espace de recherche à la taille de l'ensemble satisfiable (il n'y a donc pas de recherche à faire dans cet ensemble ainsi réduit) en contraignant cet espace par des contraintes logiques, traduction directe des connaissances requises sur la structure et sur la dynamique. Brièvement, le principe est le suivant : les connaissances biologiques sont traduites sous forme de contraintes logiques se traduisant par exemple par des inégalités sur des entiers pour contraindre les valeurs des poids des arcs entrants d'une fonction Booléenne. L'ensemble des contraintes est traduit par un programme, le *grounder*, en forme normale contrainte (CNF), une formulation globale du problème. La taille de cette formule, donc son occupation en mémoire dans l'ordinateur, dépend du nombre de contraintes et de leur expression. C'est la limitation de cette méthode : elle peut vite s'avérer très gourmande en mémoire. Un solveur de type SAT se charge enfin de déterminer l'ensemble satisfiable à partir de la CNF. Pour plus de détails, je renvoie aux articles cités ci-dessous.

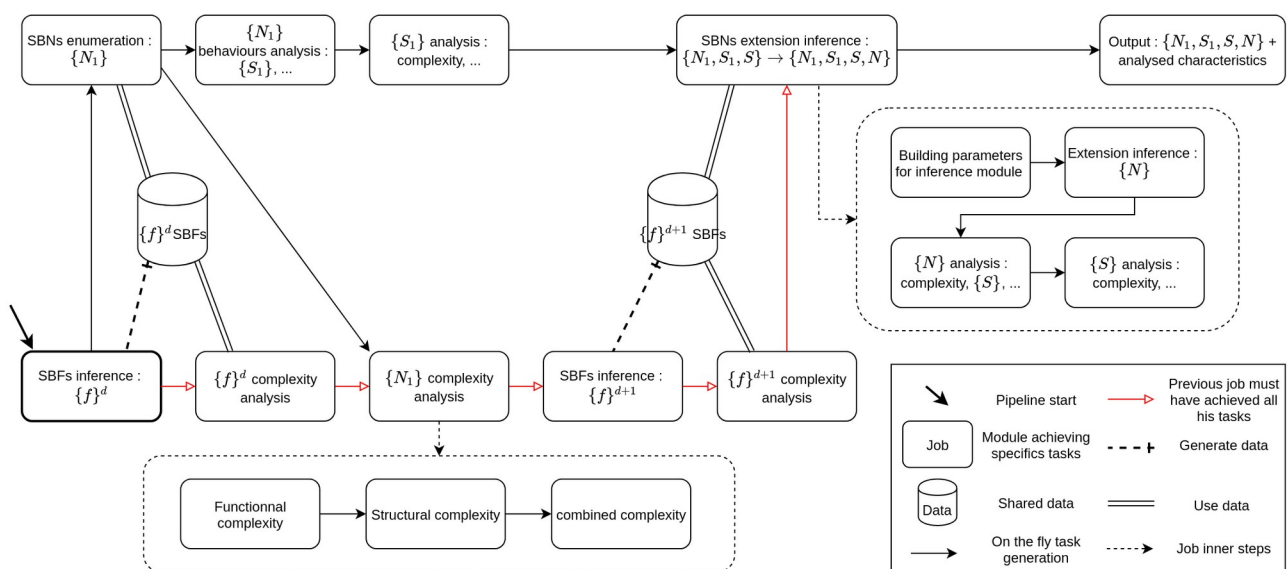


Figure 8.35. Chaîne logicielle de traitement des SBNs utilisant la bibliothèque *Java Parallel Pipeline*.

En 2009, F. Corblin, alors en thèse avec E. Fanchon et L. Trilling dans notre équipe à TIMC-IMAG a montré que l'on pouvait utiliser une approche déclarative implémentée sous la forme de programmation logique par contrainte, en Prolog d'abord, en ASP⁴⁰ [Gebster, 2011] plus tard, pour inférer des ensembles de réseaux Booléens satisfiables, ou des propriétés les concernant, à partir de connaissances structurales et dynamiques (comportement) [Corblin, 2009, 2010, 2011, 2012]. Le formalisme utilisé était alors celui des réseaux de Thomas. En 2013, H. Ben Amor, alors en thèse avec J. Demongeot et moi, et F. Corblin, avons adapté cette méthode pour les réseaux Booléens à seuil (TBNs) [Ben Amor, 2013]. Nous l'avons également appliquée à l'inférence des réseaux pouvant modéliser le réseau de régulation de la morphogenèse florale d'*Arabidopsis thaliana* : résultat, 532 TBNs satisfiables (tous différents, tous minimaux) au lieu du modèle unique retenu par

⁴⁰ Answer Set Programming

Mendoza, l'auteur à l'origine de ce travail de modélisation [Mendoza, 1998]. Par la suite, L. Trilling, S. Ivanov et moi-même, avec l'aide précieuse de plusieurs étudiants (Q.T. Vuong, U. Christen, R. Segretain), avons amélioré la technique en utilisant le langage ASP, un langage de programmation par contrainte avec logique non monotone, dialoguant avec un langage impératif (Java) via un pipeline de traitement optimisé [JPP], intégrant la limitation aux réseaux minimaux, la non-redondance des réseaux par permutation, mais également des traitements *a posteriori* comme le calcul des graphes de transition, les calculs de complexité, *etc*, pour traiter ce type de problèmes sur les TBNs [Vuong, 2017 ; Christen, 2019] puis les SBNs [Segretain, 2020a et 2020b] (figure 8.35).

Le nombre de réseaux satisfiables pour une connaissance biologique donnée (structure + dynamique) peut être potentiellement très grand ; par exemple, rappelons les 532 modèles satisfiables pour le réseaux de morphogenèse florale d'*Arabidopsis thaliana* obtenus à partir des connaissances biologiques énoncées par Mendoza *et al* [Mendoza, 1998 ; Ben Amor, 2013]. Rien ne permet, sans plus de contraintes sur la structure ou la dynamique, d'affirmer qu'un réseau vaut mieux qu'un autre ! Pourquoi dès lors ne pas essayer de considérer plutôt l'ensemble satisfiable comme un modèle en soit ? Le comportement du système serait alors vu comme un comportement moyen de l'ensemble des modèles satisfiables. J'ai avancé cette idée de système biologique simulé par un ensemble de modèles vers 2009 alors que nous commençons à travailler sur l'apport de l'approche déclarative pour inférer des ensembles de réseaux, et nous en avons déjà touché un mot dans [Ben Amor, 2011, 2013]. Cette idée a en fait été proposée dès 2004 par un des pères des réseaux Booléens, S. Kauffman [Kauffman, 2004]. Elle a été utilisée pour étudier statistiquement la dynamique collective d'un ensemble de modèles [Krawitz, 2007 ; Schweiger, 2017]. Récemment, Chevalier *et al* [Chevalier, 2020] ont proposé une stratégie de simulation de réseaux Booléens fondée sur des ensembles de réseaux satisfiables déterminés par inférence sur les connaissances *a priori* sur la structure et la dynamique des réseaux, en ASP, d'une manière très semblable à la notre. Je reviendrai par la suite sur cette vision ensembliste de la dynamique cellulaire à laquelle j'adhère puisqu'elle permet de rendre compte des variations possibles du modèle décrivant à un instant donné l'agencement des éléments cellulaire et leur état, soit sous la forme d'une dynamique moyenne résultant de l'activité superposée de l'ensemble des réseaux satisfiables, soit comme une trajectoire dans un paysage de réseaux, dont certains sont satisfiables. Je développerai particulièrement ce dernier point en incluant les notions de robustesse des comportements et de résilience.

8.3.5. Premières questions sur les réseaux (SBNs)

A présent, je vais énumérer un certain nombre de questions que nous nous posons sur les systèmes biologiques et que nous permettent d'explorer les techniques et les formalismes mentionnés plus haut. Dans la section qui suivra et qui sera dédiée à des réflexions plus larges sur la nature du vivant et son fonctionnement, ces questions trouveront une résonance.

8.3.5.1. Complexité - Relation structure-fonction

Lorsqu'on travaille sur des ensembles de réseaux, on est tenté de les comparer entre eux et de comparer leur dynamiques. Disposer de mesures permet leur classification, d'établir des proximités entre réseaux et/ou comportements et d'étudier la relation structure-fonction de ces réseaux.

Il y a de très nombreux indicateurs disponibles permettant de caractériser les réseaux et leurs dynamiques, des très simples comme le nombre de nœuds et d'arcs, le nombre de bassins d'attraction et la taille respective de ces bassins [Glade, 2011], d'autres plus compliqués demandant des calculs plus lourds comme la centralité.

Les réseaux sont des machines (leurs abstractions mathématiques) qui réalisent un travail. Dans la vie de tous les jours, nous évaluons la complexité des machines à notre portée (bien que, souvent, leur apparence extérieure cache leur puissance et leur complexité réelle) tant dans leur organisation interne que dans la ou les fonctions qu'elles réalisent. Cette évaluation nous permet de les comparer entre elles ... et avec nous-mêmes, avec les choses vivantes, ces semblants de machines complexes.

Il y a de nombreuses façons de mesurer la complexité d'un objet. En s'inspirant de méthodes existantes et en réfléchissant à ce qu'elle devait exprimer, nous avons développé une manière de calculer la complexité des réseaux Booléens à signe et nous avons aussi calculé la complexité de leurs dynamiques en utilisant une technique existante. Je renvoie aux textes [Christen, 2018, 2019 ; Segretain 2020a et 2020b] pour les détails. Nous avons aussi développé une manière de calculer des distances entre réseaux en tenant compte à la fois de leur structure et de leur dynamique. Dans les deux cas, notre perception des critères à retenir pour définir un réseau de façon synthétique ou le comparer à d'autre évolue. Je discute notamment de cela dans cette section.

Complexité d'un SBN – Code offusqué. Brièvement, ce que nous retenons, c'est que le calcul de la complexité d'un SBN repose à la fois sur la complexité intrinsèque des SBFs qui le composent (une complexité dite fonctionnelle C^f) et sur la complexité de leurs arrangement entre elles (complexité dite structurale C^s).

Nous avons choisi de définir la complexité fonctionnelle C^f d'une SBF f comme la probabilité complémentaire à celle d'être dans sa classe d'équivalence $\mathcal{A}(f)$, parmi l'ensemble des classes d'équivalence de l'espace des paramètres, c'est-à-dire $C^f = 1 - P(\mathcal{A}(f))$. La probabilité de choisir une fonction appartenant à une classe d'équivalence donnée dans l'espace des paramètres est égale au volume occupé par cette classe d'équivalence dans la boule unité, divisé par le volume de la boule unité. Pour le moment, ces volumes associés aux classes d'équivalence sont évalués par un grand nombre de tirages aléatoires de vecteurs de paramètres dans la boule unité, avec une norme infinie (figure 8.36). Dans cette logique, plus une SBF est représentée (classe d'équivalence volumineuse), moins elle est complexe ; et réciproque. C'est la *rareté* qui fait la complexité ici. Derrière cela, il y a une notion de coût : il est difficile d'obtenir, donc coûteux de rechercher dans l'espace des paramètres, une fonction disposant de certains paramètres de la même manière qu'il est coûteux de fabriquer une pièce mécanique « complexe ». Je reviens en détails sur ceci dans le paragraphe *Mesurer la complexité, calculer une distance, c'est déjà modéliser* 3 pages plus loin.

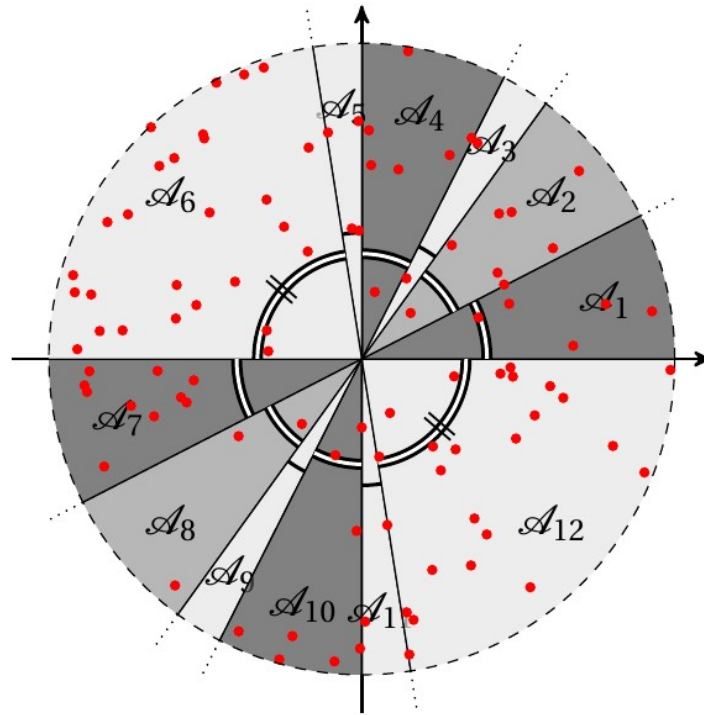


Figure 8.36. Tirages aléatoires dans la boule unité pour déterminer la proportion respective de chaque classe d'équivalence. Voir l'article [Christen, 2019] et le rapport [Christen, 2018] pour l'algorithme détaillé et les discussion sur le choix de la norme.

Pour la complexité structurale C^s , nous utilisons une forme de mesure de centralité de manière à exprimer l'influence qu'a un nœud (sous tendant une fonction) sur un autre, par l'intermédiaire des arcs dirigés et pondérés. Là aussi, cette complexité est définie comme une probabilité. Pour une étape d'un chemin comportant un arc dirigé $E_{ij} : n_i \xrightarrow{w_{ij}} n_j$, entre deux nœuds du réseau nous calculons la probabilité d'influence du nœud i sur le nœud j comme la proportion d'influence $P^I(E_{ij})$ que reçoit n_j par n_i parmi l'ensemble des influences E_{kj} qu'il reçoit :

$$P^I(E_{ij}) = \frac{|w_{ij}|}{\sum_{k=1}^d |w_{kj}|} \quad \forall i, j, k \in [1; d]$$

On peut alors définir la probabilité d'influence $P^I(Path_{ab})$ d'un chemin (*path*) plus long entre deux nœuds distants n_a et n_b comme la moyenne géométrique des influences de chaque étape

$$E_{ij} \in Path_{ab} : P^I(Path_{ab}) = \sqrt[L_{ab}]{\prod_{E \in Path_{ab}} P^I(E)}, \quad L_{ab} \text{ étant la longueur du chemin. On exprime}$$

là l'influence moyenne d'un chemin comme une probabilité moyenne afin que celle-ci puisse être comparée à celle d'un autre chemin. L'usage de la moyenne géométrique est notamment justifié par le fait que chaque SBF reçoit potentiellement un nombre différent d'entrées non nulles ; ainsi, l'influence $|w_{ij}|$ à chaque étape sur un nœud peut être normalisée par un nombre différent de poids (allant de 1 à d poids). Notons que les autres entrées que reçoit une fonction sont des compétitrices de l'entrée considérée ; elles contribuent à faire baisser le niveau de centralité d'une fonction.

Comme plusieurs chemins différents peuvent exister entre les nœuds n_a et n_b , la probabilité d'influence globale P_{ab}^I de n_a sur n_b est calculée comme la probabilité d'influence de l'union des chemins : $P_{ab}^I = P\left(\bigcup_{Path \in Paths_{ab}} Path\right)$ où $Paths_{ab}$ désigne l'ensemble des chemins entre n_a et n_b .

La centralité d'un nœud n_i (complexité structurale locale) est alors définie comme la probabilité que celui-ci influence au moins un nœud n_k , lui inclus : $C_i^s = P\left(\bigcup_{k=1}^d \left(\bigcup_{Path \in Paths_{sa}} Path\right)\right)$

Pour calculer la complexité globale (synthétique) d'un réseau SBN, on utilise les 2 complexités, structurale C^s et fonctionnelle C^f à la fois. La centralité C_i^s de chaque nœud module sa complexité fonctionnelle C_i^f , de telle manière que la complexité d'un réseau N est obtenue comme suit :

$$C(N) = \prod_{i=1}^d \begin{cases} C_i^f \times C_i^s & \text{si } C_i^s > 0 \\ 1 & \text{sinon} \end{cases}$$

Complexité d'une « musique du réseau » (comportement unitaire). Commençons par définir ce que nous appelons « comportement unitaire » ou « musique » ici. La dynamique des SBNs génère des états attracteurs, soit sous forme de cycles limites discrets (donc passant par plus d'un état du réseau), soit sous forme de points fixes (des cycles de période 1). Les états qui ne font pas partie des attracteurs sont des états transitoires qui, normalement (sauf erreur du réseau pouvant survenir de façon stochastique par exemple), ne sont plus visités une fois passés. Il est alors cohérent d'associer un comportement biologique à un état attracteur du réseau.

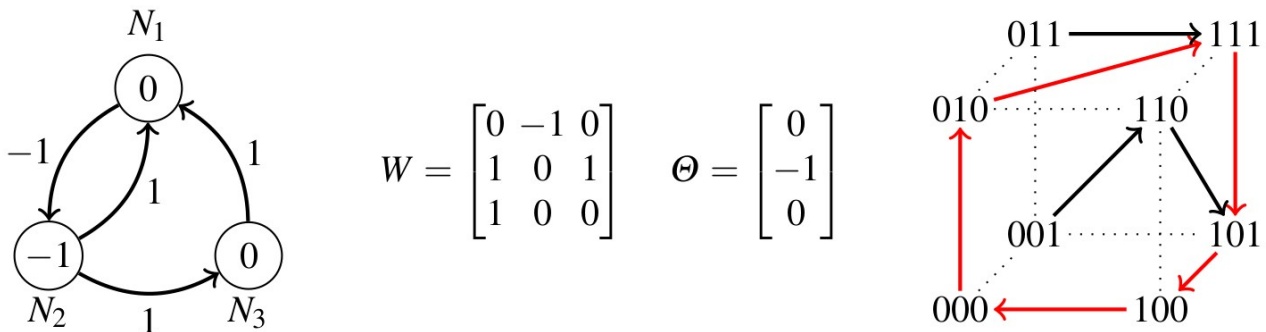


Figure 8.37. Un TBN, ses paramètres (poids et seuils) et son graphe de transition. Le graphe de transition indique en chaque nœud du graphe les valeurs des états des nœuds (N_1, N_2, N_3). Ce réseau possède un unique attracteur indiqué en rouge. Les autres transitions indiquées en noir sont transitoires.

Nous avons cependant décidé de détailler les comportements nœud par nœud. En effet, la réalisation d'une fonction biologique pourrait en effet n'être vue que comme la manifestation d'une succession d'états d'un seul nœud. Il est donc utile d'avoir une description à ce niveau. Nous parlerons alors d'un comportement unitaire, ou plus simplement de « musique », comme la succession périodique (de période comprise entre 1 et 2^d) des états d'un nœud (*joués* par un nœud), ces états appartenant nécessairement à un attracteur du réseau, par exemple la musique $(10011)^*$ jouée par le nœud N_1 du

réseau TBN montré figure 8.37. Rien n'empêche par contre de considérer une fonction biologique nécessitant plusieurs musiques (sur autant de nœuds) jouées en même temps.

La complexité de Kolmogorov [Kolmogorov, 1968] semblait la plus indiquée pour estimer la complexité des séquences binaires que constituent les musiques du réseau. Définie comme la taille de plus petit programme capable de simuler une séquence, cette complexité peut en réalité être approchée par la taille du fichier de la séquence compressée. Cependant, les séquences auxquelles nous avons à faire avec nos réseaux de quelques nœuds sont de petites tailles : un réseau de 5 nœuds peut espérer une musique de $2^5 = 32$ bits, ce qui est court ; même une compression d'une séquence de moins de quelques centaines de bits ne serait pas bien significative. Comme nous travaillons sur de petits réseaux, souvent de dimension 2, 3 ou 4, nous nous sommes rabattus sur un algorithme nommé CTM-BDM (pour « Coding Theorem Method » et « Block Decomposition Method »)[soler-Toscano, 2014].

L'algorithme CTM consiste à énumérer systématiquement tous les programmes courts (toutes les machines de Turing à 5 états) et tous leurs outputs associés (les séquences que ces programmes peuvent générer). L'algorithme BDM consiste d'abord à segmenter la séquence en blocs de taille $l \leq 12$ (voir figure 8.38) et d'estimer leur complexité à partir du nombre de machines de Turing capable de les générer. On agrège ensuite les complexités de l'ensemble des blocs de séquence pour obtenir une estimation de la complexité de la séquence entière à l'aide de la fonction d'agrégation $BDM(s) = \sum_{b \in B(s)} (CTM(b) + \log_2(n(b)))$ où s désigne la séquence dont on souhaite estimer la complexité, $CTM(b)$ est la complexité du bloc b , $b \in B(s)$ l'ensemble des blocs de s , et $n(b)$ le nombre d'occurrences de b dans la séquence. On obtient ainsi une limite haute de la complexité de Kolmogorov pour une séquence s donnée.

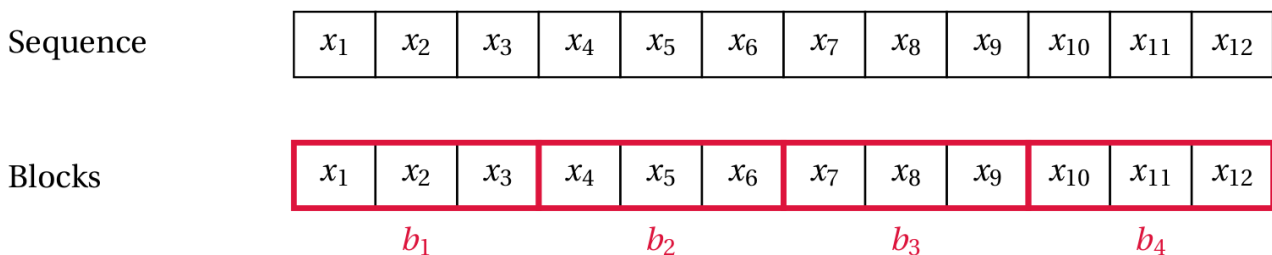


Figure 8.38. Découpage en blocs (ici de taille $l=3$) d'une séquence.

Mesurer la complexité, calculer une distance, c'est déjà modéliser.

Je vais commencer par faire remarquer quelque chose de surprenant : d'après le calcul de complexité fonctionnelle proposé, une fonction de dimension d avec des entrées nulles (voire toutes ses entrées nulles) sera plus complexe qu'une fonction dont toutes les entrées sont de poids non nuls ! En effet, dans un espace de dimension d , elle correspondra à une fonction de dimension inférieure $d-n$, n désignant le nombre d'entrées nulles de la fonction. Du fait de la perte de dimension(s), le cône de sa classe d'équivalence sera ainsi moins volumineux et une telle fonction plus dure à atteindre dans l'espace des paramètres. Dire qu'une fonction avec moins d'entrées, voire

pas d'entrées du tout, est plus complexe qu'une fonction possédant autant d'entrées non nulles que la dimension du réseau est contre-intuitif ! Une fonction comme (0,0) servant par exemple de nœud source dans un réseau de dimension 2, est assez naturellement vue comme « simple » et, *a priori*, on s'attendrait à ce qu'elle ne soit pas complexe. Je propose de prendre quelques minutes pour réfléchir à la notion de complexité, et même de mesure, dans le contexte qui nous intéresse ici, les réseaux d'interaction biologiques.

Définir un indicateur (comme ici la complexité) ou établir une distance (par exemple entre deux réseaux ou bien entre leurs dynamiques), c'est déjà faire un modèle, c'est injecter notre expérience *a priori* d'un système (ce que nous savons de ce système, où plutôt ce que nous pensons savoir, la seule connaissance qui nous est acquise en tout cas, c'est-à-dire la compréhension que nous en avons qui est à la fois partielle et à la fois obligatoirement biaisée par le ou les représentations mentales que nous nous faisons de ce système). C'est aussi faire un choix relatif à ce que l'on veut comprendre de ce système, suivre un impératif visant à répondre à des questions que nous nous posons sur ce système. Lorsqu'on parle des réseaux d'interaction biologiques, ce qu'on modélise ce sont d'une part les comportements propres, internes aux entités mises en interaction (la façon dont elles pensent étant donné ce qu'elles savent s'il s'agit d'entités pensantes, leur activité de transcription étant donné un état de régulation s'il s'agit de gènes, *etc*) et bien sûr la façon dont sont véhiculées et transmises les informations entre ces entités. Par exemple, deux personnes en train de discuter ou 3 gènes en train de contrôler mutuellement leurs activités respectives.

Faisons un peu le point sur ces modèles de systèmes d'interaction biologique ; comme déjà évoqué dans la section 8.3.2.2.e, la façon dont ils sont encodés importe pour la suite de ce paragraphe. La figure 8.39 montre deux façon cohérentes mais bien différentes de concevoir la formalisation d'un réseau d'interaction. Celle qui semble la plus évidente *a priori* et qui est celle que nous utilisons consiste à confondre l'existence des interactions avec les paramètres (poids) non nuls de la matrice d'interaction. C'est ce qui correspond au réseau **C** de la figure ; il n'existe aucun arc spécifié de poids nul. Une autre manière de modéliser ce même système d'interactions biologiques, est de s'intéresser d'une part à l'existence d'interactions, d'autre part à leur efficacité. Dans ce cas, comme le montre le réseau **B** de la figure, il est cohérent de signifier l'existence d'un arc d'auto-interaction inopérant sur la fonction **a** en le représentant avec poids nul, tout en ne représentant pas d'interaction (pas d'arc) entre les fonctions **b** et **a**, aucune interaction de ce type n'ayant été décrite dans le système d'interaction biologique considéré. Dans ce type de modèle que je qualifie de modèle d'adjacence explicite à paramètres séparés, la représentation requiert 2 matrices, une matrice d'adjacence et une matrice contenant les paramètres. Notons que ce modèle est moins homogène. Soulignons aussi que pour éviter le stockage d'une matrice d'adjacence en plus de la matrice des paramètres, on pourrait être tenté de connecter complètement le réseau et d'affecter des poids nuls aux interactions inopérantes comme aux interactions inexistantes, comme dans l'exemple du modèle **A** (dont on supposera l'inexistence d'interaction entre les fonctions **b** et **a**). Une telle façon de voir les choses est incohérente car les arcs ne servent alors plus à rien vu qu'ils ne

permettent plus de distinguer ce qui est inopérant de ce qui est inexistant. Seuls les paramètres servent, donc autant directement transformer un tel réseau en réseau C où seules les interactions de poids non nulles sont représentées.

Cette différence, comme je le disais, revêt une importance non négligeable lorsqu'on parle de complexité. Dans un cas (B), il existe un arc aa , même s'il est de poids nul sur la fonction a , ce qui ajoute une information sur cette fonction et doit affecter inévitablement sa complexité.

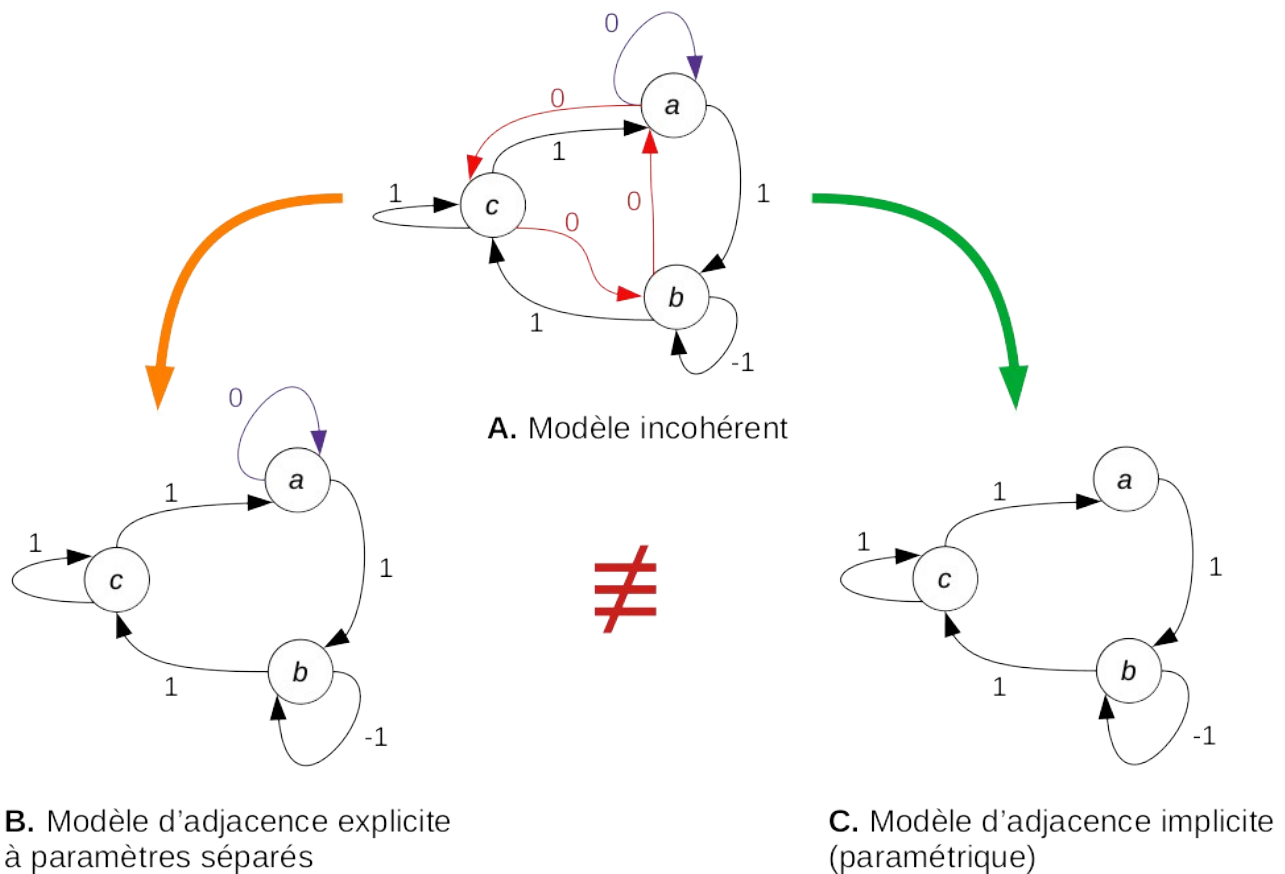


Figure 8.39. Cohérence des modèles. Cette figure montre 3 réseaux. Le réseau C est un SBN du type de ceux que nous utilisons. Son adjacence est complètement connue par le poids des arcs. Un poids nul signifie l'absence d'un arc, l'absence de relation. Une autre façon de modéliser un système d'interactions biologiques est de signifier quelles sont les interactions possibles et d'y ajouter une information supplémentaire, une information d'efficacité de cette interaction sous la forme de poids. Dans ce deuxième cas (B) le réseau possède deux informations séparées. On voit que ce réseau B n'a pas d'arc reliant les fonctions b à a (ni c vers b , ni a vers c); cela signifie qu'aucune interaction n'est décrite entre ces entités. En revanche, la fonction a du réseau B présente un arc d'auto-activation de poids nul. Cela signifie qu'une interaction est décrite mais qu'en l'état elle est inopérante. Le réseau A en revanche est ambigu et potentiellement incohérent : hyperconnecté, il a pourtant quatre arcs de poids nul. Il peut s'agir, comme dans le réseau B, de deux interactions décrites mais inopérantes. Dans ce cas, un tel réseau est valide dans une telle modélisation. Si maintenant certaines de ces interactions (ici les arcs b vers a , c vers b et a vers c montrés en rouge) n'existent pas réellement (ce qu'on ne peut deviner à la lecture du graphe), alors cela signifie que ce modèle utilise la représentation d'adjacence de façon incohérente : il est inutile de représenter des arcs avec un poids nul si c'est pour signifier l'absence d'interaction. Ce modèle doit donc être transformé en modèle B (modèle d'adjacence explicite à paramètres séparés, en spécifiant ce qui est adjacence et ce qui ne l'est pas) ou en modèle C (modèle d'adjacence implicite paramétrique).

Revenons à la complexité. La complexité dynamique du réseau, c'est la complexité de ce qui est produit par tout ou partie de la structure du réseau, un comportement phénotypique, une idée collective, *etc.* La complexité structurale du réseau, c'est celle de la façon dont les arrangements entre les entités sont faits. La complexité fonctionnelle du réseau, elle, peut être vue comme la difficulté d'obtenir une entité donnée dans un réseau, sa rareté, sa difficulté à être produite. Une telle définition de la complexité fonctionnelle est un choix ; c'est bien un modèle que nous proposons là. Voyons s'il est cohérent. Pour y répondre, peut-être suffit-il de s'interroger sur la probabilité de trouver des entités ne recevant aucune interaction afférente (ou très peu) et n'étant connectées au système d'interaction que comme sources, et sur celle de conserver de tels systèmes d'interaction au cours de l'évolution des systèmes biologiques, qu'il s'agisse de gènes, de neurones dans les cerveaux, d'espèces dans les écosystèmes. C'est probablement presque impossible de trouver de tels éléments dans les réseaux biologiques, car la vie, son évolution, créent du branchement, de l'interaction, des relations, forçant les éléments des systèmes à échanger informations ou matière et à se recalibrer par rapport à ces nouvelles entrées. La rareté comme critère de mesure prend alors tout son sens. Mais on peut aussi y voir une notion de coût : il est *a priori* coûteux de maintenir un élément utile (sinon *non maintenu* par l'évolution) mais non ou faiblement connecté en entrée au cours de l'évolution, au même titre qu'il est coûteux de mettre en place une chaîne de production spécifique pour la pièce unique d'un jeu de Lego ou d'un réveil mécanique, pièce qui joue un rôle unique dans l'assemblage auquel elle est destinée.

La complexité, au sens commun, se retrouve dans cette difficulté d'obtention et de maintenance, dans la complexité du système générateur : nous trouvons ces entités avec peu d'entrées, « simples » par référence à ce que nous connaissons, les machines, dont nous savons apprécier la complexité des pièces, par rapport à leur fonctionnement interne faisant possiblement intervenir de nombreux sous-mécanismes, mais aussi parce qu'elles sont dures à fabriquer. La fonction Booléenne à signe est assez peu réductible ; autrement dit, elle ne contient pas de sous-mécanismes. La pièce de machine l'est. Les parties qui la composent ne sont pas nécessairement *complexes* au sens commun. Si elles le restent, c'est uniquement parce qu'elles ne sont pas faciles à fabriquer. La complexité de conception ... On revient là à une forme assez canonique de ce que peut être la « complexité », terme éminemment subjectif !

Je vais aller un peu plus loin. Cette définition de la complexité d'un réseau fondée d'une part sur la centralité pondérée des fonctions entre elles et, d'autre part, sur la rareté de ces fonctions, me semble incontournable. Lorsqu'on souhaite étudier sans *a priori* la relation structure-fonction (ou structure-comportement) d'un système (ex : génotype-phénotype), il faut s'abstenir d'intégrer dans une description structurelle toute connaissance *a priori* que l'on pourrait avoir sur la dynamique (le comportement) du système. Avoir la capacité mentale de représenter la structure d'un système et de simuler son comportement (comme c'est le cas pour toute personne habituée à travailler sur des réseaux Booléens, qui plus est de petites tailles) est un piège ! Ce qui est rendu possible par la simplicité des éléments qui composent le système (les fonctions Booléennes à signe ici) et celle de leurs relations, ne l'est pas pour la grande majorité des systèmes d'interaction biologique (systèmes

génétiqes, séquence et fonction des protéines, écosystèmes, *etc*) ou même systèmes artificiels (ex : un moteur de voiture, un système électronique) ; sauf à avoir une connaissance *a priori* de ce qui, dans la structure d'un système permet telle ou telle dynamique, il n'est pas possible d'inférer un fonctionnement à partir d'une structure donnée, ni l'inverse. Le génotype, séquence de bases nucléiques, est peu compréhensible (il faudrait pour la traduire en séquence d'acides aminés connaître par cœur le code génétique ; la structure secondaire puis tertiaire de la protéine ainsi traduite est encore moins prédictible, même par un ordinateur) et ne permet pas de deviner le phénotype (la ou les fonctions émergeant de la structure primaire jusqu'à la structure tertiaire voire quaternaire de la protéine). Sans connaissance *a priori* de liens entre structure et fonction, il n'est pas possible d'injecter (malgré nous), sauf par erreur, des connaissances concernant la fonction dans un descripteur de la structure. Malheureusement, c'est tout à fait possible dans notre cas lorsque nous décidons de mesures concernant les réseaux Booléens. Je vais être plus précis. Du point de vue dynamique, une fonction de paramètres $(-1,-1)$ est strictement équivalente à une fonction $(-1,0)$ ou $(0,-1)$ ou même $(0,0)$: ce sont des fonctions avec la même table de vérité, donc ayant la même dynamique. Fonder un calcul de complexité des fonctions sur des classes d'équivalence dynamique est une erreur car c'est injecter dans cette complexité des fonctions la connaissance de leur dynamique. Étant donné un formalisme permettant une description du monde physique ou biologique (intégrant notamment la possibilité d'exprimer l'existence de relations non fonctionnelles, mais aussi un potentiel évolutif), si l'on n'a pas la moindre idée de la façon dont se comporte dynamiquement une fonction, alors la seule chose que l'on puisse dire sur la structure c'est que $(-1,-1)$ possède deux entrées exprimant toutes deux une relation négative (inhibitrice), tandis que $(0,0)$ n'en possède pas. Ces fonctions étant irréductibles (pas de sous-mécanismes), la seule façon d'évaluer leur complexité, c'est de procéder comme on le fait pour une pièce mécanique ou une pièce électronique élémentaire, c'est d'en évaluer la difficulté de production, autrement dit la rareté.

Notre mesure de complexité-rareté fonctionnelle a donc ainsi une certaine cohérence. Il est évident que d'autres définitions, adaptées à d'autres problèmes, puissent être considérées comme une complexité fonctionnelle (plus haut niveau) décrivant la machinerie interne de l'entité (ex : un cerveau si l'entité est un humain), ce qui transforme les informations entrantes en état de l'entité. Dans le cadre des SBFs cette machinerie interne se résume à l'équation de la somme des états entrants pondérés, donc aux poids entrants sur une fonction. En particulier pour les fonctions de faible arité⁴¹, dans les réseaux de petite taille, la différence de complexité entre fonctions aurait été trop faible. Notons toutefois que l'arité est bien prise en compte dans le calcul de complexité structurelle – centralité.

Complexité et compaction. De la même manière qu'un réveil matin mécanique est complexe s'il est fait de pièces complexes (rares / coûteuses à concevoir) agencées de manière compliquée et non de pièces simples agencées en file indienne, un réseau est complexe s'il est fait de fonctions complexes subtilement agencées entre elles. Suivant le principe de Somolonoff décrit dans [Zenil, 2018], cette

41 L'arité d'une fonction est le nombre d'arguments ou d'opérandes qu'elle contient. Dans le cas des fonctions Booléennes, ce sont leurs paramètres.

a 6 degrés de liberté et sa physique de vol est gérée) d'un Piper Cherokee évoluant dans un paysage en fil de fer.

Une autre façon de voir le problème est celle du positionnement de l'ingénieur vs la Nature quant à la conception / mise en place de réseaux réalisant une tâche de complexité donnée. L'homme, ingénieur devra beaucoup réfléchir pour concevoir un réseau très complexe, chargé de boucles et aux paramètres précis, surtout si le réseau requis nécessite de nombreux nœuds ; l'ingénieur préférera assurément une architecture modulaire, couplant de petits éléments de réseau dont il connaît le comportement, à la manière dont il conçoit un programme informatique. Pour compacter son code, le programmeur doit avoir une solide expérience en programmation mais cela peut lui assurer d'être le grand gagnant annuel de l'IOCCC⁴² (ou tout autre concours de code offusqué) avec un code savamment offusqué (figure 8.40). Néanmoins, le programmeur préférera, pour des questions de maintenabilité des programmes notamment, un code clair, lisible, modulaire. Cette digression sur la programmation offusquée, qu'on pourrait décliner à la miniaturisation et à l'optimisation des systèmes électroniques, soulève donc la question de où se situe le curseur de la Nature. Là où l'ingénieur met à profit son expérience accumulée au cours du temps pour parvenir à rendre des machines plus compactes, optimalement architecturées, l'évolution de la vie tend-elle aussi à faire évoluer les systèmes du vivant vers des systèmes plus compacts, plus complexes ? Les opérons en sont certainement une manifestation, mais plus encore le regroupement de protéines aux fonctions élémentaires sous la forme de domaines intégrés dans des protéines complexes. Il en est très certainement de même pour les réseaux biologiques, génétiques et métaboliques, mais aussi de neurones ou écosystémiques. C'est une discussion que nous aurons sous peu dans une section dédiée. Voyons déjà s'il y a des enseignements à tirer dans l'étude des relations structure-fonction.

Relation structure-fonction. Dans deux articles nous avons énuméré tous les réseaux (TBNs dans [Christen, 2019] puis SBNs dans [Segretain, 2020 a et b]) d'une dimension d donnée et, pour chaque réseau, toutes les musiques qu'il est capable de générer, et systématiquement calculé la complexité de l'un et de l'autre.

Une série d'histogrammes des complexités des TBNs de dimensions $d=\{2,3,4,5\}$ est donnée figure 8.41. Plus que la répartition des complexités des réseaux, c'est la mise en regard de la complexité des réseaux et des complexité des musiques qu'ils étaient capables de générer qui est intéressante.

Une analyse rapide de la figure 8.42 montre que les musiques complexes ne sont bien entendu jouées que par des réseaux d'une complexité suffisante, mais que l'immense majorité des réseaux, complexes surtout, joue des musiques simples, soit courtes, soit comportant beaucoup de répétitions comme (10000)*.

42 IOCCC : International Obfuscated C Code Contest (<https://www.ioccc.org/>), un concours renommé de programmation dans lequel les virtuoses de la programmation montrent leur talent en terme d'offuscation de code.

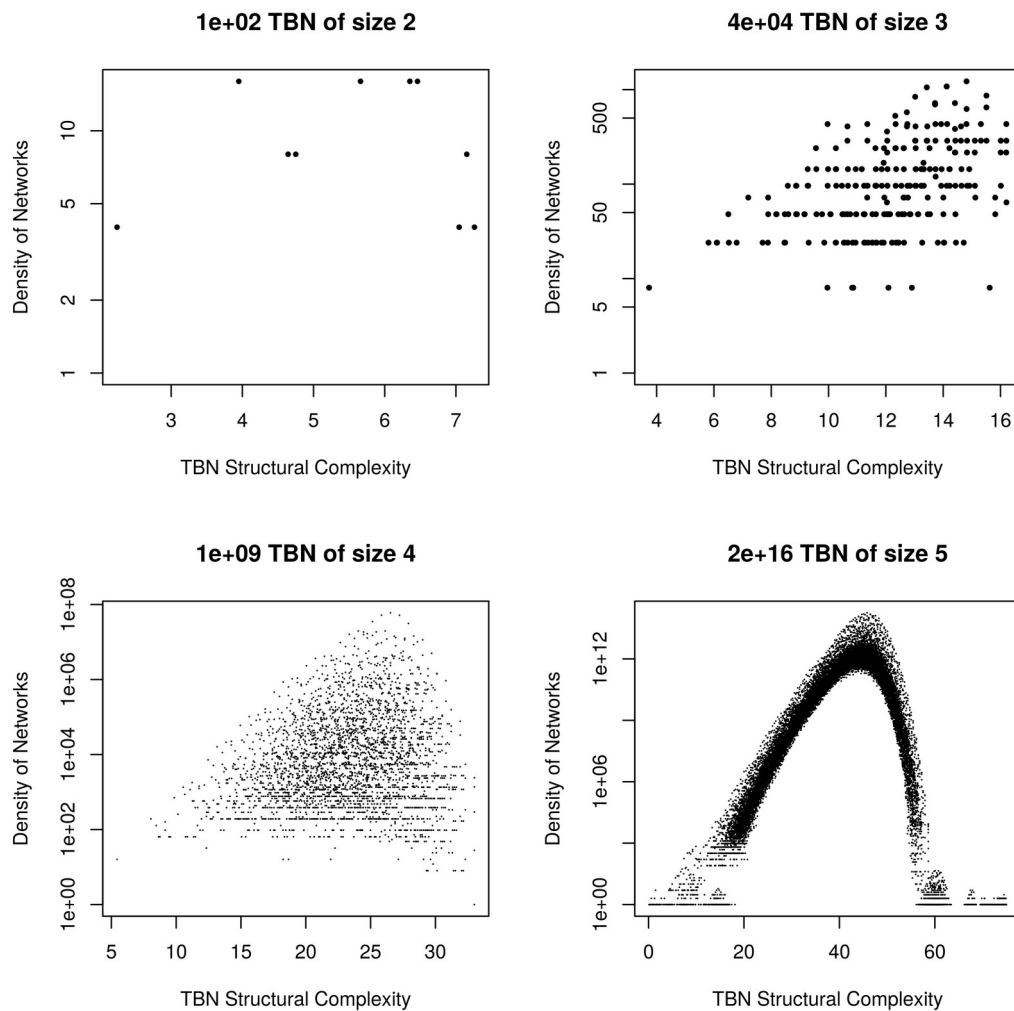


Figure 8.41. Histogrammes des complexités des TBNs de dimension $d=\{2,3,4,5\}$. Le nombre de TBNs d'un niveau de complexité donné est compté (voir [Christen, 2019] pour l'algorithme non trivial de calcul de ces histogrammes). Le pic de densité de TBNs se situe dans le dernier tiers des niveaux de complexité. Il y a donc beaucoup de réseaux d'une complexité comprise entre la complexité moyenne et la complexité maximale. NB : les complexités ne sont pas comprises entre 0 et 1 car le calcul de celles-ci différerait de celui présenté dans ce manuscrit. Voir [Christen, 2019].

Une analyse plus détaillée reste à faire pour déterminer dans quelle mesure la complexité de ces réseaux est allouée à générer plusieurs musiques différentes (appartenant à plusieurs attracteurs), éventuellement à générer des musiques longues mais répétitives plutôt que longues et complexes, ou encore à assurer une robustesse élevée à des musiques simples. Cela reste aussi à confirmer pour des réseaux de plus grande dimension. Enfin et surtout, la question reste ouverte concernant les réseaux du vivant : l'évolution a-t-elle poussé le curseur dans une direction plutôt qu'une autre ? Différentes stratégies pourraient avoir été adoptées par le vivant selon les milieux dans lesquels le vivant se développe, les règnes du vivant ou plus localement les espèces, selon les époques de la vie sur Terre, selon l'importance vitale du système considéré et bien entendu selon le type de système d'interaction biologique.

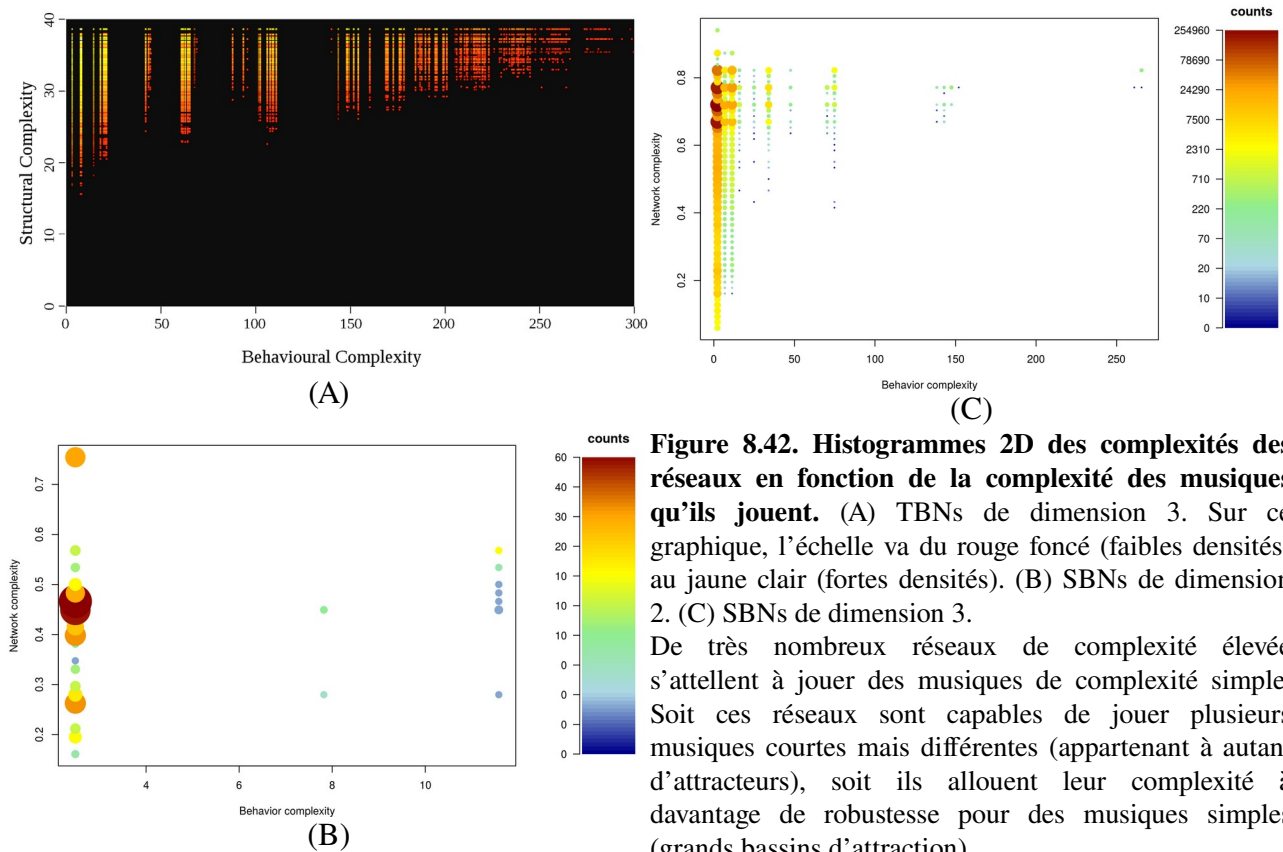


Figure 8.42. Histogrammes 2D des complexités des réseaux en fonction de la complexité des musiques qu'ils jouent. (A) TBNs de dimension 3. Sur ce graphique, l'échelle va du rouge foncé (faibles densités) au jaune clair (fortes densités). (B) SBNs de dimension 2. (C) SBNs de dimension 3.

De très nombreux réseaux de complexité élevée s'attellent à jouer des musiques de complexité simple. Soit ces réseaux sont capables de jouer plusieurs musiques courtes mais différentes (appartenant à autant d'attracteurs), soit ils allouent leur complexité à davantage de robustesse pour des musiques simples (grands bassins d'attraction).

Évidemment, les musiques complexes ne sont jouées que par des réseaux de complexité modérée à élevée.

Nous verrons que l'évolutivité des réseaux est peut-être un guide déterminant ces stratégies. Selon leur complexité et la complexité des comportements qu'ils assurent, certains réseaux et fonctions associées sont possiblement plus évolutifs que d'autres. Nous pensons notamment que ce sont les réseaux de complexité modérée qui sont les plus à même d'évoluer en maintenant viable la fonction qui leur est associée. Il est possible alors que l'impossibilité pour certains systèmes d'évoluer (donc l'impossibilité du vivant à produire des versions d'un tel système adaptables à un environnement changeant) les condamne, restreignant ainsi le vivant à des réseaux appartenant à une certaine classe de complexité (par exemple aux réseaux de complexité modérée). Travailler avec les SBNs permet une certaine exploration du caractère évolutif des réseaux en lien avec la complexité de leur structure et de leur dynamique associée. Nous y reviendrons dans la section 8.3.5.5 et dans la conclusion en 9.1.

8.3.5.2. Distances - Classification des réseaux

Établir une distance dépend du contexte dans lequel on souhaite l'appliquer. Le paragraphe précédent laisse clairement entrevoir mon positionnement quant à l'usage en recherche en biologie théorique que je souhaite faire des réseaux Booléens. Ce qui m'intéresse particulièrement, c'est la façon dont une fonction émerge dans le vivant, depuis ses origines même, puis évolue, une fonction

qui peut peut s'avérer capitale pour le vivant qui dès lors en dispose et ne peut plus s'en séparer, une fonction soutenue par un système fragile composée de plusieurs éléments en interaction, tous sujets aux changements. Mais certains systèmes, ainsi que leurs fonctions associées, peuvent dans la mesure du raisonnable, c'est-à-dire tant que la viabilité de l'organisme est maintenue, être amenés à changer, à muter. Dans ce contexte, encore une fois, les réseaux Booléens que nous utilisons sont un bel outil pour étudier l'évolution.

Il est essentiel de pouvoir comparer ces systèmes, d'en mesurer la proximité ou au contraire l'éloignement, et même éventuellement déterminer des vitesses d'éloignement dans un cadre évolutif ou géographique (ex : éloignement géographique d'espèces vivantes). Établir de telles distances, c'est le lot des théoriciens souhaitant mesurer la proximité des espèces vivantes, ou de certains de leurs caractères (des protéines particulières, des lipides ...) ; ces distances sont établies sur la base de comparaisons de séquences génétiques ou d'acides aminés [Lespinat, 2020]. C'est ce que je souhaite obtenir également, mais avec des systèmes d'interaction biologique modélisés par des réseaux Booléens.

Établir une meilleure distance que celle, simpliste, proposée sur la base de la taille des bassins d'attraction (section 8.3.3.1), requiert de tenir compte de plusieurs aspects concernant la dynamique et la structure des réseaux. Le comportement dynamique d'un réseau reflète complètement sa structure. Une distance entre réseaux devrait donc pouvoir être établie, *a priori*, à partir de la connaissance du graphe de transition. Idéalement, il faudrait qu'une telle distance soit par exemple telle que :

- un état transitoire dont l'image est changée, mais reste dans le même bassin d'attraction, ne coûte pas trop cher
- un état transitoire changeant de bassin d'attraction coûte davantage
- un cycle (un attracteur) légèrement modifié coûte encore davantage
- la disparition d'un attracteur avec un petit bassin d'attraction coûte relativement peu en comparaison de celle d'un attracteur ayant un grand bassin d'attraction

Pour évaluer ceci quantitativement, il faut pour cela définir ce qu'on entend par « même bassin d'attraction » et « cycle (ou attracteur) modifié ».

Il est difficile de tenir compte de tout cela et une partie de notre travail actuel est concentré sur cette tâche. Illustrons cette difficulté avec quelques exemples montrés figure 8.43. Le graphe (a) comporte un seul attracteur, un cycle limite de 4 états, en rouge, les 4 autres états formant le reste de son bassin d'attraction. En (b), ce bassin d'attraction est coupé en deux avec la création d'un nouveau bassin d'attraction, un cycle de taille 3, en bleu. De même entre d'un côté (a) et de l'autre (c) qui voit la création d'un bassin d'attraction (bleu) identique à celui de (b) et dont les trajectoires sont assez semblables mais dont l'attracteur se réduit à un seul état (point fixe). On comprend aisément comment le comportement de (b) a pu évoluer en celui de (c) ; une telle évolution a pu être graduelle. De même entre (a) et (d), on comprend une évolution possible des états transitoires de (a) en 2 attracteurs et leurs trajectoires associées en (d). Ou bien s'agit-il d'une évolution de (c) vers (d). Quel réseau et son graphe de transition associé est-il plus proche d'un autre ? Là, on commence à comprendre qu'établir une distance entre ces réseaux pour les positionner dans un paysage

« morphogénétique » s'avère ardu. Le graphe (e) semble presque une copie de (a) dans laquelle une partie du cycle est modifiée (l'état (000) devient (101)). L'évolution de (a) vers (f) est en revanche moins triviale bien que (a), (e) et (f) conservent le même bassin d'attraction.

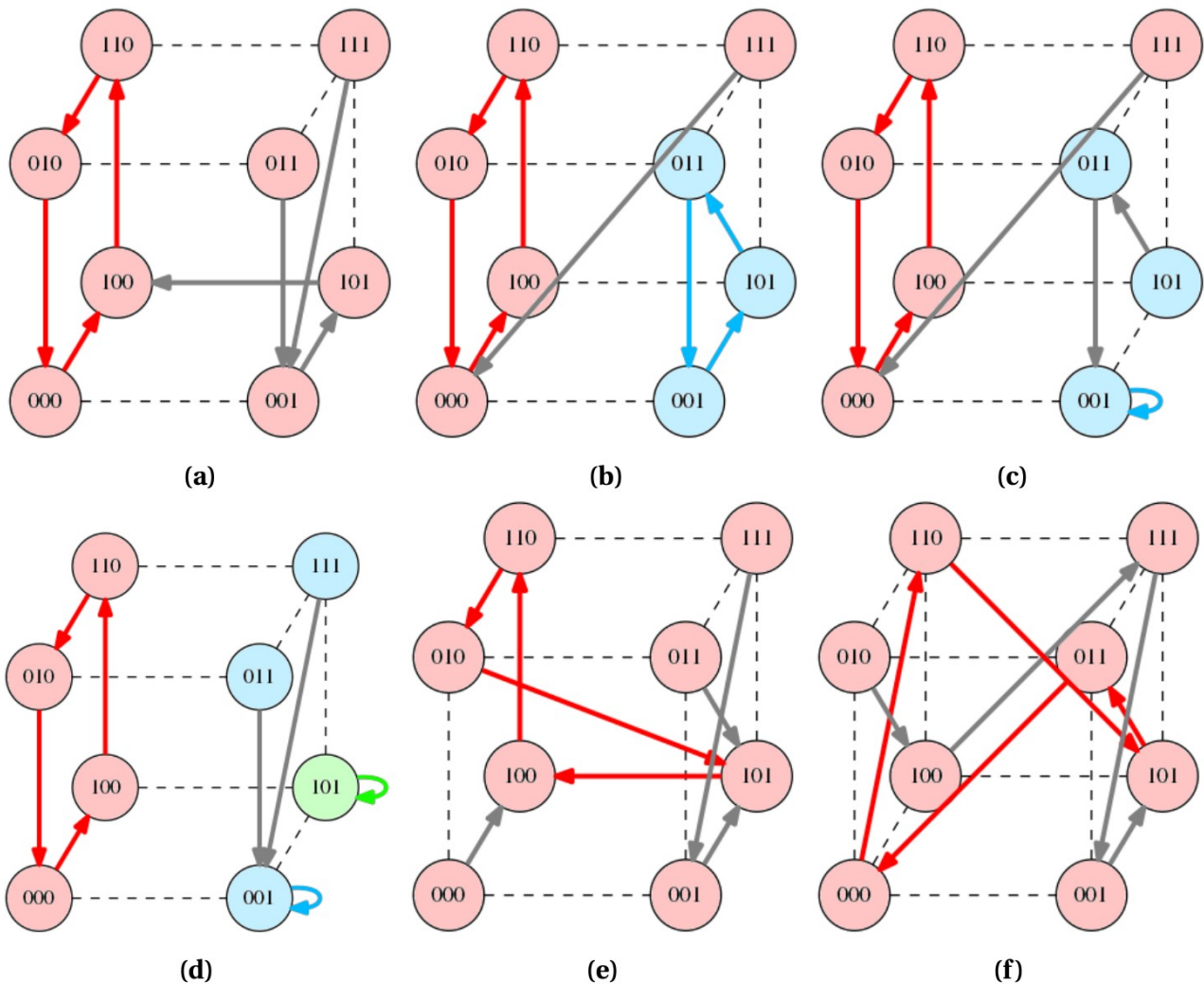


Figure 8.43. Exemples de changements que peuvent subir les graphes de transitions (réseaux).

Ne prendre en compte que la dimension dynamique des réseaux pour les différencier risque de faire passer à côté d'un point important : si la dynamique d'un réseau est l'expression « phénotypique » de sa structure, c'est bien cette dernière qui modélise le système biologique sujet à une éventuelle évolution. Or, qu'est ce que ces évolutions d'une dynamique à une autre impliquent en termes d'évolutions du réseau lui-même ? Sont-elles aussi graduelles ? Il apparaît dès lors essentiel d'établir en parallèle une distance structurelle entre réseaux pour dresser ce paysage morphogénétique peuplé des réseaux et pouvant servir de matrice à l'évolution.

Comme je l'ai déjà dit, la distance entre deux instances d'un système physique doit refléter ce qui fait sens en regard du fonctionnement de ce système physique ; en cela, une distance est un modèle.

La distance structurelle entre deux SBNs, modèles d'un système biologique, devra intégrer des éléments de sa réalité physique. Pour un SBN modélisant un réseau de régulation génétique par

exemple, le système biologique sous-jacent est formé de gènes et de régions promotrices et régulatrices, le tout supporté par une chaîne de bases azotées possédant une structure tridimensionnelle, malléable, de facteurs de transcription, *etc.* Les changements que subissent ces éléments physico-chimiques sont restreint : les mutations des bases azotées sont ponctuelles et la double robustesse du code génétique (voir la partie précédente sur les origines du vivant) empêche des changements trop abrupts, induisant des variations graduelles de la structure accompagnées de variations graduelles de la fonction biologique. Ces questions sur l'aspect graduel de l'évolution restent d'actualité comme le montre cet article récent de Pagel *et al* [Pagel, 2022]. Comment cela se traduit-il dans le modèle de réseaux Booléens ? Même question pour des mutations affectant les régions de régulation (activation ou inhibition changeante), ou même les régions promotrices (augmentation ou diminution de l'expression du gène). Les changements plus brutaux comme les insertions de morceaux de génomes (parfois des gènes entiers) ou des délétions vont, eux, amener possiblement à des extensions de fonctions biologiques : par exemple la composition de gènes ainsi faite peut amener à la formation de nouvelles protéines composites. Leurs domaines sont issus des gènes ou parties de gènes insérés en aval d'une région promotrice. Dans ce cas, on pourra éventuellement considérer qu'une fonction est partiellement conservée au cours du processus évolutif. Dans les graphes de transition des SBNs, cela se traduit par une extension des attracteurs correspondants. Mais dans leur structure, c'est plus compliqué : du point de vue des SBNs, un gène ou une partie de gène inséré dans un gène, pour reprendre un tel exemple, c'est une composition de fonctions Booléennes qui, rappelons le, représentent les gènes, telle que cette composition doit générer une composition de dynamique représentant, elle, le phénotype composite. Nous développerons ce point lors de la présentation du travail que nous avons fait, R. Segretain, S. Ivanov, L. Trilling et moi, sur l'extensibilité des réseaux (section 8.3.5.5). En attendant, exprimer un coût dans le cas des délétions, copies, compositions n'est pas aisé. Le mieux est probablement de s'inspirer de la façon dont sont calculées les distances entre séquences génétiques et les nombreuses pondérations qu'elles comportent sachant une réalité biologique, comme par exemple la prise en compte de biais compositionnels aux niveaux nucléiques et protéiques en travaillant sur une pondération habile des matrices de substitution dans le cas de l'alignement de séquences nucléiques relatives à des espèces dont on connaît les fréquences naturelles nucléiques et protéiques [Bastien, 2006].

Une distance étant un modèle, chargé d'une part de subjectivité, il n'est pas possible d'en définir une idéale. Il est cependant tout à fait recommandable d'en choisir une, d'en énoncer les limites et de travailler avec sur un sujet d'étude, en ce qui nous concerne les SBNs. Ce faisant, on apprend nécessairement des choses qui aident à réfléchir sur ce système et ce qu'il modélise (*cf.* la discussion à ce propos dans le préambule et le chapitre introductif). C'est ce que je propose de faire : définir une distance structurelle et une distance dynamique intégrant au mieux un certain nombre de critères énoncés ici, d'autres éventuellement et, de façon pragmatique, calculer dans une population de réseaux, les distances entre leurs structures et entre leurs dynamiques. Par exemple, une première étude que nous allons réaliser sous peu, consiste à calculer le couple de distances structurelle et

fonctionnelle (dynamique) pour chaque couple réseau-musique (en se limitant à des SBNs de dimension d donnée) dont nous connaissons déjà les complexités respectives (*cf.* figure 8.41). D'une telle matrice des distances, nous pourrions alors déterminer une classification hiérarchique des réseaux, les plus proches en structure et distance formant des groupes.

Dans la section 8.3.6, nous développerons davantage cette idée et nous ajouterons à cette double distance structuro-fonctionnelle, une contrainte imposée par l'environnement (au sens large : environnement cellulaire, biochimique ...) : celle de viabilité.

8.3.5.3. Expressivité, homogénéisation et décomposition en réseaux de bas niveau

Un point que je trouve intéressant de souligner à propos des graphes de transition des réseaux PBN, TBN et SBN concerne l'expansion qu'ils subissent lorsqu'on développe un PBN en TBNs puis SBNs. Exprimer un PBN en TBN ou un TBN en SBN requiert la plupart du temps (sauf fonctions très simples ; voir 8.3.3.2.c et 8.3.3.2.d) un nombre plus important de fonctions Booléennes. La conséquence est une expansion inévitable du graphe de transition. De plus, un réseau exprimé dans un formalisme de haut niveau (ex : PBN) peut être développé dans un formalisme de bas niveau (ex : SBN) de plusieurs manières différentes (il est simulé de plusieurs manières différentes).

Modéliser un réseau biologique par un PBN, ou même un réseau propositionnel multivalué (non Booléen), c'est utiliser un langage de haut niveau, à forte expressivité, bien utile pour exprimer aisément l'agencement perçu du système biologique, mais dont certaines des mécaniques internes sont cachées. Par exemple, exprimer la fonction $\neg(x \oplus y)$ (c'est-à-dire : *not* (x xor y)) requiert une seule fonction PBF et peut se traduire (par exemple, car cette solution n'est pas unique) par un SBN formé d'au minimum 4 SBFs (voir figure 8.30). Là, il y a quatre questions (liées) qui m'intéressent :

- La première est la relation de ceci avec le vivant et ce qu'on peut en tirer. Descendre à une description plus fine des fonctions Booléennes de haut niveau a potentiellement une équivalence biologique : c'est considérer les mécanismes microscopiques sous-jacents, jusque là négligés, au processus biologique. Au plus bas niveau de la biologie, l'échelle biochimique, ce sont des centaines d'interactions précises entre espèces en présences (acides aminés des enzymes, bases nucléiques, ions, solvant ...), de structures tridimensionnelles (structures protéiques, ribonucléiques, dépliement de la chaîne d'ADN, de celle des ARNs en présence ...) finement agencées qui agissent et constituent autant de mécanismes microscopiques. Revenons à l'échelle plus raisonnable peut-être des processus biologiques successifs impliqués dans la vie d'un réseau de régulation génétique. Ces processus, la transcription et ses étapes détaillées de la promotion d'un gène influencée par des facteurs de transcription à l'exportation de l'ARN messenger vers son site de traduction (le réticulum endoplasmique dans le cas des cellules eucaryotes ; dans le cas des procaryotes, la traduction et la transcription sont contiguës), la traduction et ses étapes répétées faisant intervenir ribosomes, ARN de transfert, *etc.*, et l'ensemble des étapes qui font que le produit de ce gène devienne interaction directe ou plus communément indirecte avec un autre gène, sont autant de processus qui ne sont normalement pas modélisés dans un réseau Booléen qu'on utilise

habituellement pour apporter une lumière sur un phénotype et ce qui l'influence.

Tous ces processus et mécanismes fins, comptent-ils tout de même dans la modélisation ? Cela dépend évidemment de ce qui est en jeu : Modéliser des interactions génétiques ? Modéliser ce même réseau dans un contexte cellulaire ou environnemental fluctuant ? Modéliser l'évolution possible de ce réseau ? Étant donné les considérations que je défends depuis le début de mon engagement en recherche et dans ce manuscrit, je ne peux ignorer cette question de la prise en compte des différentes échelles lors de la projection de l'objet biologique dans un formalisme particulier. C'est, en ce qui me concerne, une question qui revêt plus d'intérêt que la compréhension de tel ou tel système biologique.

- Quelle diversité de réseaux exprimés dans un formalisme de bas niveau obtient-on selon le réseau d'origine exprimé, lui, dans un formalisme de haut niveau et les comportements dynamiques (musiques, attracteurs, bassins) qu'il réalise ? Cette diversité est-elle fiable partiellement à la complexité des comportements attendus, à leur multiplicité, *etc* ? (voir figure 8.44 et figure 8.45)
- Chacun des mécanismes fins évoqués ci-dessus est sujet à la fluctuation, à l'erreur, et à l'évolution. Ne pas prendre en compte ces échelles, c'est laisser de côté une possibilité d'exprimer ces variations au cours d'une dynamique ou au fil des âges. Dans la diversité des possibles déclinaisons microscopiques d'un même réseau de haut niveau, on peut parier que certaines posséderont une robustesse plus élevée à ces changements ponctuels ou permanents. Et pour rejoindre ce qui a été développé dans la section 8.3.5.2 précédente, on peut se demander si certains d'entre eux verront leur mutations graduelles davantage accompagnées d'une gradation tout autant graduelle de leurs dynamiques associées. Explorer la robustesse et l'évolutivité de ces déclinaisons est à mon sens digne d'intérêt.
- Enfin, avec l'expansion de la structure d'un réseau vers plus de détails se fait le dépliement du graphe de transition correspondant (voir figure 8.44). L'attracteur augmente: on reste parfois pendant plusieurs étapes (nœuds du graphe de transition) sur un état des gènes d'intérêt suivis dans le PBN, le temps de réaliser les mécanismes internes pour passer à l'étape suivante. De plus, l'extension du graphe de transition peut amener une plus grande richesse de comportements comme montré figure 8.44. Dans la diversité des solutions, quelles solutions sont les plus robustes ? Quand on déplie une fonction « complexe », comment le graphe de transition se déplie-t-il localement ? Qu'est ce qui, dans la structure détaillée du réseau fait apparaître de nouveaux attracteurs ? Quand on détaille les mécanismes d'un réseau d'interaction, quel pourrait être le sens biologique de la révélation de nouveaux comportements tels que le sont les nouveaux attracteurs apparus en développant un PBN en SBN ? Ces questions là aussi seront une source évidente de réflexions pour moi et pour les étudiants qui voudront bien tourner leur regard vers des sujets fondamentaux de la modélisation, de la biologie théorique et du vivant.

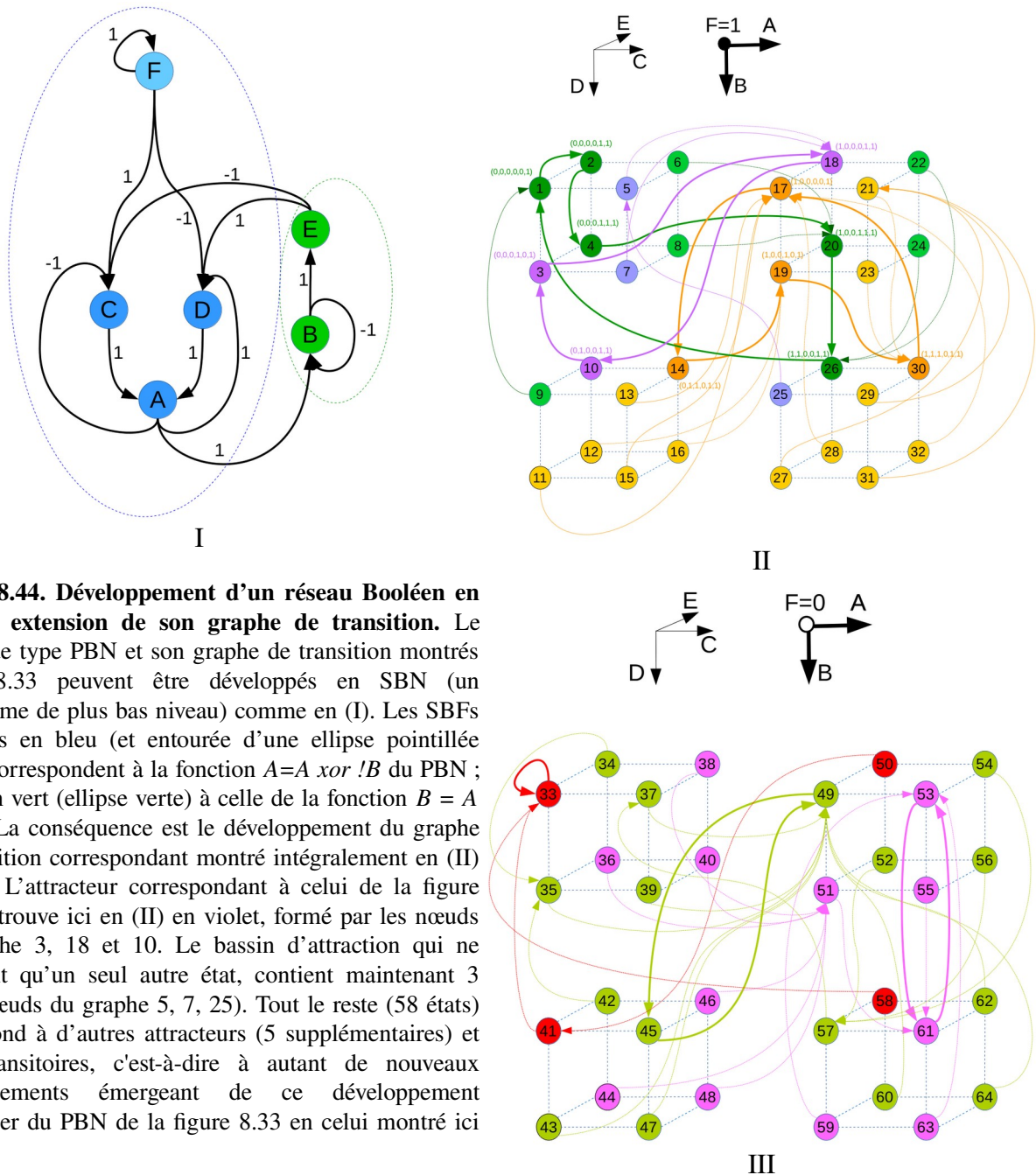


Figure 8.44. Développement d'un réseau Booléen en SBN et extension de son graphe de transition. Le réseau de type PBN et son graphe de transition montrés figure 8.33 peuvent être développés en SBN (un formalisme de plus bas niveau) comme en (I). Les SBFs montrées en bleu (et entourée d'une ellipse pointillée bleue) correspondent à la fonction $A=A \text{ xor } !B$ du PBN ; celles en vert (ellipse verte) à celle de la fonction $B = A \text{ ET } !B$. La conséquence est le développement du graphe de transition correspondant montré intégralement en (II) et (III). L'attracteur correspondant à celui de la figure 8.33 se trouve ici en (II) en violet, formé par les nœuds du graphe 3, 18 et 10. Le bassin d'attraction qui ne contenait qu'un seul autre état, contient maintenant 3 états (nœuds du graphe 5, 7, 25). Tout le reste (58 états) correspond à d'autres attracteurs (5 supplémentaires) et états transitoires, c'est-à-dire à autant de nouveaux comportements émergeant de ce développement particulier du PBN de la figure 8.33 en celui montré ici en (I).

Ajoutons sur la table le plan de mise à jour. Jusque là nous ne parlons que de mises à jour synchrones.

La description d'un réseau pourrait être totalement auto-contenue dans sa matrice d'interactions s'il ne fallait pas rajouter le plan de mise à jour. Ceci constitue une information supplémentaire qui rend le langage non homogène. Cependant, il existe systématiquement une possible expression synchrone de tout réseau asynchrone. Cela se fait par introduction de nouvelles fonctions Booléennes qui introduisent autant de délais (voir figure 8.45). De la même manière que pour passer des TBNs aux SBNs, il fallait rajouter des SBFs pour encoder l'information contenue dans le seuil, l'information

donnée par le plan de mise à jour requiert l'ajout de fonctions Booléennes. Et de la même manière, intégrer cette information au sein même de la structure du réseau est une source de diversité : il n'existe pas une manière unique d'introduire ces délais.

Remarquons tout de suite que ceci aussi aura pour effet de provoquer une extension du graphe de transition, faisant possiblement apparaître de nouveaux comportements et, tout du moins, une nouvelle diversité de trajectoires dans les bassins d'attraction existants.

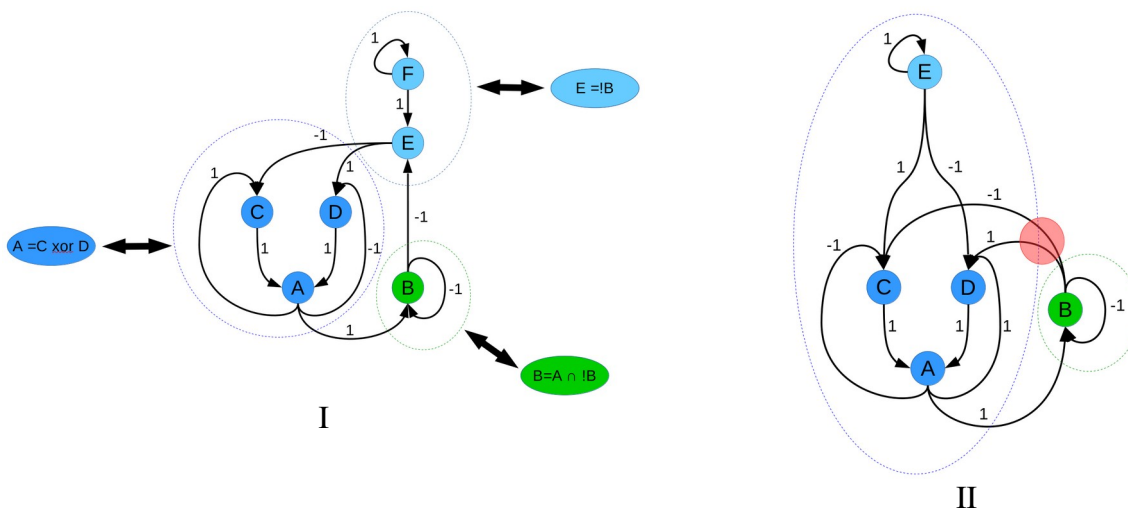


Figure 8.45. Autres développements en SBN. Le SBN montré en (I) est un autre développement valide du PBN de la figure 8.33. En revanche, celui montré en (II), qui ressemble beaucoup à celui montré dans la figure 8.44.I avec une simplification concernant l'action du nœud B sur les nœuds C et D, ne l'est pas : il ne réalise pas les mêmes fonctions. Vus indépendamment, les blocs bleu et vert en (II) réalisent pourtant toujours les fonctions en bleu et vert du PBN, c'est-à-dire $A = A \text{ xor } !B$ et $B = A \text{ ET } !B$. La raison pour laquelle le SBN en (II) et le PBN d'origine ne correspondent pas est que le nœud E de la figure 8.44.I introduit un délai nécessaire pour que le mécanisme représenté par les fonctions entourées en bleu soit dans le bon timing (le délai manquant est marqué en (II) par une région rosée. En (II), le bloc formé par A, C, D et E est formé de 2 couches (la couche (C,D) et la couche (A)) ce qui introduit un délai de traitement de ce qui provient de (B) de 1 itération. La fonction B en revanche a toujours un temps de traitement de 1 itération donc, sans délai, elle se déphase par rapport au bloc (A,C,D,E). Cela montre cependant que l'introduction de délais dans un réseau asynchrone peut évidemment rétablir la synchronie. L'inverse est vrai aussi.

Du point de vue du système naturel ainsi modélisé, l'asynchronie provient des délais de transmission de l'information d'un événement ayant lieu quelque part dans la cellule, à un autre endroit. Les voies de transmission peuvent être diverses et multiples (en même temps) et faire intervenir de nombreux protagonistes. Encore une fois, le modélisateur d'un système biologique a tout intérêt à disposer d'une description haut niveau séparant bien en différentes modalités (structure haut niveau, mise à jour) ce qu'il souhaite exprimer. En revanche, aller dans le sens de l'homogénéisation du langage en intégrant toutes ces modalités et en réduisant la description d'un système complet à une matrice d'interactions (la mise à jour étant implicitement parallèle) permet d'établir plus aisément des mesures sur ces systèmes (complexités, distances). Je suis aussi intéressé par ce qu'apporte comme réflexions sur la modélisation (voir plus haut) ce processus de décomposition/homogénéisation non unique, en particulier concernant les aspects relatifs au potentiel évolutif de ces systèmes (voir section 8.3.6). La recherche systématique des réseaux de bas niveau à partir de l'expression de

réseaux de haut niveau modélisant un système biologique pourrait permettre d'en identifier les différents processus sous-jacents possibles, processus ayant des potentiels évolutifs différents.

8.3.5.4. Compaction et décompaction / composition et décomposition de fonctions Booléennes

En 2017, L. Trilling, S. Ivanov, Q.T. Vuong (un excellent étudiant en M1 informatique), nous sommes intéressés à la composition/décomposition de fonctions Booléennes au sein d'un même réseau (des TBFs à l'époque, mais le raisonnement est le même, bien que plus restreint, pour les SBFs), ce que nous désignons par compaction/décompaction (ou repliement/dépliage), des termes que nous trouvons plus parlants pour ce problème, ou entre réseaux. C'est une thématique difficile dans laquelle je souhaite me replonger.

Considérons par exemple les réseaux montrés figure 8.46.

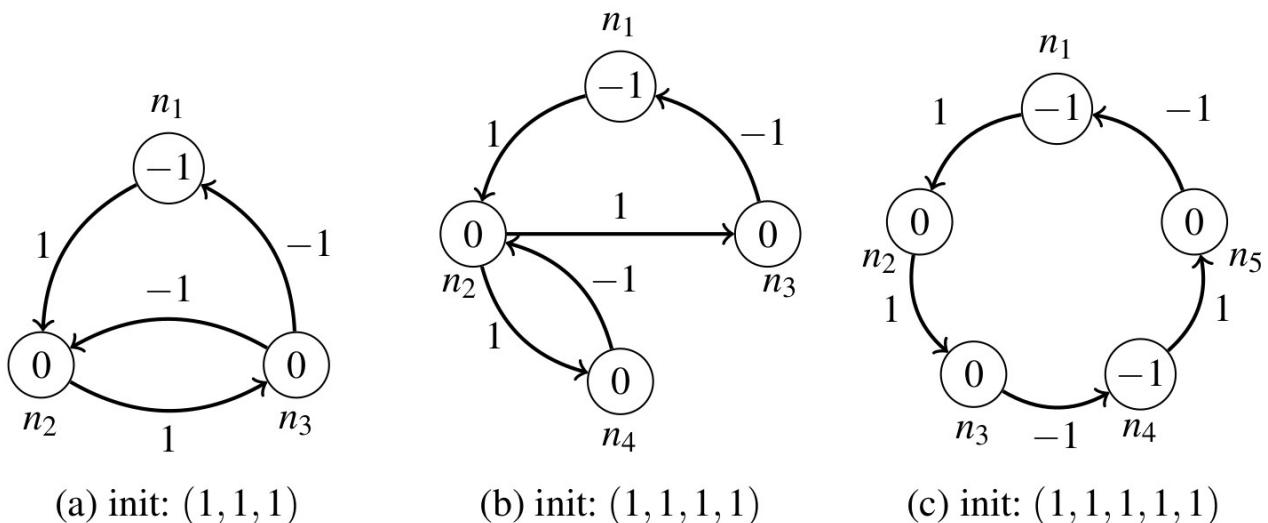


Figure 8.46. Dépliage de TBFs. Ces 3 TBNs de dimension croissante jouent tous la même musique (10011)* sur le nœud n_1 ; c'est le maintien de cette dynamique partielle (une musique) et le maintien partiel de la structure du réseau qui constituent les contraintes à respecter lors des opérations de compaction/décompaction. De telles décompactions ou compactations ne sont pas nécessairement uniques. Les réseaux montrés de (a) à (c) montrent une proposition de décompaction du réseau au niveau de certaines fonctions Booléennes. La fonction n_3 du TBN (a) est décompactée en fonctions n_3 et n_4 dans le réseau (b). De même la fonction n_2 du réseau (b) qui devient la fonction n_4 du réseau (c). Dans ces opérations certains arcs sont créés, certains paramètres changés (seuils). Vu dans l'autre sens, les fonctions absorbent les caractéristiques de ce qui est compacté : elles se complexifient.

Les questionnements à ce propos viennent d'une part de la biologie, d'autre part d'une réflexion sur l'offuscation des programmes informatiques. Au cours de l'évolution du vivant, des réseaux d'interaction assurant chacun un comportement, se combinent parfois structurellement pour former un comportement hybride. Comme nous l'avons vu section 8.3.5.1, certains programmeurs sont particulièrement doués dans la factorisation de fonctions informatiques rendant, pour des besoins d'efficacité, d'occupation en mémoire, ou parfois pour la simple démonstration de leur maîtrise, leur code plus compact, un code qu'à un certain point (à la fois subjectif et objectif) on qualifiera

d'« offusqué ». Dans un programme informatique offusqué, une même fonction peut être réutilisée de plusieurs manières différentes ; elle devient un couteau suisse tenant un rôle central dans un tel programme. Ce code offusqué est très compact mais peu lisible et très peu maintenable : faire évoluer la fonction-couteau suisse précitée relève de l'impossible.

Qu'en est-il des réseaux d'interaction du vivant ? Subissent-ils au cours de l'évolution des opérations de compaction/décompaction ? Comment ? A ce comment biologique s'ajoute le comment numérique : existe-t-il un algorithme de décompaction ? En existe-t-il un de compaction ? Si oui, quel est le critère d'arrêt de ce dernier ? c'est-à-dire comment, à partir de la seule structure d'un réseau peut-on décider qu'une compaction (maintenant une dynamique donnée, dynamique qui correspond à une fonction biologique devant être maintenue) est possible et quand ne peut-on plus compacter ? De façon générale, quel est le niveau d'offuscation du vivant ? Un réseau d'interaction très compact, surtout s'il est lié à une fonction hautement vitale de l'organisme dans lequel il agit, sera *a priori* très peu évolutif. On se doute bien qu'il existe un lien fort avec la complexité des fonctions et de leur arrangement d'une part, de la dynamique requise d'autre part. C'est d'ailleurs pour cette raison que suite à ces travaux exploratoires de 2017, nous avons cherché à mesurer les complexités structurelles et dynamiques. Dès lors, nous avons mis de côté cette thématique de compaction/décompaction.

L'intérêt de cette recherche serait donc de se doter d'un outil qui permettrait d'explorer de manière algorithmique et de comprendre certaines opérations (compaction/décompaction ou composition/décomposition) pouvant avoir lieu dans le vivant (figure 8.47). Si de telles opérations de compaction (ou inversement de décompaction) se produisent dans le vivant, c'est-à-dire, si des gènes appartenant à un réseau génétique se combinent en gènes « complexes », comment le réseau auquel la partie recombinée appartient peut-il maintenir un fonctionnement compatible avec la vie de l'organisme ? Est-ce la sélection naturelle qui élimine systématiquement les recombinaisons échouant à maintenir la viabilité ou bien peut-on avancer que le vivant, au fil de son évolution, a peu à peu retenu des gènes-fonctions plus combinables ou des processus de recombinaison favorisant le maintien de la viabilité ?

L'autre intérêt que je vois à la recherche de règles algorithmiques de compaction est un certain nombre d'interrogations qui surgissent lorsqu'on se rappelle que les SBNs sont une forme de langage de programmation ; ils sont l'abstraction d'une machine. Si l'on impose le maintien d'une partie de la fonctionnalité d'une fonction d'une telle machine-programme (fonction ici représentée par une musique) est-il possible de déterminer si la fonction est compactable avec moins d'instructions/éléments ? De même, comment combiner deux machines de façon à ce qu'elles ne soient qu'une produisant un comportement particulier ? Si oui comment ? Quelles instructions/pièces peuvent être ainsi factorisées ? Comment le programmeur parvient-il en pratique à offusquer son code ? Quel processus cérébral le lui permet ? Dit autrement : si un programmeur parvient avec l'habitude à offusquer du code, à factoriser des commandes, cela signifie-t-il qu'il a su tirer de son expérience des règles (dont il n'est pas nécessairement conscient) ? Cela sous-entend-t-il

finalement une possible existence d'algorithmes pour ces tâches ?

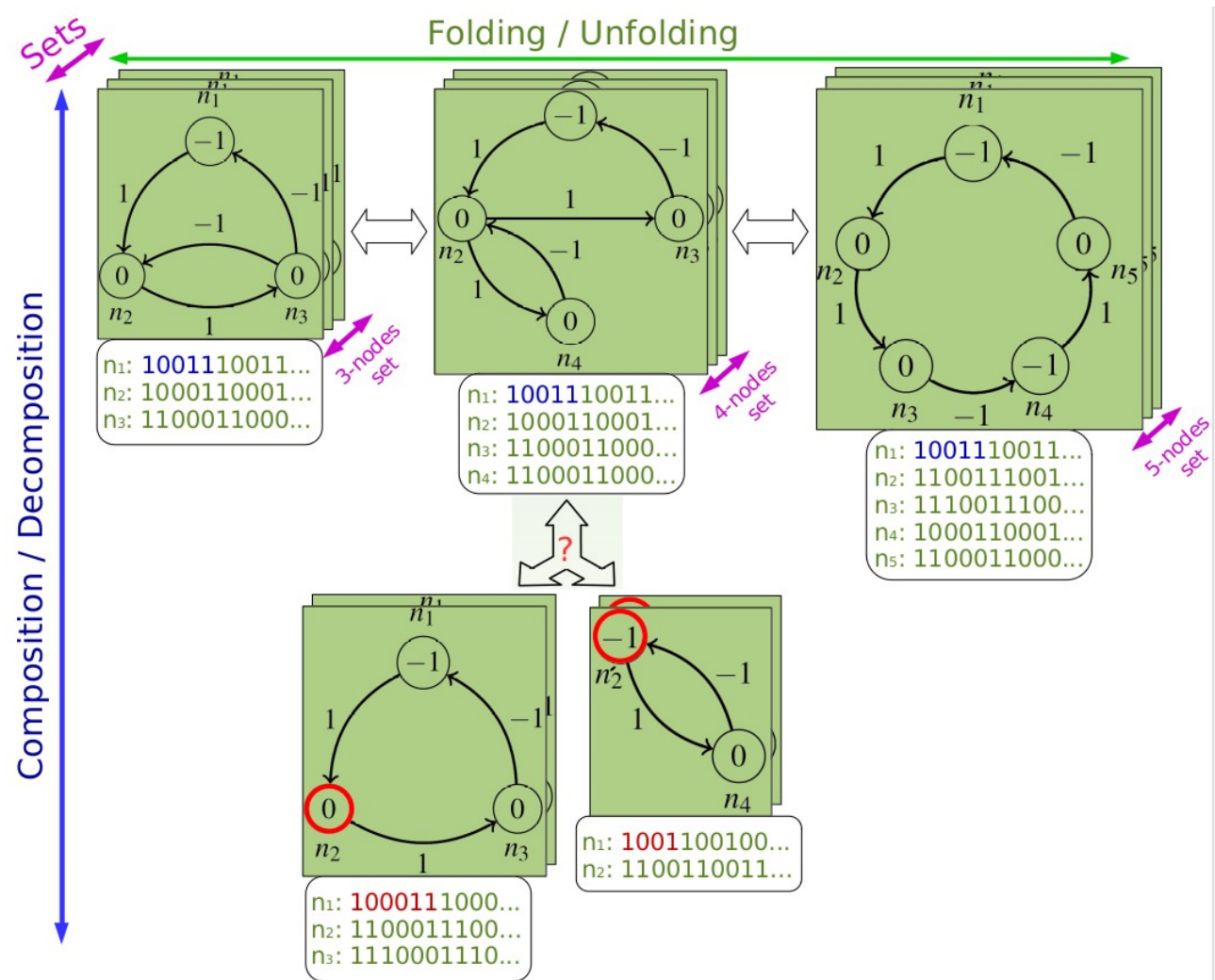


Figure 8.47. Repliement/Dépliement ou Compaction/Décompaction de TBFs.

Avant de continuer, cassons tout de suite l'ambiance : il n'existe certainement pas d'algorithme proprement dit pour compacter des fonctions Booléennes car cela signifierait vraisemblablement résoudre le problème de l'arrêt⁴³. Trouver un programme qui détermine de façon générale quand un autre programme, quel qu'il soit (tous les réseaux possibles), peut s'arrêter (arrêter la compaction, celle-ci n'étant plus possible au bout d'un moment car le comportement sélectionné, une musique donnée, n'est plus maintenu), est indécidable. En revanche, à la vue des programmeurs capables d'offusquer leur code ou des ingénieurs capables de concevoir des machines factorisant plusieurs éléments fonctionnels, il est possible d'imaginer qu'il existe des règles empiriques pour cela. Plus simplement, il sera peut-être possible d'identifier des règles sous la forme d'une liste de primitives compactables ou non. Une comportement du réseau (une musique jouée par le réseau) résulte en effet des caractéristiques des fonctions Booléennes du réseau (activité locale), de leurs arrangements respectifs (interactions) et des flux de bits qu'elles reçoivent respectivement, autrement dit leur synchronie qui est due à la distance (taille du chemin) des fonctions Booléennes entre elles. Un

43 Voir https://fr.wikipedia.org/wiki/Probl%C3%A8me_de_l%27arr%C3%AAt

programmeur, lorsqu'il compacte son code, tient compte de ces 3 points : il sait comment compacter des instructions (activité locale), comment ces instructions et ce qui les remplace sont liés au reste du programme (interactions) et respecte l'ordre de mise à jour des états du processeur effectués par ces instructions en tenant compte des retards ou avances que risque de créer un changement de programme (il corrige alors ces défauts de synchronie soit en prenant en compte la correction dans la nouvelle instruction, soit en restructurant le programme). Si un de ces points ne peut être respecté, un code réalisant une fonction particulière ne peut être compacté. J'avancerais que ceci est peut-être rendu possible, pour le programmeur, parce que les langages de programmation s'expriment dans un langage qui est (volontairement) proche du langage naturel, langage dans lequel nous pouvons mentalement visualiser des processus complexes et éventuellement les condenser.

Remarquons aussi que je parle d'absence d'algorithme pour la compaction, mais pas pour la décompaction. Dans ce dernier cas, je suis plus circonspect et, intuitivement, je pense qu'il est envisageable de toujours trouver une solution de décompaction, l'arrêt étant contrôlé par la complexité des fonctions qui ne peuvent plus descendre en deçà d'un niveau élémentaire.

En pratique maintenant, comme on le comprend il n'est pas possible *a priori* d'extraire de telles règles de compaction en se fondant uniquement sur la connaissance de la structure du réseau. Il est nécessaire de connaître en même temps la dynamique. La table de vérité du réseau contient à la fois sa connaissance structurelle et dynamique ; c'est par elle qu'il faut vraisemblablement passer si l'on cherche de telles primitives de compaction. Pour l'instant, nous ne savons pas s'il est possible d'extraire des tables de vérité des règles de compaction. Une piste possible, bien que peu claire pour l'instant, est d'exploiter une idée de L. Trilling, la décomposition des fonctions Booléennes en cascade, faisant passer une fonction Booléenne de dimension d à d fonctions Booléennes de dimension $d-1$, pour voir si l'on trouve dans ces tables de vérité décomposées en cascade des critères permettant de décider si deux fonctions Booléennes peuvent être compactées et si oui comment. Ce qui est très envisageable en tout cas, c'est qu'il y ait un lien très fort entre d'une part la complexité du réseau à compacter, d'autre part la complexité de la musique à maintenir après compaction. On sait d'emblée que d'un réseau de dimension d dont la musique à maintenir est de taille supérieure ou égale à $2^{d-1} - 1$, ne pourra être compacté en un réseau de dimension $d-1$ car un tel réseau ne peut accueillir un attracteur de taille supérieure à cela (l'état nul du graphe de transition est un point fixe et utilise un état du graphe de transition). En addition, il faudra aussi vraisemblablement beaucoup d'intéresser à la réécriture des graphes, un sujet connexe⁴⁴. Un autre excellent étudiant de M1, ENS Paris, Vincent Ferrari, a dans ce cadre élaboré une preuve d'équivalence (bi-simulation) entre SBNS et Boolean P-Systems (un système de réécriture adapté pour les réseaux Booléens), ainsi qu'un mécanisme expliquant comment passer de l'un à l'autre. Ce travail est en cours de rédaction pour publication.

Une autre manière de procéder pour déterminer des réseaux plus compacts jouant une même musique et ayant des caractéristiques proches, si aucune règle ne peut être (aisément) déterminée, est

⁴⁴ Voir [https://fr.wikipedia.org/wiki/R%C3%A9criture_\(informatique\)](https://fr.wikipedia.org/wiki/R%C3%A9criture_(informatique)) et https://en.wikipedia.org/wiki/Graph_rewriting

de procéder comme nous l'avons fait pour d'autres problèmes : en utilisant la programmation logique par contraintes. En spécifiant une musique devant être lue par au moins un nœud et des caractéristiques comprenant au moins la dimension du réseau, éventuellement le maintien de certains paramètres, on pourra trouver des ensembles satisfiables de réseaux plus compacts. L'observation systématique de ces réseaux par rapport à d'autres dont ils pourraient être issus par compaction de fonctions, donnera des pistes sur d'éventuelles règles de compaction.

8.3.5.5. Extensibilité des réseaux

Nous avons vu plusieurs manières de faire évoluer un réseau d'interaction: par mutations ponctuelles, par factorisation (décompaction, ou au contraire décompaction) de sa structure, mais une manière commune par laquelle les réseaux biologiques se sont complexifiés (ainsi que les fonctions qu'ils supportent) est l'ajout de nouveaux éléments comme des gènes (par duplication puis mutations par exemple, ou par copies de parties entières d'un génome), des espèces animales (arrivée d'une nouvelle espèce dans un écosystème), ou l'apparition au cours de l'évolution de nouvelles structures anatomiques cérébrales ou plus simplement de nouvelles connexions de neurones faisant communiquer des parties initialement disjointes. Que provoque dans un système biologique l'ajout d'un nouvel élément ? Comme dans le cas des mutations, il paraît essentiel que l'ajout d'un élément n'affecte pas (trop) le fonctionnement général du système biologique, c'est-à-dire la viabilité de l'organisme vis-à-vis de cette fonction biologique.

Une première façon de concevoir l'évolution des organismes suite à l'extension de leurs réseaux d'interaction par ajout d'éléments est purement fondée sur la sélection naturelle : si la fonction qui émerge d'une extension d'un réseau n'est pas viable, ce réseau étendu n'est simplement pas retenu par l'évolution.

Une autre est de faire l'hypothèse que certains réseaux ont des propriétés évolutives particulières évitant au vivant de rebattre totalement ses cartes. Et j'ajouterais une hypothèse encore plus forte : l'évolution a sélectionné des réseaux qui disposent de ces propriétés évolutives ; elle maintient le vivant évolutif. Nous avons déjà évoqué cette possibilité en parlant du potentiel de certains réseaux à évoluer de façon graduelle tant dans leur structure que dans leur dynamique.

R. Segretain, S. Ivanov, L. Trilling et moi-même nous intéressons également à l'évolution par extension du réseau avec extension de la fonction biologique ; nous avons expérimenté cette question avec des SBNs et la question que nous posons est la suivante :

Considérons :

- un réseau R_1 de dimension d_1
- une extension de réseau R_2 de dimension d_2 à rajouter au réseau R_1 (pour des raisons de simplicité, nous avons réduit R_2 à sa plus simple expression, c'est-à-dire une seule fonction Booléenne, donc $d_2=1$)
- une musique M_1 jouée par ce réseau R_1 , avec la longueur de la musique $L(M_1) \leq 2^{d_1}$
- une musique étendue M telle que $M=M_1.M'$, avec M' une extension de M_1 avec

$$L(M) \leq 2^{d_1+d_2}$$

Quels sont tous les réseaux R tels que $R = R_1 \circ R_2$ (relation de composition), R de dimension $d=d_1+d_2$, tels que les contraintes précitées soient respectées et tels que R_1 reste inchangé ? Par inchangé, on entend par là que si l'on retire R_2 de R (extrusion) ainsi que tous les arcs qui le reliait à R_1 , on obtient un réseau R_1' équivalent à R_1 , c'est-à-dire possédant le même graphe de transition. Autrement dit, ses SBFs, même si elles n'ont pas les mêmes valeurs de paramètres (poids des arcs entrants), restent équivalentes à celles de R_1 et sont reliées de la même manière entre elles. Nous avons ainsi déterminé toutes les extensions possibles de tous les réseaux R_1 de dimension $d_1=2$ pour toutes les musiques M_1 et toutes les extensions M' possibles.

Et à cette question, nous adjoignons celle de leurs caractéristiques : complexité structurelle du réseau R_1 initial et du réseau R étendu, complexité des musiques initiale M_1 et étendues M , nombre de bassins d'attraction des réseaux initiaux et étendus, taille moyenne des bassins d'attraction, ...

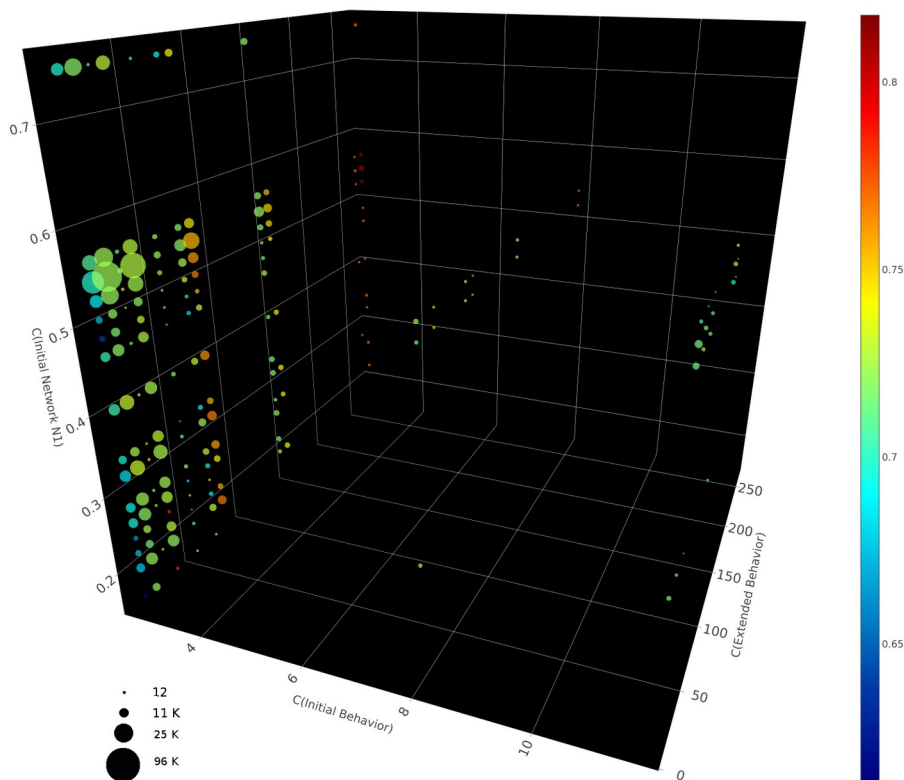


Figure 8.48. Histogramme 3D des complexités structurelle et dynamique des extensions de SBNs. Dans cet histogramme, les complexités du réseau initial R_1 (NB : la figure est en version anglaise. Ici $R_1=N_1$ avec N pour *Network*), des musiques initiales M_1 et des musiques étendues M ont été discrétisées en une cinquantaine de classe chacune. La taille des points (échelle logarithmique) représente le comptage des 777216 quadruplets $\{R_1, M_1, R, M\}$ obtenus par extension, réparti dans les classes 3D de complexité $\{C(R_1), C(M_1), C(M)\}$. La couleur des points représente la complexité moyenne des réseaux étendus $N R$ dans ces classes.

Nous avons publié la partie technique de ce travail et un résultat préliminaire concernant l'extension des réseaux de dimension 2 vers 3, dans la conférence *Complex Networks 2020* et des suppléments

dans *bioRxiv* [Segretain, 2020a et b]. Dans cet article, nous avançons une hypothèse : les réseaux de complexité moyenne et leurs musiques associées seront les plus à même d'évoluer comme décrit ici, plutôt que des réseaux très complexes dont le paramétrage fin permet la réalisation de dynamiques complexes, ou que des réseaux trop simples qu'il est difficile de complexifier pour obtenir un comportement d'une complexité suffisante. L'analyse de tels résultats est compliquée car d'une part il y a beaucoup de dimensions à explorer (les différentes caractéristiques que nous étudions: complexités, bassins d'attraction, etc.), d'autre part les objets manipulés, de petits réseaux et de petites séquences binaires, de part leur petite taille, ne diffèrent *de facto* que peu entre eux ; leurs caractéristiques ne diffèrent donc que très peu. Je ne vais pas faire ici une description exhaustive des résultats que l'on obtient ainsi mais plus souligner le potentiel de telles études systématiques.

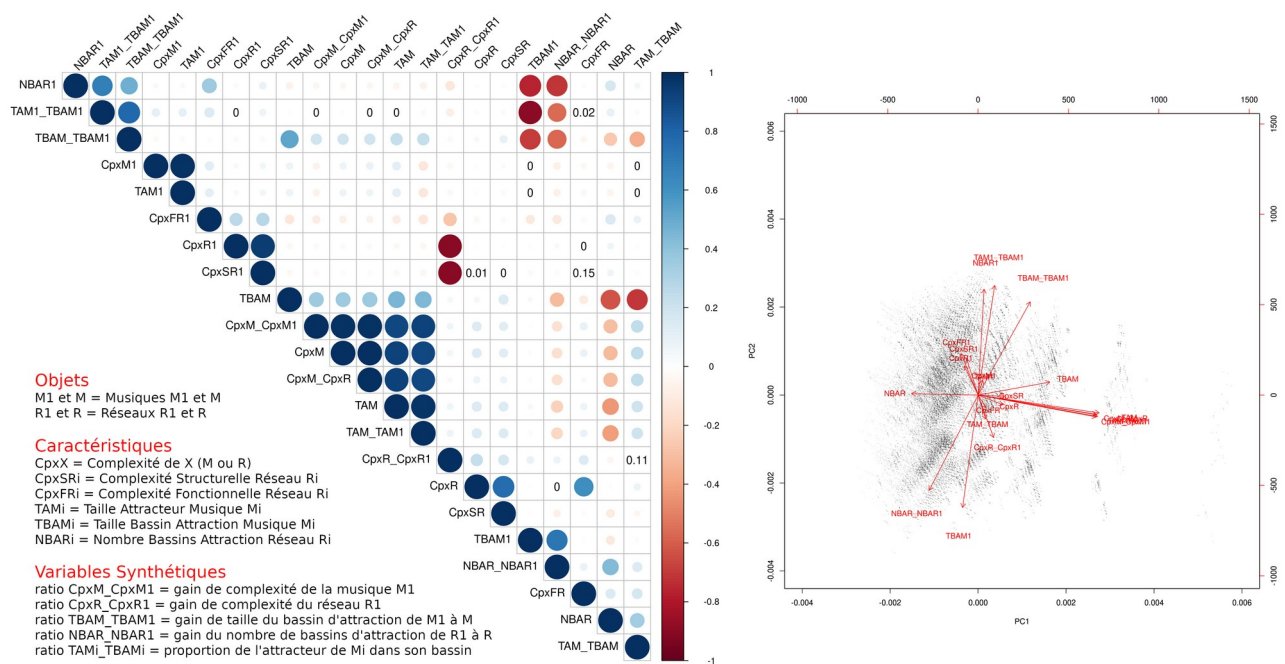


Figure 8.49. Corrélations des complexités et caractéristiques des réseaux et musiques. Ce corrélogramme (à gauche) est obtenu par calcul des corrélations deux à deux des caractéristiques des réseaux et musiques, ou de leurs relations, obtenues pour chaque quadruplet $\{R_i, M_j, R_k, M_l\}$ calculé. On voit apparaître certaines corrélations fortes, positives ou négatives entre complexités des réseaux et musiques, et les connaissances sur les bassins d'attraction (nombre de bassins d'attraction, taille des attracteurs, tailles des bassins d'attraction). Le code couleur et la taille des points indique le degré de corrélation (coefficient de corrélation). Les valeurs indiquées sont les p-valeurs supérieures à 10^{-20} ; toutes les autres corrélations, même faibles, sont significatives au seuil de 10^{-20} . On en tire par analyse en composantes principales (à droite) des groupes de variables cohérentes, mais aussi la contribution de chaque extension (points noirs) à ces composantes. De là, il est envisageable de faire de la classification des réseaux et extensions

Nous avons par exemple commencé par calculer les histogrammes 3D ou 4D (comptage des occurrences pour 3 ou 4 variables à la fois) pour représenter en fonction de complexités structurales et dynamiques le comptage des extensions (quadruplets $\{R_i, M_j, R_k, M_l\}$) comme montré dans la figure 8.48. On apprend ainsi qu'il y a beaucoup de réseaux R_i de complexité moyenne, que la plupart jouent des musiques M_j peu complexes, que les musiques complexes sont jouées par des réseaux

étendus complexes, que les musiques initiales complexes ou modérées sont étendues par des réseaux de complexité plus modérée, ... Néanmoins, ce type de figure reste difficile à analyser. De plus, ils ne montrent pas les relations possibles avec le nombre d'attracteurs des réseaux initiaux et étendus, ni la taille de l'attracteur des musiques initiales et étendues et de leur bassins d'attraction respectifs. Faire des statistiques comme étudier la corrélation linéaire entre ces complexités et ces renseignements associés sur les attracteurs (figure 8.49), réduire le nombre de dimensions par ACP⁴⁵, ... sont autant de précieux outils dont nous nous servons pour comprendre comment ces ensembles de réseaux se comportent, quelles relations les unissent, *etc.* Ce choix que j'ai fait il y a quelque temps de ne pas me focaliser sur un réseau particulier mais davantage de tous les étudier comme on étudie les êtres vivants qui constituent un écosystème devrait s'avérer payant. De la même manière qu'on ne peut comprendre correctement le fonctionnement d'un écosystème en focalisant son attention sur quelques espèces, même celles que l'on pense « clef », je crois qu'on a tout intérêt à ne pas rester rivé sur un cas particulier, tout intéressant soit-il, si l'on veut se faire une idée plus globale des grandes relations à l'œuvre dans ces espaces de réseaux, espaces dans lesquels les réseaux diffèrent par leur structure, les fonctions qu'ils mettent en œuvre, leur dimension, *etc.*, espaces dans lesquels ils évoluent.

Pour ne pas quitter ce chapitre sur une note trop générale, voici maintenant quelques résultats que je trouve déjà fascinants, quelques questions qui en émergent et leur apport en biologie des systèmes. En calculant les projections à 1 dimension des histogrammes à 3 ou 4 dimensions, on tire de très précieux enseignements sur la co-répartition des extensions dans les différentes complexité et leur corrélation avec d'autres observables (bassins d'attraction ...). Les figures 8.50, 8.51 et 8.52 en donnent un aperçu.

On voit très vite qu'il y a un lien fort entre la complexité des musiques, la taille de leurs attracteurs respectifs et celle de leurs bassins d'attraction, ce qui est attendu. En revanche on ne voit pas facilement de lien entre la tailles des musiques M_I (ou celle de leurs bassins d'attraction) et celle des musiques M (resp. leur bassin d'attraction). On remarque tout de même, d'après les figures 8.51 *haut* et *milieu* que les musiques M_I complexes ne se retrouvent prises en charge que dans les extensions de complexité moyenne à élevée, et qu'il semble y avoir un lien peu marqué entre la taille des bassins d'attraction de M_I et la complexité des réseaux étendus R , seuls les plus simples provenant de réseaux R_I ayant de grands bassins d'attraction pour M_I .

45 Analyse en Composantes Principales.

Projections from the 3D histogram (R1,M1,M)

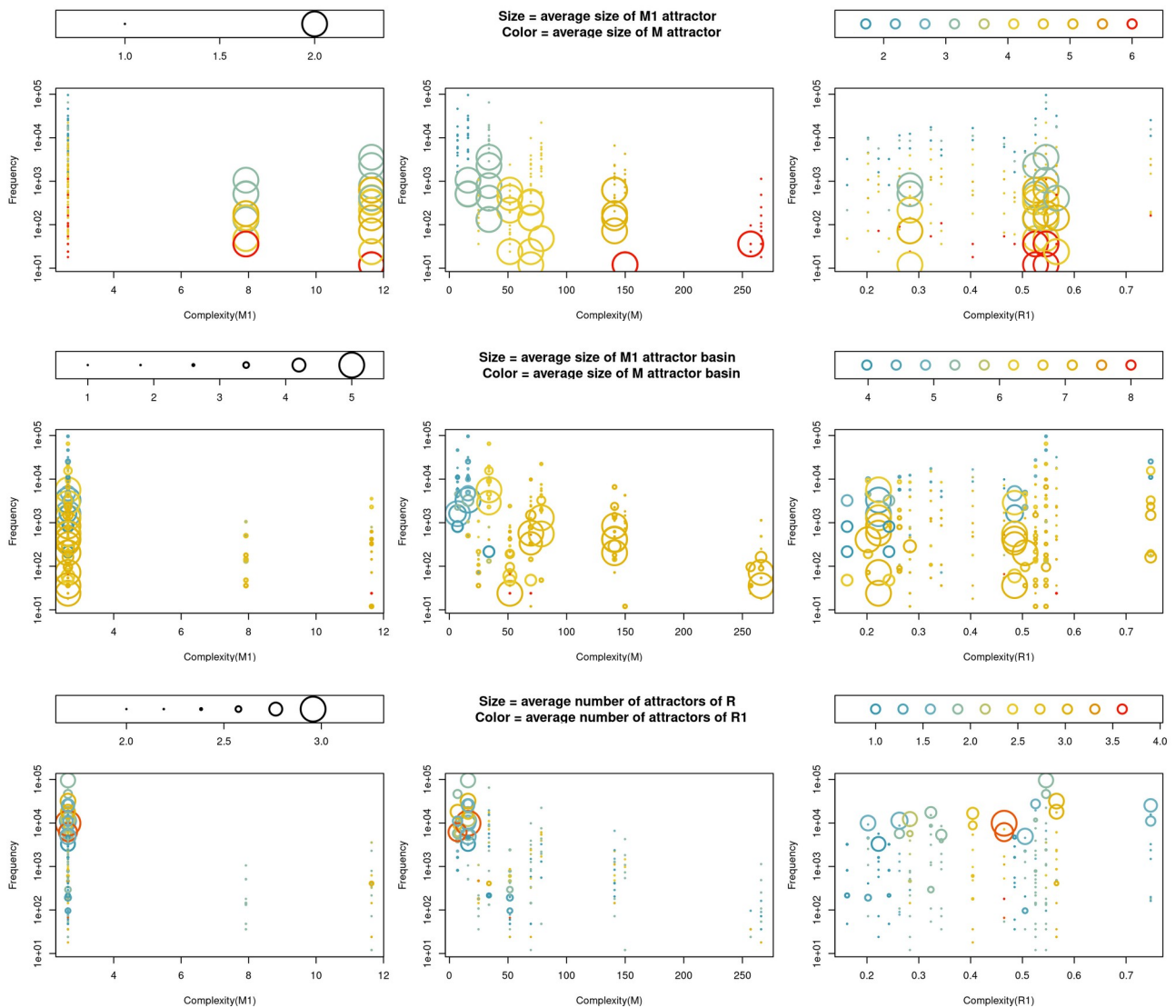


Figure 8.50. Projections 1D de l’histogramme 3D (R1,M1,M). Ces diagrammes sont des projections histogrammiques d’histogrammes 3D, ici celui montré figure 8.48. En plus des fréquences d’histogrammes classiques, ces 3 projections montrées en ligne donnent une information partagée entre eux. Dans chaque classe de ces histogrammes, il y a en effet plusieurs points, donc plusieurs fréquences possibles. Ces différentes fréquences résultent de la projection de l’ensemble des classes selon les 2 autres axes de l’histogramme 3D. En plus de ces informations, on affiche 2 autres valeurs via la taille des points et leur couleur. Chaque point représente la moyenne, dans une classe d’histogramme 3D $\{R_l, M_l, M\}$, de la variable représentée (les écarts types sont calculés mais non représentés). Il est important de comprendre que dans les 3 projections montrées sur une ligne, les fréquences projetées de l’histogramme 3D, montrées sous la forme de points colorés de tailles différentes, sont conservées (les tailles et couleurs des points sont conservées) mais dispersées différemment dans ces 3 projections. Cela permet de tracer une même information dans ces co-histogrammes, par exemple comment se distribuent les tailles des bassins d’attraction de M_l et M en fonction de la complexité de M_l , de M et de R_l . De plus, verticalement, dans une même projection, les points-fréquences sont positionnés au même endroit. Cette information permet donc de comprendre comment différentes observables (taille d’attracteurs, tailles de bassins, nombre de bassins, complexités structurale et fonctionnelle ...) sont corrélées à l’intérieur des projections. Ici on ne montre que 3 groupes de projections, mais d’autres sont possibles. L’interprétation de ce graphique se trouve dans le texte.

Parmi les observations les plus surprenantes qu'on puisse faire (figure 8.50 en *haut*), il y a le fait que les musiques M_i complexes (qui sont les plus longues en moyenne) sont jouées par des réseaux R_i de complexités variables, mais bien déterminées : on les retrouve jouées par les réseaux de complexité 0.27 puis par les réseaux de complexité comprise entre 0.5 et 0.6 ; tous les autres jouent des musiques simples. Cette observation est encore plus marquée avec la taille des bassins d'attraction (qui contiennent tous les états menant à l'attracteur ainsi que ceux de l'attracteur) : il y a des classes de complexité particulières de réseaux dont la taille moyenne des bassins d'attraction est grande (voir figure 8.50 au *milieu*). Ces classes ne sont pas les mêmes que celles qui produisent des musiques complexes. Les réseaux produisant des grands bassins d'attraction (un marqueur de robustesse) sont situés principalement dans les tranches de complexité de R_i de 0.22 et 0.48.

Projections from the 3D histogram (R,M1,M)

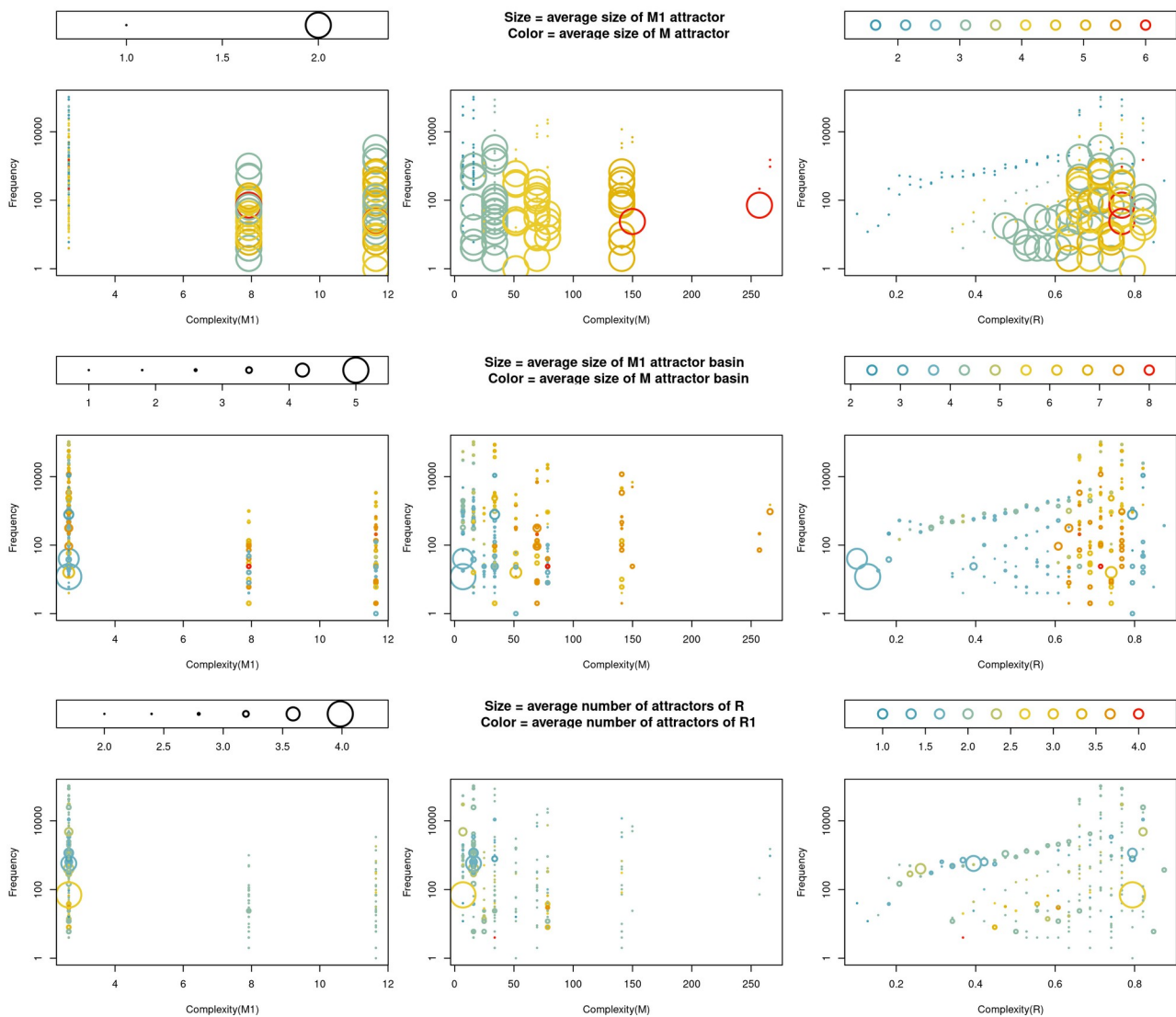


Figure 8.51. Projections 1D de l'histogramme 3D (R,M1,M). Voir figure 8.50.

On observe aussi que d'autres réseaux R_l assurent une robustesse d'importance intermédiaire et que ces classes de réseaux R_l sont adjacentes aux classes produisant les plus grands bassins d'attraction.

On retrouve ce comportement concernant les réseaux étendus R , leurs musiques et tailles de bassin d'attraction : ce sont les réseaux R de complexité comprise entre 0.6 et 0.8 qui jouent les musiques les plus complexes et qui ont les bassins d'attraction les plus importants.

Le nombre de bassins d'attraction également semble avoir un lien non trivial avec les complexités des réseaux qui les génèrent. C'est bien visible pour les réseaux R_l (figure 8.50 *bas*) : il y a des tranches de complexité de réseaux R_l (par exemple 0.4, 0.46, 0.57) qui génèrent un nombre moyen d'attracteurs importants ; d'autres au contraire ne génèrent qu'un nombre faible d'attracteurs (ex : les tranches 0-0.26 ou 0.48-0.55). Cela semble être le cas mais c'est moins marqué pour les réseaux R (qui sont de dimension supérieure, $d=3$). Entre les deux, une différence importante vient de l'augmentation très importante de l'étendue des complexités fonctionnelles (celles liées exclusivement à la complexité des fonctions Booléennes). En dimension 2, il y a 4 complexités possibles pour les fonctions Booléennes, mais elles ne diffèrent que d'une valeur infime ; en dimension 3, il y a un nombre bien plus important de complexités fonctionnelles.

Pour moi, cette distribution inattendue des tailles de musiques, de bassins d'attraction et de nombre d'attracteurs est une information cruciale qu'il faut explorer davantage. Ces distributions font penser qu'il puisse exister de grandes catégories de réseaux R structurés d'une certaine manière (arrangement des fonctions entre elles) ou possédant un jeu particulier de fonctions (certaines n'ayant que peu d'entrées tandis que d'autres en ont plus, permettant à certains arrangements de se faire, d'autres n'étant pas possibles), fonctions qui contrôlent certains aspects de la dynamique, de telle manière que ces réseaux adoptent une stratégie de robustesse (grand bassin d'attraction) ou plutôt de complexité comportementale. C'est une hypothèse qu'il faudra vérifier en analysant la composition et la structure des réseaux. D'après la figure 8.52, on voit que la complexité d'un réseau est principalement due à sa complexité structurelle, mais on voit aussi que les complexités fonctionnelles ne se répartissent pas non plus de façon uniforme. On retrouve des bandes de complexités de réseaux ayant une faible complexité fonctionnelle (petits points) et d'autres une complexité fonctionnelle importante (grands points). Il faut donc s'intéresser à ces groupes de réseaux particuliers qui ont de grands bassins d'attraction (ou ceux qui jouent des musiques M_l complexes) et voir comment ils sont constitués (Quels sont leurs arrangements ? Quelles fonctions Booléennes les composent?) et les comparer à ceux qui ne montrent pas de comportement semblable.

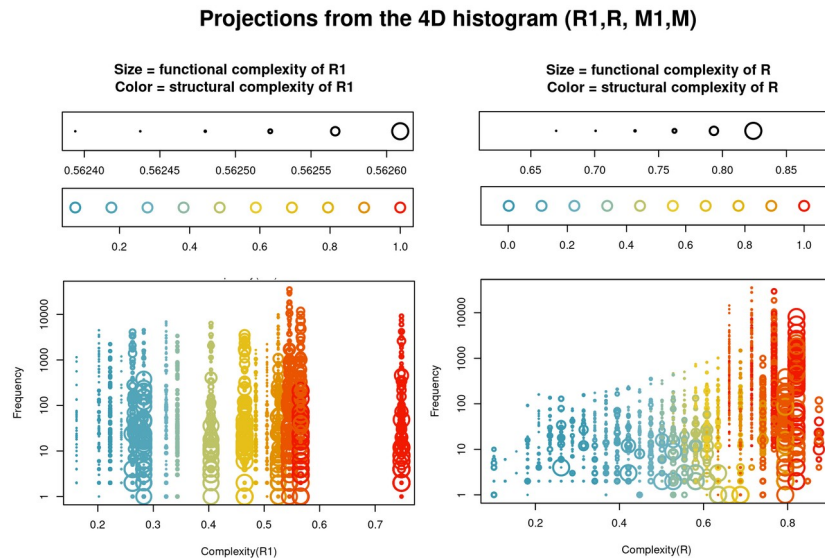


Figure 8.52. Projections 1D de l’histogramme 4D (R1,R,M1,M). Voir figure 8.50. Ici les projections ne se font que sur les axes de complexité des réseaux R_1 et R . Les 2 dimensions supplémentaires (taille et couleur des points) concernent les complexités fonctionnelle et structurelle de ces réseaux.

Revenons au problème posé, celui de l’extension. Dans ce cadre, il y a de nombreuses autres questions qui peuvent être posées, par exemple : est ce qu’un réseau produisant un grand bassin d’attraction continue, après extension, de produire un grand bassin d’attraction ? Ou bien est ce que sa complexité est mise à profit d’un autre critère ? Mêmes questions pour les autres observables. Combien persistent dans leur comportement (robustesse vs multi-attractivité vs complexité dynamique) ? Ceux qui affichent une telle persistance se ressemblent-ils entre eux ? ... De même, est-ce qu’un réseau possédant une certaine combinaison {*complexité fonctionnelle*, *complexité structurelle*} sera étendu (avec la contrainte de conservation partielle de la dynamique, comme dans la problématique présentée ici) en un réseau de plus grande dimension conservant une telle combinaison ?

Ces questions nombreuses qui touchent à la relation structure-fonction des machines sont ici traitées avec des réseaux de faible dimension pour la simple raison que l’intégralité des extensions est calculable. Ce n’est plus le cas dès qu’on passe sur des extensions de dimension 3 vers 4 ou plus ; il y en a trop. Pourtant, il faudra bien que nous explorions ces plus grands réseaux pour éviter des effets liés à leur faible dimension (par exemple la faible diversité des complexités fonctionnelles des réseaux de petite taille) et à la taille trop limitée de leurs dynamiques (taille des musiques) qui, là aussi, diffèrent trop peu entre elles. Parmi les solutions évoquées dans notre article [Segretain ; 2020 a et b] nous évoquons la possibilité de réaliser de tels calculs, soit par tirages aléatoires dans l’espace des réseaux (espace de paramètres de fonction et espace des layouts), soit par tranches de complexités de réseaux. C’est de telles études que nous pourrions vraiment tirer des enseignements sur l’évolutivité des réseaux par extension de structure et de dynamique, et comprendre quelles relations unissent complexité structurelle, fonctionnelle, robustesse, multi-attractivité, évolutivité.

L’autre option qui permet d’explorer l’évolutivité des réseaux et comportements, c’est celle de la vie

artificielle ; c'est une manière aussi d'appréhender certains aspects du passage d'un possible vivant fragile aux origines du vivant à un vivant plus complexe, plus robuste parfois ; c'est enfin une manière de rentrer peu à peu dans la théorie de l'atavisme du vivant, en particulier celle liée au cancer. C'est tout ceci que je présenterai dans la section suivante.

8.3.6. Il y a du mou dans les réseaux

De la série de questions élémentaires sur les réseaux qui a précédé dans la section 8.3.5 et dans ce chapitre en général, émerge une vision d'ensemble dont je vais maintenant peindre plus précisément le tableau. Le mou dans les réseaux, c'est des réseaux plus variés dans leur fonctionnement physique, mais aussi moins nets qu'il n'y paraît, moins forts, plus diffus. C'est aussi les voir dans un cadre évolutionnaire, à savoir suivre la trajectoire de systèmes d'interaction biologique au cours de leur complexification ou leur emboîtements successifs depuis leur apparition pour former des systèmes plus complexes. C'est donc réfléchir à cette vision du vivant et de son évolution ; c'est réfléchir aux conséquences que cela peut avoir dans la conception de problèmes actuels comme le cancer par exemple.

C'est cette vision qui servira de contexte à mes futures recherches et à celles que je proposerai à mes étudiants. Ce sont ces propositions qu'il faudra évaluer, confronter à la simulation, à l'observation, mais ce sont ces mêmes propositions qui auront peut-être pour mérite de stimuler chez les jeunes chercheurs l'imagination et la volonté de déterminer *une vérité* sur la nature du vivant.

8.3.6.1. Réseaux flous, interactions diffuses, canaux de communications amorphes

Précédemment, nous avons longuement parlé réseaux, mais pas du milieu qui véhicule les informations entre les éléments de ce que l'on identifie comme réseaux d'interaction biologique, ni même de la façon dont se propagent ces informations. Avant de revenir aux réseaux, sans trop développer cette partie (il y aurait trop à dire sur chaque situation que je vais présenter), je vais donc commencer par discuter un peu du substrat et des moyens de communication dans le vivant.

Le sujet est vaste évidemment car, par exemple, sont inclus dans les réseaux d'interaction les écosystèmes (et sociétés dans ces écosystèmes) avec de multiples individus de mêmes espèces ou d'espèces différentes qui communiquent par un très grand nombre de moyens (sonores, chimiques, lumineux, mécaniques, *etc*) qui font chacune l'objet d'études poussées, ou encore les réseaux de neurones. Par habitude (il paraît évident que les individus d'une société ou les espèces d'un écosystème constituent un réseau) ou parce que la morphologie ne laisse pas de doute là dessus (les neurones sont en réseau, mais aussi les *mycelia*) beaucoup de systèmes naturels tels que ceux précités nous apparaissent en réseau et leur fonctionnement (communication animale, neurotransmetteurs, *etc*) nous semblent bien identifiés et clairs. Ce n'est pas toujours aussi simple ni évident qu'on ne le pense (on en apprend toujours beaucoup sur la communication animale par exemple, même chez des espèces qu'on connaît bien comme les abeilles), mais je vais laisser ces systèmes plutôt bien explorés pour mettre davantage l'accent sur d'autres qui le sont moins et dont

le mode de communication est moins évident.

Quand j'étais à l'université, dans les cours de biologie j'aimais bien les cours de signalisation intracellulaire. Je trouvais autant fascinant que gênant qu'on me présente le système d'adressage comme un mini système postal avec ses adresses protéiques portées par des vésicules, des modifications post-translationnelles permettant une reconnaissance spécifique par le receveur, et même les postiers qui pédalent de maison en maison pour amener les lettres à bonne adresse, bien guidés par un système d'aiguillage (protéines AP1, AP2, COP ...). Ce qui était formidable effectivement, en plus des adresses, c'était le rôle joué par le cytosquelette de microtubules ; avec eux, on avait en plus des autoroutes sur lesquelles se promènent, dans 2 directions (sens centrifuge, rétrograde et sens centripète, antérograde ; voir [Hirokawa, 1998]), les vésicules chargées de matériaux et portées par des moteurs moléculaires (et leur cortège d'adaptateurs) comme la kinésine et la dynéine. Cette belle image a d'ailleurs ému beaucoup de monde, pas forcément en bien, lorsqu'elle fut illustrée en 2012 dans la vidéo *BioVision « The inner Life in a Cell »* par l'université de Harvard [BioVision, 2012], une superbe infographie qui s'est vue reprocher de transmettre une vue très mécaniste de l'activité interne de la cellule, un ballet de micro-machines bien programmées. Finalement c'est une pensée rassurante de savoir que beaucoup n'ont pas adhéré à cette vision pourtant très jolie et séduisante [Nicholson, 2013, 2014]. Ces systèmes d'adressage et de transport intra-cellulaire existent vraiment, sont plutôt bien compris et marchent effectivement très bien. Si j'ai volontairement choisi ce ton un rien piquant c'est parce qu'il s'agit là d'une machinerie « évidente » à rechercher et qu'aurait pu imaginer un ingénieur s'il avait du penser les transports dans la cellule. Pourquoi pas après tout ! L'évolution est fort inventive et il n'est pas si étonnant que de tels systèmes très raffinés soient apparus. Je voudrais toutefois attirer l'attention sur d'autres modes de communication intra-cellulaires possibles, des modes plus simples, moins spécifiques, mais aussi compatibles avec des formes de vie plus primitives et ayant pu émerger très tôt dans le vivant avec les premières structures fibrillaires notamment. C'est ce dont je vais discuter ici.

8.3.6.1.a. *Trail systems*

Le vivant (les sociétés, les organismes, les cellules, les molécules), par son action, laisse des traces de son passage et ces traces peuvent être exploitées en retour. Ceci est connu sous le nom de *trail systems*. Cousins des *boids* (modèles numériques de systèmes collectifs comme les nuées d'oiseaux, *fish schools*, etc), il en existe à de nombreuses échelles et nous en connaissons d'ailleurs tous plusieurs à l'échelle humaine ou animale [Helbing, 1997 ; Theraulaz, 2002 ; Kroo, 2004 ; Goldstone, 2006 ; Jonhson, 2006 ; Nishinaria, 2006 ; Davies, 2007 ; Schulte-Pelkum , 2007; Glade, 2009] : les chemins émergents par piétinement de l'herbe, les chemins dynamiques marqués par des cairns en haute montagne, les colonnes de fourmis, etc (voir figure 8.53). Ils se forment parce que les individus (les agents du système biologique) laissent une trace (chimique, visuelle, ...) de leur passage dans l'environnement dans lequel ils se déplacent et que ces traces sont perçues par d'autres et mises à profit pour orienter leurs futurs comportements, leurs déplacements en particulier. Cette

trace (*trail*), soumise aux aléas de l'environnement (diffusion, repousse de l'herbe, dégradation de la trace chimique, intempéries, *etc*), finit par se dissiper et disparaître peu à peu. Au contraire, lorsque un agent passe sur une trace existante, il a tendance à la suivre (l'agent dispose toutefois d'un certain libre arbitre et n'est pas totalement contraint de suivre cette trace ; cela lui permet d'explorer le reste de l'environnement, là où il n'y a pas de traces) et à renforcer celle-ci (par exemple à reconstruire un cairn ou rajouter des cailloux dessus, réémettre des phéromones, écraser un peu plus l'herbe, *etc*). La nature éphémère de la trace est un avantage ; elle rend le système dynamique et non figé. Une trace peu utilisée (car désignant une trajectoire devenue dangereuse ou inaccessible, par exemple en montagne, ou ne menant plus à de la nourriture dans le cas des fourmis, ou plus simplement parce qu'elle devient imperceptible) donc peu renforcée disparaîtra, ne sera donc plus utilisée et en conséquence évitera à des agents du système de se mettre en danger (montagne) ou de se mobiliser inutilement pour chercher de la nourriture qui n'est plus présente.

C'est une forme de système de réaction-diffusion avec mémoire. Les agents du système naturel « écrivent » dans l'environnement et leur trace constitue une forme de mémoire éphémère, mémoire confirmée par d'autres agents qui, collectivement « perçoivent » l'environnement.



Figure 8.53. Exemples de *trail systems* à 3 échelles différentes. (A) Échelle 10-1000 m : *trail systems* humains : à gauche des chemins émergents dans un parc [Helbing, 1997], (B) Échelle 10^{-3} - 10^{-1} m : une structure émergente (fourmilière) et le *trail system* associé, une colonne de fourmis.

Certains systèmes à l'échelle cellulaire ou sub-cellulaire sont susceptibles de se comporter comme des *trail systems* (figure 8.54). Il semble bien que certaines amibes comme *Dictyostelium discoideum* [Bagorda, 2004] produisent et suivent effectivement des traces persistantes d'AMP cyclique à l'instar de fourmis avec des traînées de phéromones. Certaines bactéries comme *Listeria monocytogenes* (et même certains virus comme *Vaccinia virus*) expriment des récepteurs pour l'actine (ARP2-3) à leur surface et peuvent ainsi détourner le cytosquelette d'actine à leur profit [Alberts, 2004 ; Boukella, 2004 ; Welch, 2013 ; Theriot ; 2004 & 2016]. Propulsées par l'ajout de fibres entre elles et le faisceau d'actine, ces particules d'à peine 1 à 2 μm de long se déplacent à une vitesse d'environ 0.2 $\mu\text{m}/\text{sec}$ selon des trajectoires harmonieuses qui peuvent les amener à pousser sur la membrane et, à terme, sortir de la cellule en rompant la membrane cellulaire. Le faisceau d'actine polymérisée, formé de milliers de filaments d'actine filamenteuse (F) reliés par des

protéines intermédiaires, se dépolymérise lentement à l'arrière de la traînée, libérant de grandes quantités locales d'actine globulaire (G) qui peut, après rechargement par une molécule d'ATP, être remise à contribution pour former de nouveau filaments. C'est en cela que ce système pourrait s'apparenter à un trail system : si la trace d'actine dépolymérisée est suffisamment marquée et persistante, alors elle pourrait contribuer à recruter une particule microbienne capable d'initier la polymérisation d'une comète d'actine ; une telle particule pourrait se voir propulsée préférentiellement dans la trace persistante d'une comète passée. Les vidéos réalisées par J. Theriot sont en ce sens troublantes (voir figure 8.55) : elles montrent des particules adoptant des comportements de regroupement, de suivi, d'évitement, semblables à ceux que j'avais proposés (à tort) pour expliquer l'auto-organisation des microtubules *in vitro* par des processus de réaction-diffusion (voir mes travaux antérieurs). Notons qu'il existe des cas de comètes fondées sur les faisceaux de microtubules [Brito, 2005].

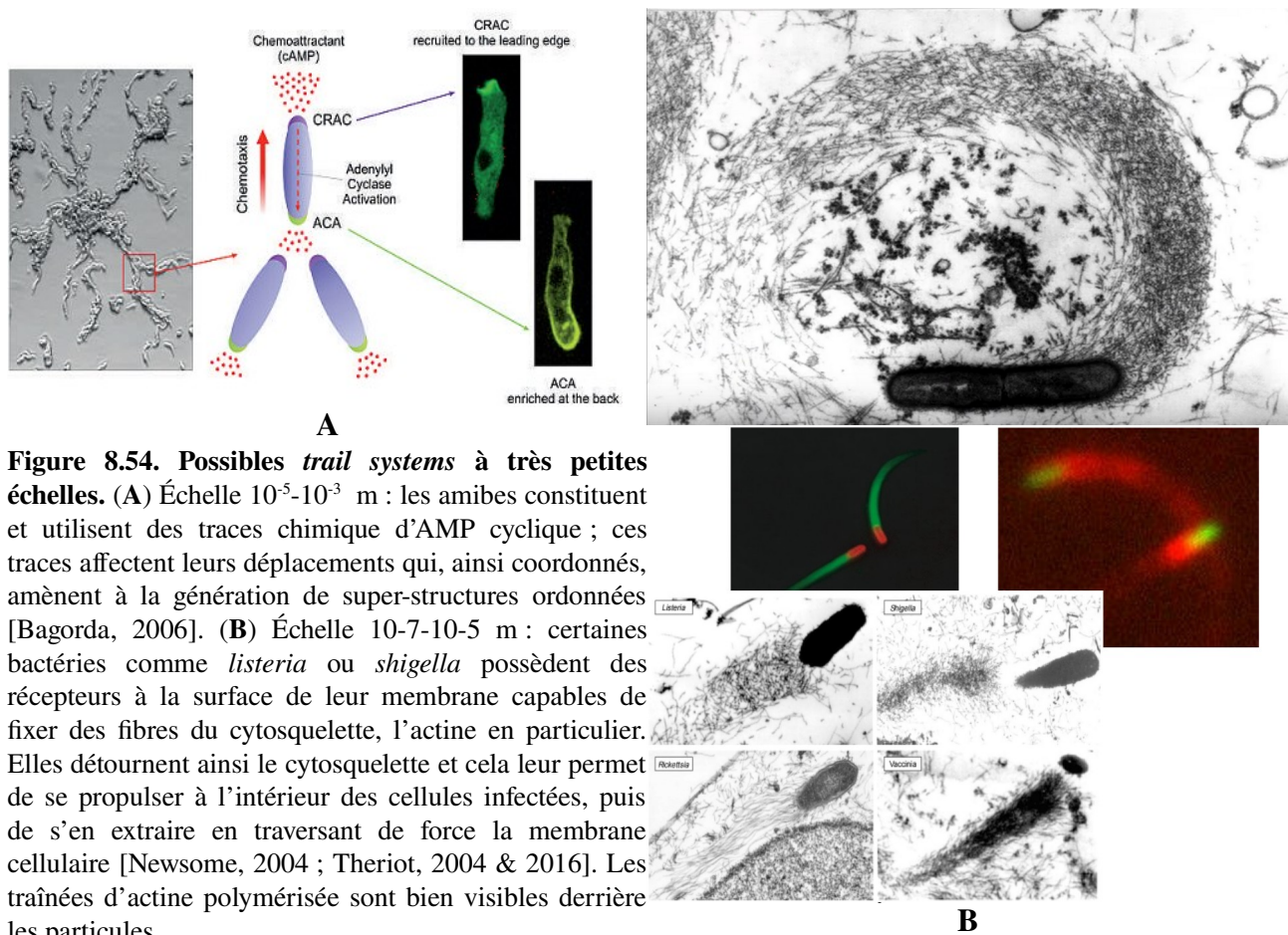
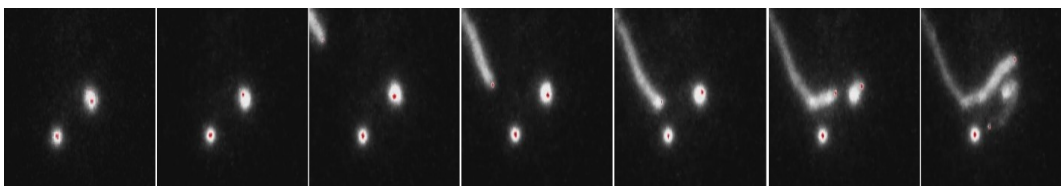
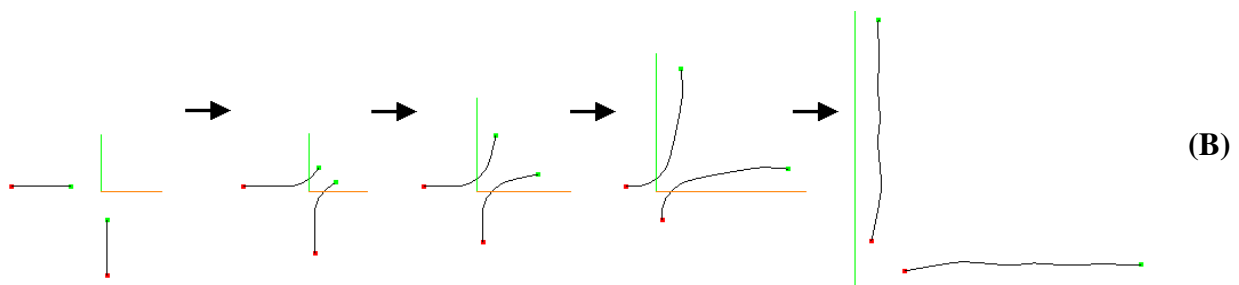
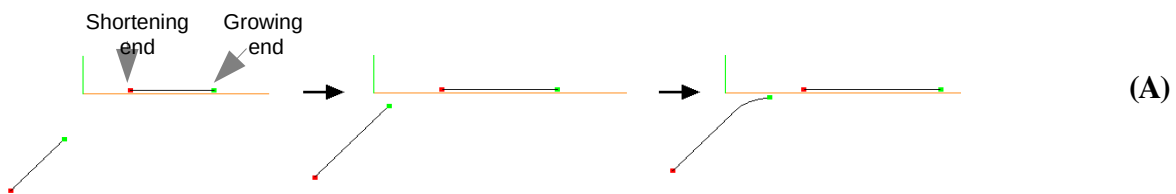
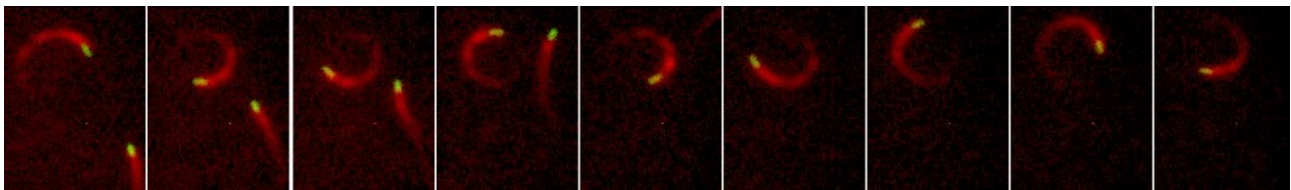


Figure 8.54. Possibles trail systems à très petites échelles. (A) Échelle 10^{-5} - 10^{-3} m : les amibes constituent et utilisent des traces chimiques d'AMP cyclique ; ces traces affectent leurs déplacements qui, ainsi coordonnés, amènent à la génération de super-structures ordonnées [Bagorda, 2006]. (B) Échelle 10^{-7} - 10^{-5} m : certaines bactéries comme *Listeria* ou *Shigella* possèdent des récepteurs à la surface de leur membrane capables de fixer des fibres du cytosquelette, l'actine en particulier. Elles détournent ainsi le cytosquelette et cela leur permet de se propulser à l'intérieur des cellules infectées, puis de s'en extraire en traversant de force la membrane cellulaire [Newsome, 2004 ; Theriot, 2004 & 2016]. Les traînées d'actine polymérisée sont bien visibles derrière les particules.

Si l'on pousse ce raisonnement à une échelle plus petite, ce que j'ai fait il y a une vingtaine d'années durant ma thèse sur l'auto-organisation des microtubules, on peut imaginer les systèmes microtubulaires comme de possibles trail systems : les microtubules, ces tubes de 20 nm de diamètre et pouvant atteindre plusieurs micromètre de long, sont des assemblages supramoléculaires de tubuline, une hétéroprotéine globulaire liée à un nucléotide, le GTP ou le GDP. L'assemblage des

hétérodimères de tubuline-GTP se fait aux deux extrémités, de même que le relargage de tubuline-GDP. Les extrémités des microtubules sont en instabilité dynamique [Mitchison, 1984], c'est-à-dire que, brutalement, elles peuvent passer d'une phase d'assemblage à une phase de désassemblage (*catastrophes*), et réciproquement (*rescues*). Cette instabilité dynamique commence à être plutôt bien comprise et est liée d'une part à des phénomènes de capping par la tubuline-GTP (accumulation de tubuline-GTP stabilisante), d'autre part (et c'est lié) à une structure et un comportement biomécanique local différent des extrémités [Margolin, 2012 ; VanVuren, 2005]. Les deux extrémités du tube n'ont pas le même comportement : l'instabilité dynamique est plus forte à l'extrémité dite négative, tandis que l'autre est plus stable. Il en résulte que dans certaines conditions, les microtubules adoptent un comportement dit de tapis roulant (*treadmilling* [Rodionov, 1997]) au cours duquel les microtubules croissent par assemblage au niveau des extrémités positives, tandis qu'il décroissent en même temps par désassemblage au niveau de leurs extrémités négatives.



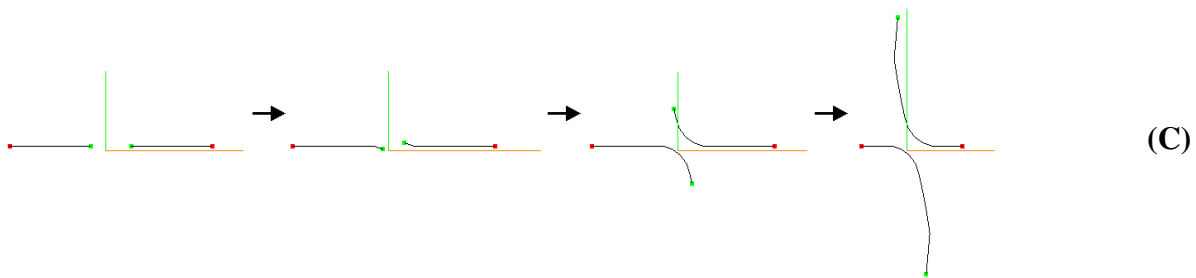


Figure 8.55. Extraits de vidéos de comètes d'actine. Les images sont extraites de vidéos réalisées dans le laboratoire de Julie Theriot. Sous chaque série d'images, je montre des simulations utilisant le même modèle de trail systems que celui fait pour les microtubules [Glade, 2002], redimensionné pour des comètes d'actines. Ces simulations ne reproduisent pas exactement le comportement mais illustrent les interactions entre particules-comètes. **(A)** Une comète tourne sur elle-même comme si elle suivait sa propre trace. La simulation montre un phénomène de suivi. **(B)** Deux comètes convergent vers un même point et, avant de s'atteindre, se repoussent. Ceci peut s'expliquer par le fait que le milieu autour des particules en train de recruter de l'actine G est déplété en actine (phénomène de pompage). C'est ce que reproduit parfaitement la simulation. **(C)** 2 particules sont dans un état stationnaire (il n'y a pas de direction de poussée préférentielle de l'actine polymérisée) lorsqu'une 3ème particule arrive à proximité. A ce moment, la particule en déplacement change brutalement de direction tandis qu'une particule stationnaire se met aussi brutalement à bouger, peut-être par un phénomène d'évitement (comme en **B** ou montré dans la simulation associée).

Ce processus pourrait, dans certaines conditions de réaction (l'instabilité dynamique des microtubules étant très sensible à certains ions comme le magnésium ou le calcium), amener des microtubules ou des groupes de microtubules (*bundles*) à adopter un comportement de trail systems, y compris dans les cellules vivantes pour lesquelles une partie des microtubules sont sous forme libre (l'autre étant reliée au centrosome). C'est ce qui a servi d'hypothèse de travail pour la modélisation que j'avais proposée pour expliquer l'auto-organisation des microtubules (voir mes travaux antérieurs). Bien que j'ai montré que les traces de tubuline-GDP laissées par les microtubules se désassemblant sont trop faibles pour supporter cette hypothèse lorsque la réactivité est faible (réactivité microtubulaire réduite à cause de faibles concentrations en magnésium et présence d'eau lourde D_2O stabilisante) [Glade, 2012], de telles traces pourraient bien être plus importantes lorsque la réactivité des microtubules augmente (concentrations plus fortes en magnésium) et que les microtubules sont groupés (*bundles*). Ces conditions pourraient être celles dans lesquelles s'étaient placés Pirollet *et al* et Carlier *et al* [Carlier, 1987 ; Pirollet, 1987] puis Mandelkow *et al* [Mandelkow, 1989]. En particulier, l'équipe de Mandelkow obtenait la formation de vagues de concentration en microtubules dans sa solution et les détails microscopiques montraient que les microtubules, d'une vague d'assemblage à l'autre, se formaient au même endroit (les nouveaux microtubules occupaient la même place que leur prédécesseurs désassemblés (figure 8.56)).

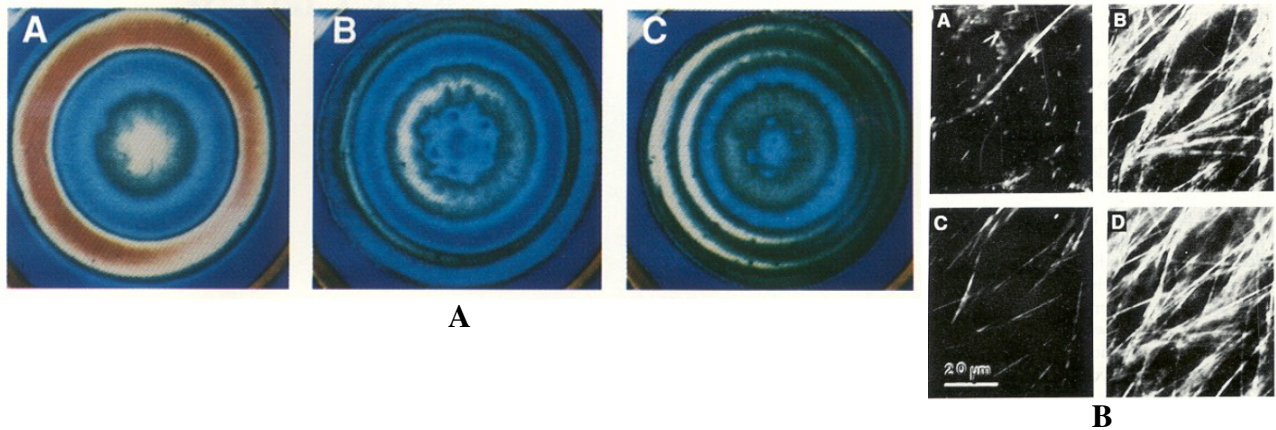
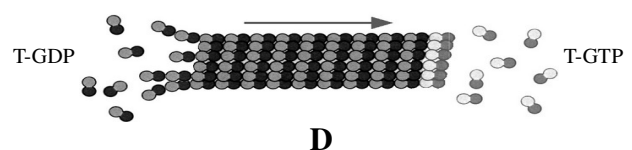


Figure 8.56. Les microtubules, des trail systems ? (A) et (B) sont extraits de [Mandelkow, 1989]. On y voit (A) les motifs de vagues de concentrations de microtubules se formant dans une solution de faible épaisseur (200 μl de solution dans un container de 22 mm de diamètre) très concentrée en tubuline (46 g/l), les photos étant prises à 189s, 499s et 949s du début de la réaction, initiée à la mise en température (37°C) de la solution. En (B) on y voit un gros plan (large de 100 μm) pris sous microscope à 4 moments différents alternant périodes de solution désassemblée (photos A et C) et assemblée (photos B et D). On constate que les microtubules assemblés en B puis désassemblés en C, sont assemblés avec la même disposition en D.

C'est notamment ceci, le fait que Tabony et Job avaient montré que le processus d'auto-organisation des microtubules était dissipatif et le fait que Mandelkow lui-même soutenait l'idée qu'il puisse s'agir d'un phénomène semblable à celui de Belousov-Zhabotinski, c'est-à-dire un processus de réaction-diffusion hors-équilibre thermodynamique, qui a motivé le choix d'un modèle de réaction-diffusion pour expliquer l'auto-organisation des microtubules. Un modèle réalisé par Sept et al [Sept, 1999a & 1999b], fondé sur un processus de réaction diffusion reproduit d'ailleurs très bien ce que Mandelkow *et al* observaient (phénomènes spatiaux et dynamiques). Nous (J. Tabony et moi) avons donc proposé que les microtubules, à l'échelle microscopique, pourraient se comporter comme des fourmis ou comme décrit précédemment pour les comètes d'actine : un microtubule s'assemble en consommant localement de la tubuline-GTP, relâche derrière lui une trace de tubuline-GDP qui, par échange du GDP par du GTP libre présent dans le milieu (soit en mettant dans la solution de départ une concentration importante de GTP, soit en permettant sa régénération par une enzyme), permet la formation d'une trace de tubuline-GTP pouvant permettre la nucléation et l'assemblage (ou l'élongation, ou pouvant affecter la réactivité) dans une direction préférentielle de nouveaux microtubules (resp. de microtubules déjà formés) et, peu à peu, biaiser l'orientation générale des microtubules dans une solution jusqu'à former une organisation complexe comme celles montrées figure 8.57.



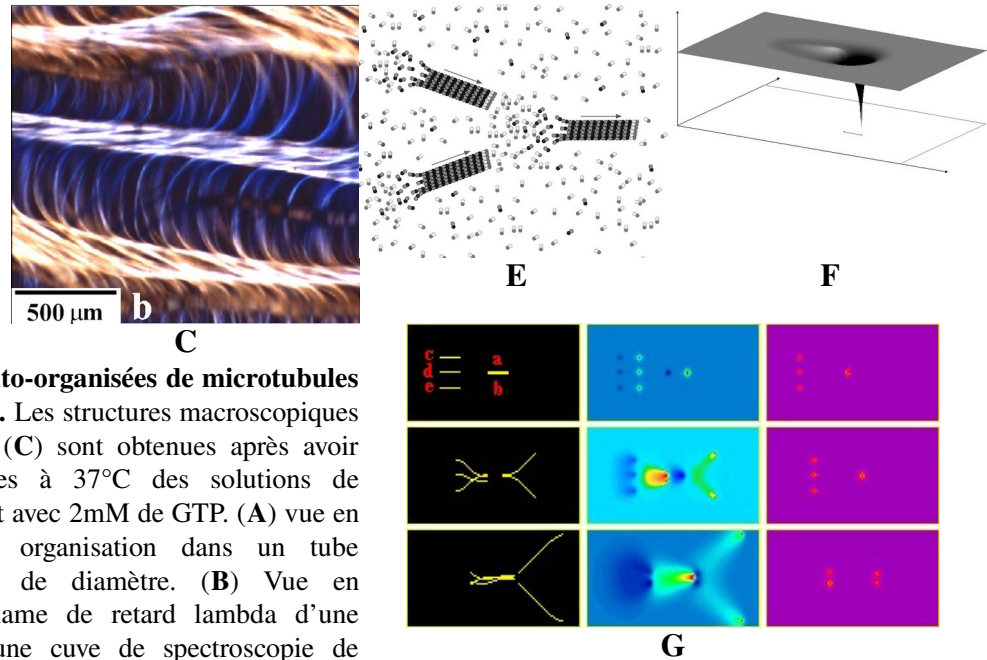


Figure 8.57. Solutions auto-organisées de microtubules et modèle de trail system. Les structures macroscopiques montrées en (A), (B) et (C) sont obtenues après avoir maintenu quelques heures à 37°C des solutions de tubuline à 10 g/l contenant avec 2mM de GTP. (A) vue en lumière réfléchiée d'une organisation dans un tube cylindrique de 1.5 cm de diamètre. (B) Vue en biréfringence avec une lame de retard lambda d'une organisation faite dans une cuve de spectroscopie de 40x10x1 mm.

Les couleurs jaune et bleue renseignent sur l'orientation à 45 ou 135° des microtubules dans la solution. (C) Un gros plan en biréfringence + lambda sous microscope.

En (D) à (G) des simulations de microtubules fondées sur le modèle de trail systems [Glade, 2002]. (D) Schéma de la dynamique d'un microtubule. (E) Schéma d'interaction de microtubules en train de croître dans la traînée laissée par un microtubule en amont. (F) Simulation d'un microtubule en plein treadmilling : la surface montre la concentration en tubuline-GTP. A l'avant du microtubule (à droite), on voit une déplétion en tubuline-GTP (T-GTP) dans la solution. A l'arrière on voit un léger renflement blanc qui montre une région un peu plus concentrée en T-GTP, après régénération de la tubuline-GDP. (G) Avec le même modèle, simulation à 3 instants (3 lignes d'images) différents de l'interaction de 5 microtubules. La colonne de gauche montre les microtubules (des solides articulés), la colonne du milieu montre la concentration en T-GTP, celle de droite la T-GDP. Les concentrations locales en T-GTP et T-GDP permettent un dialogue entre les agents-microtubules.

Par simulation, dans les conditions de réaction de nos expériences d'auto-organisation microtubulaires (c'est-à-dire finalement peu réactives par rapport à celles de Mandelkow, de Carlier et de Pirollet), j'ai montré que les traces de tubuline-GDP laissées par un ou plusieurs microtubules rangés en faisceau sont trop peu concentrées et se dissipent trop vite (voir figures 7.3 et 7.4). En revanche, comme discuté dans l'article [Glade, 2012], il est envisageable que des traînées formées par des faisceaux de microtubules puissent avoir une influence dans des situations plus réactives comme celles dans lesquelles Mandelkow *et al* ont obtenu un milieu excitable avec des microtubules.

Je m'interroge donc toujours sur l'effectivité et le possible rôle que pourraient jouer de tels comportements de communication intra-cellulaire. Selon les périodes de la vie cellulaire, la concentration en microtubules mais également la stabilité réactionnelle dans le cytoplasme varie. Ceci peut être étudié, en particulier en reprenant mes anciennes simulations à l'échelle microscopique, en les appliquant à des concentrations importantes en microtubules et en testant diverses situations réactives.

8.3.6.1.b. Diffusion anisotrope

Les trails systems à l'échelle sub-cellulaire restent hypothétiques. Il y a un autre moyen par lequel les fibres du cytosquelette peuvent contribuer à l'établissement d'une communication intra-cellulaire : la diffusion biaisée.

D'abord, un tube, c'est un canal ! Chaque microtubule constitue un canal d'environ 16 nm de diamètre et de plusieurs micromètres de long. Dans un tel canal, la diffusion de petites molécules, les ions notamment, mais également des molécules de quelques centaines de Daltons⁴⁶ (vitamines, drogues ...) à ... quelques centaines de kDa (tubuline), se fait de façon quasiment unidimensionnelle. Cette idée n'est pas nouvelle et a été développée et simulée par D. Odde en 1998 [Odde, 1998] : Odde a montré, par simulation, que le paclitaxel (taxol), une molécule d'environ 850 Da à usage anti-cancéreux, stabilisatrice des microtubules (elle réduit l'instabilité dynamique) dont le site de fixation se trouve sur la Béta-tubuline à l'intérieur du tube du microtubule [Alushin, 2014], pouvait diffuser à l'intérieur des microtubules. Il a fait de même pour la tubuline, bien plus grosse que le taxol puisque sa masse est d'environ 100 kDa (l'hétérodimère a une dimension d'environ 8x4x4nm), ou des immunoglobulines G (IgG) des molécules d'environ 150 kDa. Il montre que ces molécules diffusent assez bien à l'intérieur des microtubules et que l'équilibre de concentration avec le milieu extérieur peut être atteint en quelques minutes seulement pour des microtubules de quelques dizaines de micromètres.

Ensuite, l'ensemble des fibres du cytosquelette sont souvent organisées en faisceaux (d'actine et de microtubules). Dans un faisceau de microtubules, les tubes sont alignés et espacés d'environ 20 à 30 nm (équivalent à une largeur de tube). Ces faisceaux de fibres ne sont pas distribués au hasard dans les cellules. Ils sont au contraire organisés. En particulier, les microtubules sont approximativement organisés de façon radiaire, depuis le centre cellulaire (depuis un organite, le centrosome, formé de tubuline) vers la périphérie (membrane). L'actine filamenteuse, elle, est souvent organisée en forme de cage, ou panier, sur tout le pourtour de la cellule.

Alignés comme ils le sont, en faisceaux, ces ensembles de fibres et de tubes peuvent constituer des guides de diffusion : la diffusion dans les axes orthogonaux aux fibres est ralentie de fait par la présence des fibres, mais celle qui se fait dans l'axe des fibres ne l'est pas. Ainsi, des molécules « piégées » dans de tels faisceaux (et qui plus est à l'intérieur des tubes des microtubules) sont guidées le long de leur axe, délivrant des molécules dans des régions préférentielles de la cellule, là où les faisceaux les mènent, de façon passive.

Des expériences numériques préliminaires que j'avais faites (section 7.1.7), simulant quelques dizaines de microtubules et la diffusion de quelques centaines de milliers d'agents, les molécules de tubuline, selon un modèle intégrant la forme géométrique des molécules (tubuline et microtubules), leur diffusions translationnelle et rotationnelle (temps de corrélation) sur 3 axes, les collisions élastiques entre ces particules, semblaient aller dans ce sens. Ces simulations étaient limitées par la lourdeur des calculs. Je projette de reprendre ces calculs et de moderniser leur implémentation (calcul GPU), pour explorer cette possibilité d'un transport par diffusion biaisée (anisotrope).

⁴⁶ Le Dalton est une unité utilisée en biologie cellulaire et moléculaire. Elle est telle que 1 Dalton (1 Da) = 1 g/mol. On utilise souvent des kDa (pour kiloDaltons).

Ceci peut également être vérifié expérimentalement. On sait que les microtubules peuvent être alignés de façon très efficace en les faisant croître sous champs magnétiques intenses ou sous champs électriques [Vassilev, 1982 ; Glade, 2005 & 2006](et section 7.1.2). Dès lors, il est tout à fait envisageable de réaliser des mesures de diffusion (par photo-bleaching) d'un traceur fluorescent comme la rhodamine injectée en haut d'un tube contenant une solution de microtubules préalablement assemblés soit sous champ magnétique intense (alignés dans l'axe du tube à essai), soit hors champ (et dans ce cas ils s'auto-organisent à cause de l'interaction avec la gravité, donc sont alignés localement et organisés en super-structure [Tabony, 2002b]), soit sous clinostat pour mimer des conditions d'apesanteur et empêcher l'auto-organisation et tout alignement global de se faire [Glade, 2001 ; Tabony, 2001 & 2002a]. De plus, en simulation comme sous champs magnétiques intenses, il est possible de contrôler le degré d'alignement des microtubules et en conséquence le biais de diffusion qu'ils provoquent.

8.3.6.1.c. Propagations d'informations structurales – chenaux de faible entropie

Si l'on reste sur le cytosquelette, mentionnons également une possibilité de communications d'ondes solitoniques et de trains ioniques se propageant respectivement dans les microtubules et à la surface des fibres d'actine. Ces questions initiées par l'anesthésiologiste S. Hameroff et le physicien et philosophe des sciences (prix Nobel de physique 2020) R. Penrose sont sujet à débat et peu acceptées par la communauté du cytosquelette. Quelques personnes cependant poursuivent et développent ces idées de propagation rapide de signaux le long de ces fibres, des ondes solitoniques (ondes solitaires qui se propagent sans se déformer ~ sans dissipation) dues à des changements de conformation de la tubuline se propageant de proche en proche, ou des déplacements d'ions à la surface de ces fibres. Hameroff et Penrose, comme ceux qui les succèdent dans ces travaux, s'intéressent non seulement à la propagation d'un signal le long des fibres, au passage de signaux de fibres en fibres, mais aussi aux collisions de signaux à la surface même des microtubules (ou à travers le réseau de fibres ou de tubes), les amenant même à les proposer comme substrat d'une « pré-computation », les surfaces microtubulaires servant alors d'intégrateurs de signaux [Hameroff, 1982, 1996, 1998 ; Tuszynski, 1995, 2004; Hagan, 2002 ; Georgiev, 2004, 2007]. Hameroff proposait que les microtubules qui constituent 10 % des protéines neuronales et qui sont alignés le long des axones, puissent pré-intégrer le signal se propageant dans les axones, de telle manière que les microtubules se comportent comme de petits ordinateurs pré-calculant les signaux dans le cerveau, avant même que les interactions neuronales elle-mêmes fassent leur œuvre et produisent l'activité cérébrale que l'on connaît (... assez mal). C'est ce que dans sa théorie Orch-Or (*Orchestrated Objective Reduction*) de la conscience, Hameroff, appelait *pre-consciousness*. J'aime beaucoup ces idées, très poussées dans les réflexions sur les relations entre très petites échelles (moléculaires, voire atomiques) et les grandes échelles (macro et mésoscopiques). Il y a d'ailleurs certainement des choses intéressantes à exploiter et à tester expérimentalement à ce propos et j'ajouterais que, même si de tels processus ne s'avéraient pas fondés, il demeurent de bonnes sources d'inspiration, en particulier dans le cadre du calcul non-conventionnel. Ainsi, développer de telles idées et tenter de

les pousser jusqu'à la réalisation expérimentale reste justifiable scientifiquement.

Ce modèle de propagation de signaux par changements de configurations de proche en proche le long des fibres du cytosquelette m'a inspiré un autre mode de propagation possible : une propagation de niveaux entropiques, de niveaux de structuration spatiale. J'aimerais ainsi profiter de cette séquence libre pour livrer une dernière idée qui m'intéresse, bien qu'à considérer avec prudence. Je prends le risque de me faire remonter les bretelles sur ce point, mais si vous êtes parvenus à ce stade du document, vous accepterez plus facilement de réfléchir à une idée osée, vous en conviendrez !

Elle m'est venue il y a une dizaine d'années, un matin entre 5 et 6 h dans le RER joignant Paris à Evry et j'ai le souvenir amusant de l'avoir exposée à P. Ballet, enquiquinant dans la foulée les passagers endormis qui s'en sont plaints ; voilà pour l'anecdote qui continue de nous faire rire, P. Ballet et moi quand on se souvient de ce bon moment. Pour être précis, elle m'a été inspirée initialement par deux ensembles d'articles. Le premier, de P. Davies (l'astrophysicien qui s'intéresse beaucoup à la biologie) concerne l'application du principe d'incertitude d'Heisenberg⁴⁷ à la vitesse de repliement des protéines [Davies, 2004]. Cet article, dont je reparlerai, montre en se référant notamment aux travaux de Wigner sur les horloges quantiques [Pesic, 1993] que plusieurs systèmes biologiques opèrent au seuil de la stabilité quantique ; c'est-à-dire que les temps caractéristiques en biologie (ex : vitesses de repliement des protéines, vitesses des réactions et interactions moléculaires, *etc*) sont proches des limites fixées pour ces systèmes (étant données leur masse et dimensions) par l'équation d'incertitude d'Heisenberg. L'article de Davies s'attelle particulièrement à questionner le rôle possible des aspects quantiques en biologie et traite notamment de l'indétermination dans le repliement des protéines : il n'est pas possible de connaître simultanément la structure (l'équivalent de la position dans le principe d'incertitude) d'une protéine et son temps caractéristique de repliement (l'équivalent de la vitesse dans le principe d'incertitude) en une structure native. L'autre article concerne des observations sur les microtubules (encore eux !), leur surface notamment, habillée d'une protéine intrinsèquement désordonnée, la protéine *tau* qui joue plusieurs rôles majeurs comme celui de stabiliser les microtubules (réduire leur instabilité dynamique), mais aussi faciliter et stabiliser la formation de faisceaux de microtubules [Santarella, 2004 ; Melo, 2016 ; Hernandez-Vega, 2017 ; Barbier, 2019]. Les microtubules, dont les longues extrémités C-terminales seraient intrinsèquement désordonnées et ainsi recouvertes de protéines tau elles-mêmes intrinsèquement désordonnées, loin de présenter l'image bien propre d'hétérodimères protéiques de tubuline alpha et bêta rangées sous la forme d'un tube élégant, auraient davantage une structure globalement organisée en tube, mais dont la surface présente une forêt de chaînes protéiques libres, désordonnées [Roll-Mecak, 2015].

Lorsque nous nous imaginons baignant dans le cytoplasme, nous peuplons un espace d'objets entre deux échelles (nano et microscopiques) structurés, constamment agités, se rentrant les uns dans les autres et créant parfois une interaction produisant un effet. Il y a trois choses que l'esprit humain a

47 Le principe d'incertitude (ou plus exactement d'*indétermination*) qu'Heisenberg a formulé en 1927 est un principe fort de la mécanique quantique. Il stipule que la vitesse d'une particule d'une part et sa quantité de mouvement d'autre part ne peuvent en aucun cas être déterminés avec précision simultanément. Ce principe est fondamental et a été largement vérifié.

du mal à réaliser concernant ce milieu biologique : i) le niveau de structuration des molécules biologiques dans le cytoplasme, niveau dont nous allons tout de suite discuter, ii) l'omniprésence des toutes petites molécules, c'est-à-dire le solvant (l'eau) et les ions, et le fait que cet ensemble se comporte comme un continuum [voir Annexe 1.3 sur l'eau] et iii) les temps caractéristiques des réactions moléculaires pouvant être très courts avec des vitesses comprises entre $2 \cdot 10^{-13}$ s pour la plus rapide des réactions biochimiques connues (réaction de la photo-rhodopsine à un photon) à 10^{-6} à 10^{-3} s pour le repliement des polymères biologiques (ADN, ARN, protéines) par exemple ou pour de nombreuses réactions enzymatiques [Mitić, 2017 ; Yeh, 2002] !

C'est ce milieu qui sert de support à cette idée. Celle-ci repose donc sur une réflexion sur la nature structurée ou non-structurée des éléments biologiques, notamment les protéines.

Parce qu'il est peu concevable que des objets non structurés (je parle là des protéines en particulier) puissent avoir une fonction (biologique donc), parce que la cristallographie renseigne sur la structure des objets biologiques (et non sur leur *non-structure*), parce que les modèles d'interaction biologique (ex : le modèle clef-serrure) requièrent que les objets en interaction aient une forme tridimensionnelle bien définie et que les structures des objets en interaction soient complémentaires, nous soutenons logiquement la nécessité de la structure (primaire à tertiaire, éventuellement quaternaire) pour que la fonction puisse se faire. Il est hors de question de remettre en cause ce qui est ainsi établi. D'ailleurs, une structure mal formée, même de façon infime parfois, amène à un fonctionnement différent ou à l'absence de fonction.

Mais chacun de nous se préfigure aussi que la structure existe au préalable du mécanisme qui l'implique dans une réaction biologique. Comment cela ne pourrait-il être le cas ? En réalité, rien n'interdit que des objets non structurés ou peu structurés ou partiellement structurés, susceptibles d'interagir (ex : une enzyme et un ligand), se mettent à se structurer dynamiquement alors qu'ils sont en train d'entrer en contact pour aboutir au contact en deux formes capables d'effectivement interagir. Cela semble entrer en contradiction avec les mesures de structures faites par cristallographie mais je ne crois pas que ce soit le cas. En effet, quand on forme un cristal de protéines pour ensuite le placer dans un faisceau énergétique de rayons X pour obtenir une figure de diffraction à partir de laquelle on peut calculer la position respective des atomes et reconstruire la structure de la protéine, on concentre très fortement ces protéines. Dans ce contexte il n'est pas vraiment envisageable que ces chaînes d'acides aminés puissent disposer d'un quelconque degré de liberté ; elles sont contraintes d'adopter leur forme dite *native*, une conformation qui leur permet d'être actives.

Pour soutenir la suite de mon propos sur ce que l'on désigne par l'appellation *natively unfolded proteins*, plus précisément les *intrinsically disordered proteins* (IDP) ou les *intrinsically disordered regions* (IDR), je m'appuie sur l'excellent cours de biochimie, enzymologie et bioinformatique en ligne de l'université d'Anger [BiochimieAnger] (que je recommande aux étudiants) et sur quelques articles [Gast, 1995 ; Weinreb, 1996 ; Uversky, 2009 ; Melo, 2016 ; Hernandez-Vega, 2017 ; Kulkarni, 2020] qui ont avancé cette possibilité (récemment envisagée et montrée) de protéines fonctionnelles intrinsèquement désordonnées et ont posé la question de la généralisation de formes non-natives (désordonnées ou du moins partiellement désordonnées) pour les protéines *in vivo*.

Même le fameux hub cellulaire, p53, n'y échappe pas ; comme ses consœurs IDP, sa structure tertiaire, voire secondaire, n'est que partiellement définie et elle est ainsi capable d'adopter une multitude de conformations et, comme pour les autres IDP, si certains conformères sont plus probables que d'autres, le passage de l'un à l'autre est aisé, facilité par un manque de cohésion interne lié d'une part à la faible quantité de liaisons internes (ex : ponts disulfure entre les cystéines) et par l'absence ou la quasi-absence d'acides aminés apolaires regroupés habituellement en poche(s) hydrophobe(s) servant souvent de site actif. Ces protéines IDP sont nombreuses, mais encore plus nombreuses sont les protéines possédant des IDR : on estime qu'environ 50 à 70 % des protéines possèdent des IDR d'au moins 30 résidus d'acides aminés.

Dès lors, il est difficile de continuer à visualiser l'intérieur cellulaire comme une pièce d'horlogerie aux engrenages finement usinés et s'agencant parfaitement. A la place, on conçoit le fouillis peuplé de chaînes mobiles, certaines plus ramassées que d'autres, s'entrechoquant et se structurant partiellement ou complètement, dynamiquement en tout cas, puis perdant à nouveau la structure ponctuellement acquise, au fil des contacts et de séparations. En prenant un peu de recul, il persiste peu d'ordre dans tout cela, cet ordre que nous observons sous nos microscopes et qui baigne dans un désordre presque omniprésent, un milieu désordonné capable de passer très localement et très rapidement à un niveau d'ordonnement générant la fonction, puis de s'évanouir aussi vite qu'il est apparu. Dit autrement, une sorte de continuum entropique règne dans lequel des points, des petites régions, parfois des zones plus importantes de faible entropie (globalement plus structurées) émergent, disparaissent, réapparaissent, parfois persistent à cause d'éléments structurants (membranes, fibres, protéines très structurées ...). Si vous saisissez ce paysage⁴⁸, vous y verrez alors ces endroits éphémères dans lesquels se produit telle ou telle réaction biochimique comme une réaction enzymatique, certaines régions structurées paraissant ramifiées et assez persistantes car groupées en une même zone de la cellule, une zone particulièrement active comme par exemple une procession de ribosomes sur un ARN messenger au cours d'une synthèse protéique à la surface du réticulum endoplasmique ou une cascade de signalisation initiée par des protéines G sous-membranaires accolées à la partie intracellulaire de récepteurs membranaires venant de changer de conformation à la perception d'un signal moléculaire extracellulaire, certains se propageant de proche en proche au fil des réactions successives d'une même chaîne réactionnelle ou se manifestent parce qu'une augmentation locale de la densité moléculaire force les éléments en présence à se replier en conformations plus actives.

Nous parlions de réseaux ? En voici donc une forme, tout aussi amorphe que ceux dont j'ai parlé précédemment, auquel je propose de réfléchir, des réseaux que l'on pourrait désigner par « chenaux de communication entropiques » ou « pré-chemins entropiques ». Considérons par exemple l'ensemble des réactions successives qui d'un signal moléculaire externe à la cellule provoque à terme (et dans un délai très court finalement) une activité de régulation génétique ou de synthèse protéique, signal qui commence par l'interaction d'un ligand avec la partie extra-cellulaire d'un récepteur transmembranaire, continue par la transduction par sa partie intra-membranaire via des protéines G ou des messagers intermédiaires comme la calmoduline ou l'inositol triphosphate,

48 En ce qui me concerne, je visualise ce milieu entropique comme un paysage gris, dans lequel apparaissent des points ou des régions de faible entropie colorés en niveaux de rose (un peu structuré) jusqu'au rouge vif (très structuré).

jusqu'au cibles nucléaires, le tout impliquant de nombreuses protéines kinases. Dans le contexte de structuration dynamique locale proposé ci-dessus, cette cascade qui modifie le flux global du métabolisme, pourrait possiblement se manifester également comme une cascade de structurations de proche en proche, une cascade pouvant se manifester comme un chenal ou un ensemble de chenaux de faible entropie reliant la zone de réception du signal extra-cellulaire à un certain nombre de région affectées par la cascade qui en découle. Il est même envisageable que préalablement à la propagation réactive du signal, un signal de structuration mutuelle se propage légèrement en amont, plus rapidement que la succession des actions effectrices elles-mêmes. Mieux, un premier signal et sa cascade associée, une fois le milieu cellulaire traversé, pourrait bien laisser une trace de son passage, toujours sous la forme d'une sorte de chenal de faible entropie, et ce d'autant plus que ce premier signal est intense. Le terrain ainsi préparé, c'est-à-dire ses composants pré-structurés ou mieux structurés, faciliterait alors la propagation de futurs signaux. A l'échelle écosystémique, cela s'apparente fort à un terrain gardant trace d'une activité biologique précédente, tels les trails systems.

8.3.6.1.d. La part du faible du rare et de l'amorphe dans les communications

Tout cela reste léger car encore trop peu documenté, un peu spéculatif ou audacieux, c'est selon, mais l'ensemble des propositions faites ici est testable par simulation numérique et les dynamiques qui en émergent peut être confrontées avec des mesures qui pourraient être faites *in vivo*. Il y a relativement peu à redire concernant l'existence des mécanismes microscopiques décrits comme le repliement protéique, les IDP et IDR, les réactions précitées, mais il est évident que les éventuels processus émergents ici évoqués sont sujets à débat. Néanmoins, cela rejoint les réflexions de Michael Conrad sur le nombre d'interactions physiques accessibles et exploitable à et par la cellule vivante (en comparaison des systèmes électroniques) : M. Conrad voit les processus cellulaires comme des « problèmes » à traiter [Conrad, 1995]. Pour un processus opérant à une échelle donnée (par exemple une synthèse protéique impliquant ARNm, ARNt et ribosomes) il y a des interactions d'échelle immédiatement inférieure (impliquant des changements de conformation de ces particules, donc des domaines moléculaires composés parfois de nombreux acides aminés et/ou nucléotides) qui opèrent, mais on peut se poser la question du rôle individuel des interactions d'échelle encore inférieure (ex : au niveau des résidus d'acides aminés) et encore de celles d'encore en dessous (liaisons atomiques, charges ...) et ainsi de suite, dans la fonction biologique considérée (ici une synthèse protéique). On pourrait discuter de la quantité d'information apportée ou prise en charge par ces interactions d'échelles de plus en plus petites. On remarquera que c'est aussi une question que je pose dans la section 8.3.5.3 lorsque je parle de la description fine de fonctions d'un réseau d'interaction formel exprimée dans un certain formalisme qui peut être développé en un formalisme de plus bas niveau. Dans cette descente dans les échelles de plus en plus fines, on s'apercevra que certaines informations n'ont finalement que peu d'influence tandis que d'autres ont un rôle majeur ; *du point de vue du scientifique* qui cherche à identifier les contours d'un mécanisme et à lui donner un nom, les premières seront ignorées mais les autres devront être prises en compte. En fin de

compte, c'est la mesure de leur participation à un processus (une fonction biologique) et le seuil de participation que ce scientifique aura choisi qui déterminera leur légitimité à appartenir à une fonction.

Pour appuyer davantage ce propos, je vais reprendre telles quelles les propositions que P. Davies a énoncées dans son article sur le possible rôle de la mécanique quantique en biologie [Davies, 2004] et je demanderais au lecteur de bien vouloir remplacer les termes soulignés faisant référence à la mécanique quantique, par des termes comme « réseaux amorphes », « trail systems », « communications entropiques », *etc.*

- *Quantum mechanics played a significant role in the emergence of life from non-living chemical systems in the first place, but ceased to be a significant factor when life got going.*
- *Quantum information processing may have played a key role in the emergence of life, and a sporadic or subsidiary role in its subsequent development. There may be relics of ancient quantum information processing systems in extant organisms, just as there are biochemical remnants that give clues about ancient biological, or even pre-biological, processes.*
- *Life started out as a classical complex system, but later evolved some form of quantum behavior as a refinement. For example, if biological systems can process information quantum mechanically, they would gain a distinct advantage in speed and power, so it might be expected that natural selection would discover and amplify such capabilities, if they are possible. This is an extension of the dictum that whatever technology human invent, nature normally gets there first.*

Comme l'argumente là P. Davies, s'il est indéniable que de tels mécanismes existent (la mécanique quantique, mais aussi les *trail systems*, les niveaux de conformation moléculaires variables et changeant dynamiquement, *etc*) la question est de savoir s'ils ont joué un rôle (ce qui fut éventuellement possible aux origines du vivant quand peu de processus complexes existaient et surtout alors que les molécules prébiotiques, protobiotiques et biologiques ne devaient pas encore présenter de fortes affinités), s'il en persiste des reliques actuellement qui peuvent éventuellement jouer encore parfois un rôle dans les système actuels, par exemple la diffusion anisotrope dans les cellules le long des fibres du cytosquelette, ou s'ils jouent effectivement un rôle bien présent, étant même éventuellement apparus tardivement. En fin de compte, c'est la sélection naturelle qui sert de filtre et qui permet ou non le maintien de tels processus.

Cette question de la prise en compte des échelles successivement inférieures, sous-entendu des interactions *a priori* non ou peu spécifiques, faibles en tout cas, pour décrire des mécanismes est une question récurrente en sciences. C'est le cas en biologie, mais aussi en physique ou en neurophysiologie et psychologie comme en parle si bien D. Hofstadter dans son livre *Je suis une boucle étrange* [Hofstadter, 2013] lorsqu'il parle de glissement dans les niveaux de description. Que voudrait dire une description à l'échelle moléculaire, voire subatomique de la chute des dominos ou d'un lancé de dés ? En réalité, cela dépend déjà de l'accès que l'on a à ces petites échelles dans les systèmes macro et mésoscopiques considérés, mais aussi de la nécessité (sociale, scientifique, morale, *etc*) de changer une description fonctionnelle par une autre qui prendrait en compte des

échelles plus petites, des mécanismes moins évidents, moins spécifiques, moins forts, plus distribués, *etc.* C'est cette nécessité qu'Hofstadter expose dans ce paragraphe : « *Nous sommes des animaux dont la perception est limitée au monde des objets macroscopiques de tous les jours. Cela nous contraint de manière presque évidente à fonctionner sans aucune référence aux entités et processus des niveaux microscopiques. Personne n'avait en réalité la moindre idée de ce qu'est un atome jusqu'il y a seulement une centaine d'années, et personne ne s'en portait plus mal. Ferdinand Magellan a fait le tour du monde, William Shakespeare a écrit des pièces de théâtre, J.-S. Bach a composé des cantates, Jeanne d'Arc s'est fait brûler sur le bûcher, chacun pour de bonnes (ou mauvaises) raisons qui n'avaient pas le moindre rapport, de leur point de vue, avec l'ADN, l'ARN, les protéines, le carbone, l'oxygène, l'hydrogène ou l'azote ; ni avec les photons, les électrons, les protons ou les neutrons, sans parler des quarks, des gluons, des bosons W et Z, des gravitons ou des particules de Higgs.* » et dans le paragraphe qui suit « *Ce n'est une révélation pour personne que des niveaux de description différents répondent à des besoins différents selon le sujet et le contexte.* ».

J'ai exposé un certain nombre de besoins (ex : les réflexions sur les origines du vivant et l'émergence d'une génétique) mais aussi de suggestions de mécanismes et processus justifiant qu'on s'intéresse à la part de ce que j'ai appelé le rare, le faible et l'amorphe. Les modèles, du moins les modèles conceptuels, les *toy models*, permettent de considérer et d'explorer des relations d'échelles et leurs dépendances, les interactions faibles, *etc.*, là où la mesure pêche. Parfois la mesure, l'observation suggèrent, parfois c'est l'analogie entre objets d'études, ... ; c'est en tout cas un dialogue entre disciplines, entre le modélisateur et l'expérimentateur (parfois la même personne), qui autorisent en pratique la prise en compte de ces considérations.

8.3.6.2. *De la machine au promeneur dans les paysages morphogénétiques – questions de viabilité*

Revenons aux réseaux ; ceux bien formels, intelligibles, largement présentés plus haut. On peut les étudier comme on démonte une machine pour voir comment elle est faite ou d'essayer d'en tirer une description globale mais, lorsqu'on les envisage comme modèles de la machinerie du vivant ou comme modèles d'interactions entre organismes, il me semble que l'on passe à côté de quelque chose si l'on n'inclut pas systématiquement la contrainte de la ou des dynamiques qu'ils doivent assurer, et même d'y inclure la notion de noyau de viabilité, c'est-à-dire l'ensemble des états (transitoires ou attracteurs) compatibles avec la vie de l'organisme dans un environnement donné. Ce noyau de viabilité dont je parle ici est une adaptation aux systèmes discrets de la théorie de la viabilité développée par J.-P. Aubin [Aubin, 2013]⁴⁹ pour les systèmes dynamiques continus. Notre

49 S. Ivanov et moi sommes allés discuter avec lui à ce propos en 2018, dans un troquet parisien, son lieu de travail privilégié, pour voir s'il était raisonnable de considérer et d'appliquer la théorie de la viabilité dans le cadre des systèmes discrets, les réseaux Booléens en l'occurrence. De ce très agréable moment, nous sommes ressortis convaincus qu'il fallait absolument coupler tout cela, que ça avait beaucoup de sens et que ça permettrait d'utiliser nombre de connaissances (concepts, théorèmes ...) issus de cette théorie à nos travaux. Nous gardons ça « sous le coude ».

objet de travail devient donc un triplet formé par le réseau lui-même (son graphe d'interaction), sa dynamique (son graphe de transition) et son espace de viabilité, une sous-partie de la dynamique imposée par l'environnement, c'est-à-dire la partie qui rend le réseau à la fois plus vivant et dépendant de l'extérieur (d'autres systèmes biochimiques/biologiques, de conditions environnementales, du reste de l'écosystème).

De plus, comme nous l'avons déjà dit, il y a au moins trois bonnes raisons de travailler plutôt avec des ensembles de réseaux plutôt qu'avec un seul :

- la connaissance biologique est incomplète ; toujours ! S'il est vrai qu'un réseau puisse satisfaire l'ensemble des connaissances dont on dispose sur un réseau d'interaction et son comportement, et que celui-ci puisse être considéré comme un bon modèle du système biologique étudié, il est tout aussi vrai de dire que d'autres réseaux (parfois de même dimension, parfois, de dimensions plus grandes, parfois plus détaillés dans leurs mécanismes internes, *etc*) puissent aussi correspondre à ce que l'on modélise. Dès lors, il vaut mieux, bien que cela s'avère plus compliqué à utiliser dans des simulations, considérer les ensembles de réseaux satisfaisables pour les connaissances de structure et de dynamique dont on dispose.
- les éléments d'un système biologique ont des interactions multiples (ex : les miRs qui ont des centaines ou des milliers de cibles, p53 le hub moléculaire capable d'interagir avec des centaines de cibles différentes, *etc*), plus ou moins spécifiques, sont changeants, soumis à la fluctuation (ex : changements de structuration de l'ADN, changements de voisinage, de densité en facteurs de transcription ...). Un seul réseau n'est pas forcément un bon modèle d'un tel système d'interaction biologique, il n'en est que le comportement typique, celui qui ressort principalement. L'activité moyenne d'un ensemble de réseaux (par exemple) est en fait peut-être plus représentative du système biologique [Ben Amor, 2013 ; Paulevé, 2021].
- étudier un réseau particulier, que l'on conçoit comme un modèle d'un système biologique donné, c'est l'étudier en détails en mécanisant une perception que l'on a du vivant. Cela permet éventuellement d'aider à comprendre certains aspects de son fonctionnement, mais ça ne permet pas de tirer des enseignements plus généraux sur la façon dont le vivant fonctionne. Une façon de commencer à se libérer du système biologique-machine, c'est justement de considérer le système biologique comme une multitude de modèles tous exprimables à la fois, son activité résultant de leur pondération.

Nous allons donc prendre un peu de recul et considérer maintenant un espace peuplé de réseaux, un espace dans lequel ils sont différenciables par leur structure, leur composition en fonctions, leur complexité, leur dimension, *etc*. Dans cet espace, nous allons identifier ceux qui réalisent totalement ou partiellement une certaine dynamique, ceux qui ont certaines propriétés dynamiques (comme certaines tailles de bassins d'attraction, certaines complexités de comportement, ...), des dynamiques ou comportements qui pourraient permettre à des organismes de vivre dans un environnement donné. C'est ce qu'illustre la figure 8.58 et c'est ce que je désignerai par la suite

d'abord par l'expression « paysage morphogénétique », expression qui rappelle immanquablement le paysage épigénétique dans lequel le biologiste Waddington identifie des *chréodes* (pour *chemin nécessaire*) et qui décrit l'idée que les gènes influencent la morphogenèse en canalisant les cellules dans des voies de différenciation successives [Waddington, 1957 ; Verd, 2014], puis, dans la section suivante, par « mille-feuille évolutif ». Cette image est avant tout une représentation (mentale) que nous pouvons utiliser pour réfléchir sur les réseaux comme modèles du fonctionnement cellulaire et à des questions tournant autour de la robustesse, de la résilience et, plus loin, de l'évolution. Mais plus que cela, ils sont calculables ... du moins en partie. Je vais ici présenter cette idée, comment la concrétiser et ce que l'on peut en faire.

La figure 8.58 montre deux représentations de ce que pourraient être ces paysages. La première montre un ensemble de réseaux mis en relation, de proche en proche, par une distance structurelle [Segretain, 2018 ; Christen, 2018]. On part d'un réseau donné (central ici) réalisant une dynamique choisie, et on détermine de proche en proche les réseaux semblables. Parmi eux, on identifie ceux qui réalisent cette dynamique (en vert). Dans l'espace de tous les réseaux de dimension 3, ces réseaux réalisant une dynamique particulière définissent des îlots à l'intérieur desquels les réseaux sont connexes : ils constituent un réseau. De tels îlots sont représentés en rose, disséminés dans l'espace de tous les réseaux possibles (limités à une certaine dimension éventuellement) dans la figure (B). On y ajoute des aires grises qui représentent, elles, l'espace de viabilité imposé par l'environnement du système fonctionnel représenté par les réseaux. Pour un ensemble de réseaux compatibles avec la viabilité de l'organisme, c'est-à-dire possédant une dynamique satisfaisant la viabilité (aires grises), des groupes de réseaux se ressemblent, satisfont les contraintes de structure d'un système biologique donné et réalisent au moins une musique donnée (aires roses). La conjonction des aires grises et roses correspond à l'ensemble des réseaux pouvant modéliser un système biologique donné et compatibles avec la viabilité de l'organisme qui en dépend. Ce sont des noyaux de viabilité. Ce sont de tels îlots qui peuvent constituer des systèmes contigus modélisant un même système biologique.

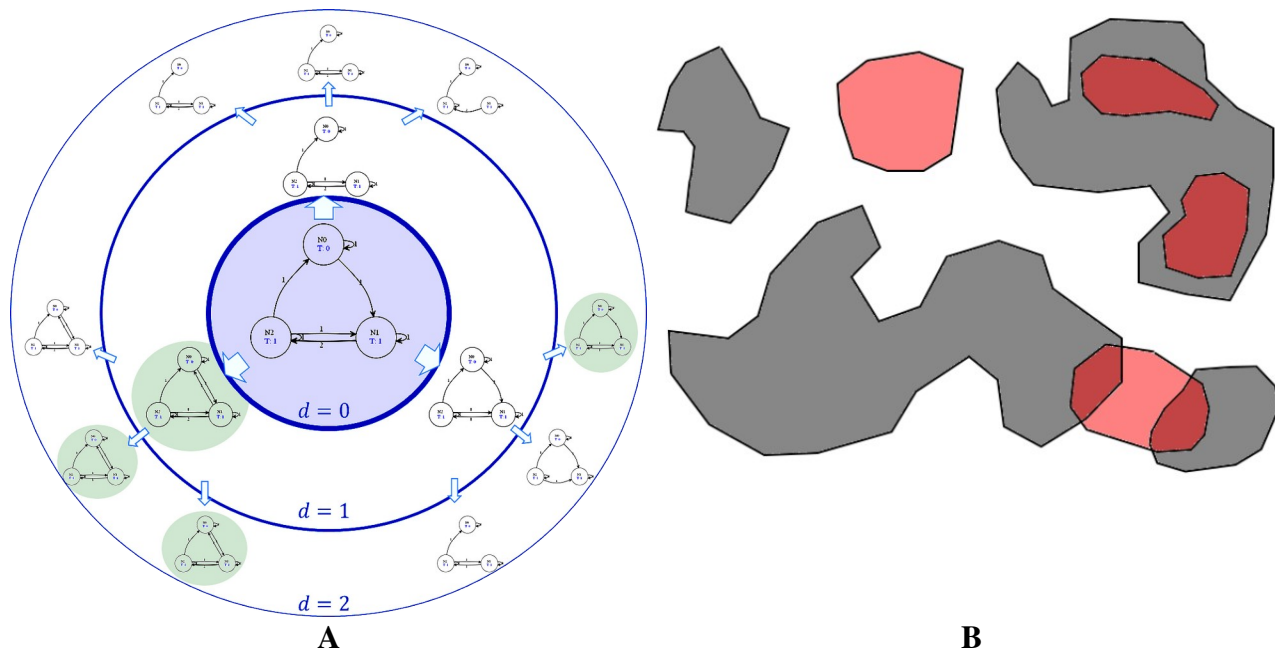


Figure 8.58. Espaces de réseaux. (A) cette figure montre un ensemble de réseaux mis en relation, de proche en proche, par une distance structurelle [Segretain, 2018 ; Christen, 2018]. On part d'un réseau donné (central ici) réalisant une dynamique choisie, et on détermine de proche en proche les réseaux semblables. Parmi eux, on identifie ceux qui réalisent cette dynamique (en vert). Dans l'espace de tous les réseaux de dimension 3, ces réseaux réalisant une dynamique particulière définissent des îlots à l'intérieur desquels les réseaux sont connexes : ils constituent un réseau. De tels îlots sont représentés en rose dans la figure (B). On y ajoute des aires grises qui représentent, elles, l'espace de viabilité imposé par l'environnement du système fonctionnel représenté par les réseaux.

De ce point de vue, une instance de réseau (un réseau particulier) représente une réalisation à un moment donné de la vie cellulaire (un instantané). Le passage d'une instance à l'autre se fait au gré des variations ponctuelles de structure ou de la dynamique. Cela sous entend que, d'une part, le système, malgré cette variation, conserve la dynamique souhaitée (le même attracteur ou une même musique au moins si cela suffit), d'autre part que l'état dans lequel se trouve le système au moment de la variation lui permet de rester dans un bassin d'attraction auquel appartient cette dynamique. C'est ce qu'illustre la figure 8.59. On peut espérer qu'une variation d'un réseau (ou d'un de ses états) puisse bénéficier de la robustesse amenée par un regroupement en îlots contigus de réseaux ayant au moins la dynamique souhaitée. Certains réseaux se trouvent en bordure de tels espaces, d'autres plus au centre. Ceux qui sont au centre sont *a priori* plus fiables que les autres en bordure : une altération ponctuelle ou durable de leur structure ou de leur dynamique a toutes les chances de rester dans le noyau de viabilité.

Nous devons vérifier par le calcul qu'il existe bien de tels noyaux de viabilité, des îlots peuplés de réseaux, éventuellement reliés par des chemins dans l'espace des réseaux. Si c'est le cas, on sera en mesure de se poser sérieusement deux questions supplémentaires :

- Qu'en est-il du vivant *réellement* ? Peut-on trouver des indices d'une telle forme de robustesse ?
- Dans ce cas, les systèmes naturels ont-ils tendance à se « positionner » dans ce qui est l'équivalent de réseaux centraux à ces noyaux de viabilité ? Le positionnement d'un système

naturel dans une configuration plus robuste pour une fonction biologique donnée implique un rôle joué par l'évolution, une des deux forces possibles (avec la pression exercée par l'environnement) capables de peu à peu positionner, par sélections favorisant naturellement la robustesse des fonctions biologiques, le système biologique dans une configuration robuste, centrale dans un îlot de viabilité.

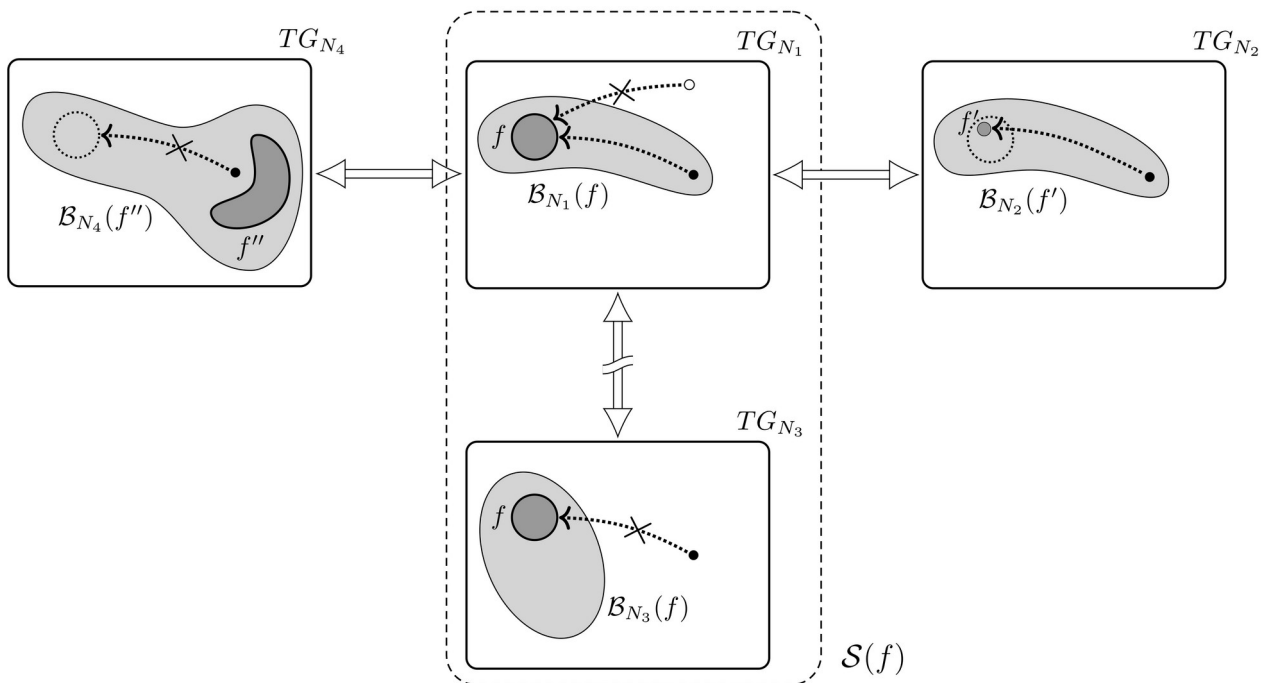


Figure 8.59. Îlots de robustesse : proximité des réseaux et de leur dynamique. Dans cette figure, chaque cadre rectangulaire en traits pleins représente l'ensemble des états (TG, le graphe de transition) d'une instance de réseau. Dans chaque cadre, l'aire gris clair représente un bassin d'attraction englobant un attracteur vu ici en gris foncé. Les points noirs et le point blanc représentent un état initial. Dans le graphe de transition du réseau N_1 , l'état initial marqué par le point noir, présent dans le bassin d'attraction, mène vers l'attracteur représentant f . En partant du point blanc en revanche, état initial externe au bassin d'attraction, on ne rejoint pas f . Cette figure illustre l'effet de diverses variations sur la dynamique des réseaux. Ici, on représente les dynamiques de 4 réseaux N_1 à N_4 . On considère une fonction biologique souhaitée f , à atteindre, représentée par les attracteurs des réseaux N_1 et N_3 . Ces réseaux sont dans l'ensemble viable satisfaisant la fonction f (set $S(f)$). Les autres réseaux N_2 et N_4 réalisent chacun une autre fonction, respectivement f' et f'' . On voit là ce qui peut se produire si une variation de structure du réseau N_1 se produit, induisant un changement de son graphe de transition. Si on fait $N_1 \rightarrow N_3$, il faut pour que rester dans f ou rejoindre f qu'on soit déjà dans l'attracteur de f ou dans un état du bassin du nouveau réseau (N_3) commun avec celui du réseau précédent (N_1). D'autres situations peuvent se présenter qui assurent une forme de robustesse : si on a une transition $N_1 \rightarrow N_2$ et qu'une transition reverse ($N_2 \rightarrow N_1$) se produit, alors comme ces deux réseaux ont le même bassin d'attraction (ou un bassin semblable au moins, par exemple celui de N_4) alors (il est probable que) la fonction f sera conservée. N_2 (et dans une moindre mesure N_4) doit donc se trouver proche de N_1 dans l'espace des réseaux car il assure une robustesse dans la viabilité de f .

Nous avons vu l'effet de variations ponctuelles sur la structure et la dynamique des réseaux. Dans la fin de ce manuscrit, lorsque vous serez arrivés à ce qui constituera mes réflexions pour l'avenir, j'aurai ajouté une dimension évolutive à ces paysages, une dimension qui fige un temps les variations et qui les teste dans un environnement imposant une contrainte de viabilité changeante, une dimension dans laquelle se poseront des problèmes d'atteignabilité.

9. DISCUSSIONS SUR L'ÉVOLUTION, LA VIE ...

Le chapitre qui vient est une ouverture. Voyez-le comme tel en tout cas. J'utilise ce qui a été dit précédemment et montre comment j'en vois l'usage pour aller plus loin dans mes réflexions sur l'évolution, ses reliques, l'atavisme et les origines du vivant, sur le faible le rare et l'amorphe. Ce sont des problèmes de recherche qui m'intéressent beaucoup, sur lesquelles j'ai commencé depuis quelques années parfois à réfléchir, à déjà explorer parfois, et ce qui pourraient bien préfigurer mon activité dans les années à venir. Il s'agit d'une succession de sections que j'ai eu un mal fou à ordonner tant elles sont inter-dépendantes. Je pense (j'espère) que cette forme-ci est compréhensible et restera agréable à lire.

9.1. Pluripotence, totipotence : réseaux composés et réseaux simulants

Au chapitre 8, nous avons vu que les SBNs et leur graphe de transition constituaient une bonne abstraction du génotype/phénotype ou plus généralement de la relation structure/fonction. Nous avons aussi disserté sur l'extensibilité (comment faire croître un réseau en taille, par ajout de nœuds, en conservant la structure du réseau initial et la fonction initiale associée dans la fonction évoluée) ; une forme de mécanisme évolutif en quelle que sorte. Je vais montrer ici une autre façon de concevoir l'augmentation de réseaux et des fonctionnalités associées : les réseaux pluripotents.

9.1.1. Simuler des réseaux avec des réseaux

L'idée générale est la suivante : supposons que l'on dispose de 2 réseaux déterministes (des SBNs par exemple) R_A et R_B , de dimensions respectives d_A et d_B , chacun avec son graphe de transition associé (son phénotype) GT_A et GT_B , chacun contenu dans un hypercube de dimension d_A pour R_A et d_B pour R_B , ces hypercubes contenant tous les états possibles que peuvent prendre les nœuds de ces réseaux (soit respectivement 2^{d_A} et 2^{d_B} états pour les réseaux R_A et R_B). Comment dès lors les combiner en un réseau R de façon à obtenir un graphe de transition qui contient les deux graphes de transition GT_A et GT_B indépendamment l'un de l'autre ? Notons d'emblée qu'un tel réseau R pourrait ne présenter aucune ressemblance ni avec R_A , ni avec R_B . Le réseau R possède les deux phénotypes qu'il peut potentiellement exprimer ; il est donc pluripotent, au sens où il contient plusieurs comportements (GT) indépendants, et simulant, au sens où il simule les graphes de R_A et R_B .

Il est bien entendu envisageable de considérer la totipotence pour une dimension donnée, c'est-à-dire l'obtention d'un réseau R de dimension d' capable de tous les phénotypes associés à l'ensemble des réseaux possibles d'une dimension d donnée, $d \ll d'$. Par exemple, il y a 21 graphes de transition (phénotypes) différents pour les SBNs de dimension 2, chacun de leurs graphes de transition nécessitant 4 états. Pour stocker 2 graphes de transition de 4 états chacun, il faut 8 états, donc deux faces d'un cube tridimensionnel (voir figure 9.1). A partir de maintenant, continuons d'appeler « graphes simulés » les graphes de transition de taille 2 plongés dans un graphe de transition de

dimension supérieure, et « faces » les hypercubes de dimension 2 contenant ces graphes simulés. Et tant qu'à faire, généralisons : appelons « hyperfaces » les hypercubes contenant les graphes de dimension d plongés dans un hypercube de dimension d' supérieure (et contenant donc un graphe de transition de dimension d' supérieure). NB : on distingue ici les hypercubes des graphes de transition eux-mêmes. Contrairement aux graphes de transition, les hypercubes n'indiquent en effet rien sur les transitions ; ils ne sont que l'ensemble de tous les états possibles que peut adopter un réseau de dimension d donnée.

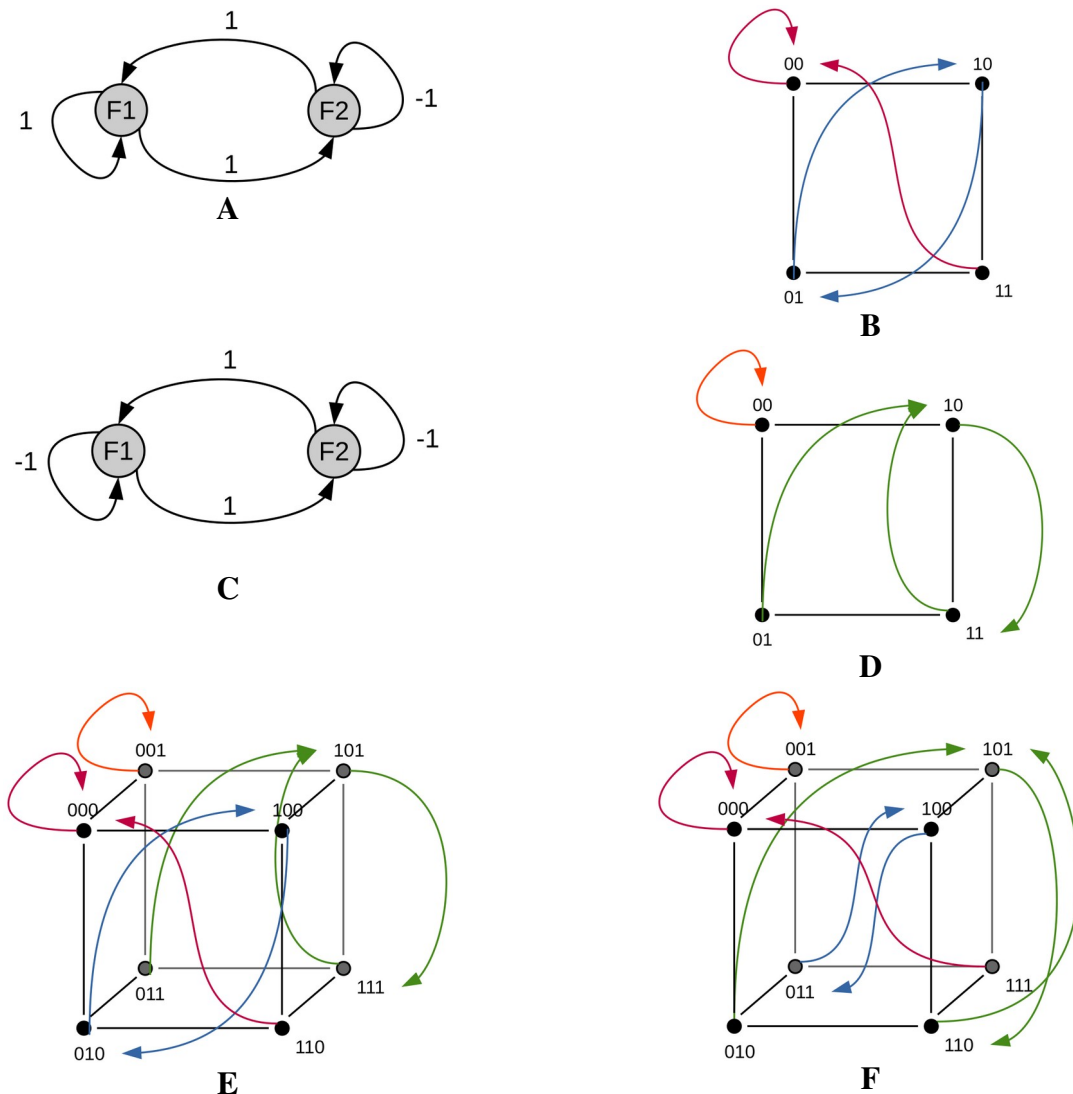


Figure 9.1. Illustration de la pluripotente. En A et C, deux SBNs R_1 et R_2 ayant respectivement pour graphes de transition les graphes B et D. Leurs bassins d'attraction sont de couleurs différentes. Est-il possible de trouver un réseau R de dimension 3 tel que les graphes de transition B et D puissent être stockés comme des faces (hyperface) du graphe de transition (cube) d'un tel réseau, comme montré en E (face avant et face arrière) ou plus 'curieusement' comme en F sur des faces inclinées d'avant en arrière ? Un tel réseau R simulerait alors sur des hyperfaces indépendantes les graphes de transition des réseaux R_1 et R_2 . Dans les deux cas (E et F), le contrôle s'effectue par le 3ème bit, donc l'état du 3ème nœud, cependant s'il est constant pour chaque hyperface simulée dans le cas E, il ne l'est pas pour le cas F ; dans ce cas, il faut changer régulièrement d'état pour parcourir les états successifs des bassins d'attraction de R_1 et R_2 . L'existence de R n'est cependant pas garantie ! NB ; dans le cas précis de cet exemple, il n'existe aucun réseau R de type SBN capable de simuler de tels graphes de transition (E

comme F). Il pourrait cependant en exister de type TBN et il en existe au moins un de type PBN.

Pour les 21 graphes de transition de dimension 2 possibles générés par les SBN de dimension 2, il faut donc 84 états stockés dans 21 hyperfaces indépendantes ; il faut donc disposer d'un graphe de transition de dimension 7, soit un hypercube de 128 états contenant 32 hyperfaces de dimension 2 indépendantes. Existe-t-il un tel réseau totipotent de dimension 7 ayant un tel graphe de transition ? La réponse est oui, avec des PBNs, étant donné qu'il est possible d'obtenir tous les graphes de transition d'une dimension donnée avec des PBNs de cette dimension. Elle est moins évidente avec les formalismes de réseaux Booléens à l'expressivité plus restreinte, les TBNs et *a fortiori* avec les SBNs.

Bien entendu, toujours sur le même principe, on peut considérer des réseaux capables de simuler l'activité de réseaux de dimensions différentes, ou tous les réseaux de toutes les dimensions inférieures ou égales à une dimension donnée. Et ainsi de suite, l'idée est déclinable à l'infini. Bien entendu aussi, tout cela, c'est seulement si les hyperfaces sont indépendantes ; on pourrait tout aussi bien chercher à éviter de dupliquer des comportements (attracteurs et leurs bassins d'attraction respectifs) en cherchant uniquement à faire croître les graphes de transition avec les comportements nouveaux uniquement. Mais restons sur le problème avec des hyperfaces indépendantes.

Telle quelle, cette idée n'est cependant pas suffisamment bien exprimée car nous n'avons encore rien dit concernant la structure de R , le *réseau simulant* ; elle peut en effet être comprise de plusieurs manières différentes. Une première manière de la formuler, la plus simple, est de ne pas spécifier la structure du réseau pluripotent R à obtenir. Une autre, à l'autre extrême, consiste à demander la conservation des réseaux R_i dans la structure du réseau R (en réalité, un tel réseau R consiste alors en une simple collection de réseaux R_i disjoints). Et bien entendu, on peut aussi considérer tout le gradient entre les deux propositions, c'est-à-dire avec un réseau R *simulant* intégrant au moins partiellement certaines parties des réseaux R_i *simulés*. Dans ce gradient, on doit pouvoir trouver des réseaux R qui sont une évolution de réseaux initiaux R_i englobant les capacités d'autres réseaux simulés R_i' .

Dans tous les cas, pour changer d'hyperface, il faut qu'un système de contrôle permette la configuration des états du réseau R l'amenant à simuler telle ou telle hyperface. Nous n'allons pas tout de suite nous concentrer sur le système de contrôle ; c'est un autre sujet que j'aborde plus loin. Néanmoins, nous allons poser la question du changement d'hyperface et des choix à faire pour l'atteindre.

Mais d'abord, à quoi sert *a priori* cette idée ? Voyons ça rapidement ; nous y reviendrons cependant plusieurs fois. D'abord, du point de vue de l'ingénierie, on pourrait imaginer fabriquer des systèmes adaptatifs capables de changer dynamiquement de comportement (graphe de transition) sous l'influence d'un contrôle externe. Là, on se pose inmanquablement la question de comment un tel système de contrôle peut permettre d'atteindre un comportement plutôt qu'un autre ; pour cela, il apparaît nécessaire de connaître par avance le contrôle à effectuer pour choisir le comportement voulu. Mais il y a peut-être un moyen de réduire cette connaissance requise ; je me réfère pour cela à

la section 8.3.6.2 qui propose de se balader dans les paysages morphogénétiques : il est envisageable qu'un certain encodage des variations du système de contrôle (donc des états à injecter dans le réseau simulé) puissent permettre de faire varier les comportements simulés en passant de proche en proche. Un petit changement dans le contrôle fait changer un comportement en un comportement proche ... proche pour le système qui l'exploite et viable pour ce système pour les conditions (internes et externes) dans lesquelles ils se trouve. On obtient ainsi un système rapidement adaptable avec une forme de robustesse : le système est moins *intelligent* (en regard de la quantité d'éléments mis en œuvre) mais bien plus robuste ; il peut s'adapter dynamiquement grâce à un système de contrôle dont l'encodage dépend des variations d'environnement que le système peut rencontrer.

Ensuite, et c'est une question ouverte évidemment : « et dans le vivant ? » Est-ce qu'il y a des systèmes ainsi modélisables sous forme de réseaux, qui permettent de simuler des classes entières de comportements pouvant être ceux de réseaux plus simples ? Les grands réseaux génétiques font déjà ça (ils sont capables de plusieurs comportements pour un même réseau ; c'est un truisme !), mais la question qui va avec est de savoir comment ces comportements (hyperfaces) ont été agrégés dans un ensemble de comportements (un graphe de transition), autrement dit comment on a pu passer de systèmes simples ayant chacun leur activité propre à de larges systèmes simulants de multiples systèmes et dans lesquels la trace des systèmes dont ils proviennent aura tendance à s'effacer avec le temps. Le reviendrai sur ces aspects (contrôle et systèmes biologiques) dans les sections suivantes.

Et dans le cerveau ? Ça aussi, c'est une question qui m'intéresse beaucoup. Son activité consciente est le fruit de la circulation d'états (davantage continus que Booléens, mais réfléchir avec l'abstraction des réseaux Booléens est utile et ne change probablement pas vraiment la nature du propos) passant dans les quelques cent milliards de neurones du cerveau (disons cent milliards exactement). Elle résulte d'influx circulant des organes des sens au corps cérébral, ré-entrant de régions du cerveau en d'autres régions, ou bouclant localement de neurone en neurone dans des noyaux ou certaines régions corticales (lire par exemple [Edelman, 2000] ou [Hofstadter, 2013]). Le cerveau a une configuration globale donnée, fixée par son anatomie macroscopique à peu près commune à chaque individu d'une même espèce, produite par un processus embryogénique semblable. Plus finement, son architecture microscopique, l'agencement de ses neurones, de ses synapses, est le fait de son expérience propre, et confère au cerveau une configuration locale. Cette configuration très complexe, faite de ces quelques cent milliards de neurones, voit son activité nécessairement délimitée à un espace d'activité (pour ne pas parler de graphe de transition) à cent milliards de dimensions, chacune représentant l'état continu ou discret d'un neurone. Quel que soit l'encodage des signaux, leur synchronie ou non, quoi qu'on fasse, le nombre de dynamiques atteignables est limité. L'ensemble des pensées et traitements faits dans notre cerveau existe dans un tel espace des dynamiques possibles. Étant donné une configuration cérébrale/neuronale, l'ensemble des configurations atteignables existe dans cet espace. Toutes les pensées de quelques milliards de dimensions, toutes complexes qu'elles soient (des milliards de dimensions c'est vraiment beaucoup ; et des milliards plus une c'est vraiment beaucoup plus), ne sont donc pas atteignables. J'aime cette idée de l'inaccessibilité à certaines perceptions ou certaines façons de penser. Mais j'aime aussi l'idée que le cerveau puisse au moins partiellement simuler des pensées 'moins complexes', qui

devraient résulter de réseaux neuronaux de dimension bien inférieure. Un cerveau est-il capable de simuler toutes les pensées d'une dimension bien inférieure ? Je ne suis pas bien certain que cette idée rencontre un franc succès auprès des personnes qui travaillent sur le fonctionnement du cerveau et la conscience, néanmoins je trouve l'expérience de pensée sur l'atteignabilité des dynamiques/comportements intéressante. Cette expérience de pensée est transposable à d'autres réseaux biologiques, génétiques notamment.

9.1.2. Recherche des réseaux pluripotents

En pratique, avec les PBNs (pour ne pas se limiter aux SBNs, donc ne pas limiter le nombre de GT), on peut confortablement expérimenter de nombreux problèmes relatifs à ce sujet. Le premier est d'explorer l'existence de réseaux pluripotents à partir de graphes de transition de réseaux de dimensions respectives inférieures. La question générale se formule ainsi : « soient k graphes de transition de dimension d (on considérera d commune à tous les graphes de transition simulés), pouvant être générés par des PBN ; existe-il au moins un tel graphe de transition de dimension minimale (c'est-à-dire intégrant tout juste les graphes de transition des réseaux à simuler en remplissant son propre graphe de transition) simulé par un PBN ?

Nous commencerons par limiter notre exploration aux petits réseaux dont voici un exemple concret avec des réseaux de dimensions fixées (ici 2 réseaux de dimension 2 dont les comportements sont simulés par un réseau de dimension 3) :

- soit R_1 avec deux nœuds N_{11} et N_{12} avec son graphe de transition associé G_1 .
- soit R_2 avec deux nœuds N_{21} et N_{22} avec son graphe de transition associé G_2 .
- soit R avec trois nœuds N_1 , N_2 et N_3 avec son graphe de transition associé G .

Est-il possible que R soit tel que son mutant avec l'état du nœud $N_3 = 0$ possède le graphe de transition G_1 et celui avec $N_3 = 1$ possède le graphe de transition G_2 ?

S'ensuit le théorème suivant à investiguer : pour tout R_1 et R_2 , de dimension 2, existe-t-il R de dimension 3 avec un mutant sur un nœud $N_3 = 0$ dont le graphe de transition est G_1 et avec le mutant $N_3 = 1$ dont le graphe de transition est G_2 . Nous explorerons systématiquement l'espace des couples de graphes $\{G_1, G_2\}$ en recherchant le ou les réseaux R capables de les simuler sur des hyperfaces indépendantes. Certains couples $\{G_1, G_2\}$ ne pourront pas être obtenus en simulation par un PBN R de dimension 3. D'où les questions : quelle proportion de couples $\{R_1, R_2\}$ peuvent être simulés par un réseau R ? par plusieurs réseaux R différents ? Et surtout, quelles caractéristiques communes ont de tels graphes de transition qui autorisent, ou au contraire ne permettent pas, leur simulation par un réseau R ? Et enfin, quelles sont les caractéristiques de ces PBN R (centralité, rareté des fonctions ...) ?

Il faudra bien entendu examiner le cas où le théorème est faux (*i.e.* il y a des cas qui demandent des R de plus grande dimension, 4 ou plus dans l'exemple ici considéré), les conditions sur G_1 et G_2 le rendant correct. On pourra donc étendre ensuite le problème avec des dimensions quelconques (ex : couples de réseaux de dimension 3 simulés par des réseaux de dimension 4) et/ou avec davantage de réseaux simulés par un même réseau R . Enfin, on pourra réaliser le même travail en utilisant des

TBNs ou des SBNs, l'intérêt étant de voir quelles fonctions TBFs ou SBFs prennent en charge les couples de graphes $\{G_1, G_2\}$ et quels sont les couples manquants liés aux limitations d'expressivité de ces formalismes.

9.1.3. Réseaux simulants – réseaux non simulants

Je vais apporter ici une précision par rapport à ce qui a été dit juste au dessus. Cette précision ajoute une restriction dans la recherche des réseaux simulants et met en évidence les nœuds et séquences de contrôle associés.

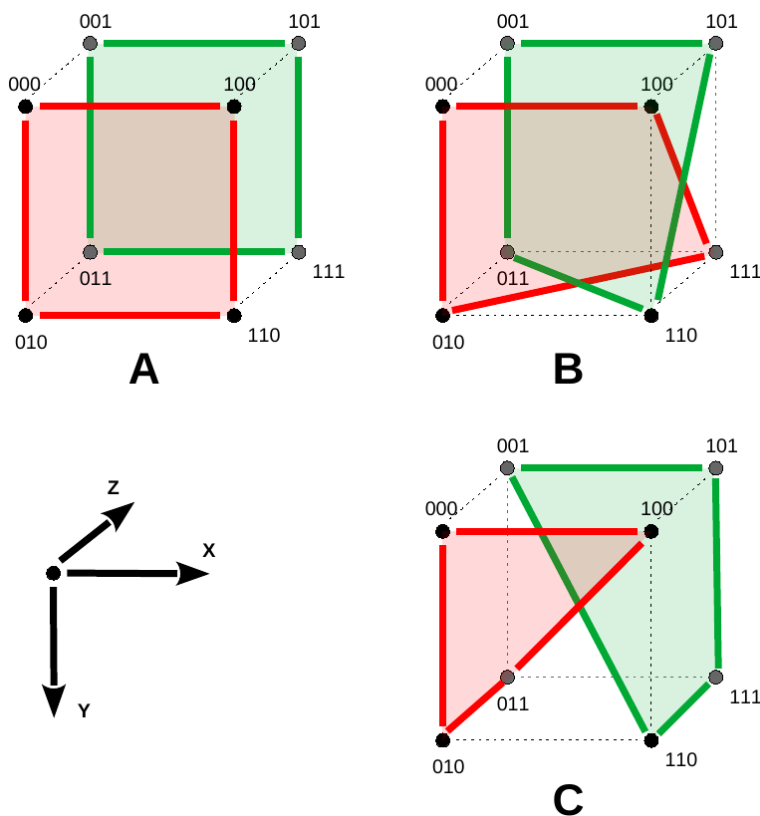


Figure 9.2. Trois hypercubes de dimension 3 contenant chacun 2 hypercubes de dimension 2. (A) et (B) Dans ces 2 exemples, une direction pure (ici z) n'est pas utilisée dans les 2 hypercubes sous-jacents. Ceux-ci sont donc simulés. En revanche, ce n'est pas le cas en (C) : il n'est pas possible d'utiliser une transition impliquant un seul nœud à la fois pour extraire les hypercubes sous-jacents. Ceux-ci ne sont donc pas simulés.

Un réseau (de dimension d' , grande) peut, comme nous le proposons, exprimer la simulation de réseaux de dimensions inférieures (dimensions d_i), des *réseaux simulés*. C'est bien entendu, comme nous l'avons dit, leurs graphes de transitions qui sont simulés dans le comportement du réseau de dimension supérieure. Cependant, il ne paraît forcément évident qu'un ensemble de 2^{d_i} états d'un réseau de dimension d (donc contenant 2^d états avec $d_i < d$) constitue le graphe de transition d'un réseau de dimension d_i . En tout cas, cela peut ne pas être toujours le cas si, comme on peut le penser idéalement, les nœuds qui servent à la simulation des réseaux simulés ne sont pas aussi les nœuds de contrôle, c'est-à-dire ceux qui décident de la simulation en cours (le comportement que le système adoptera). Prenons un exemple comme celui illustré figure 9.2. Dans cette figure, des hypercubes de dimension 3 sont représentés (8 états fonctions des valeurs Booléennes des nœuds x , y et z). Ils contiennent chacun 2 hypercubes de dimension 2 (4 états chacun, représentés respectivement en vert

et en rouge). Dans les cas des hypercubes montrés en A et en B de cette figure, on voit bien qu'une séquence de bits de contrôle appliquée à un seul nœud, indépendant de ceux impliqués dans la simulation des graphes de transition contenus dans les hypercubes simulés de dimension 2, est responsable de la sélection de l'hypercube simulé actif (le vert ou le rouge). Il est possible dans ces 2 cas de trouver au moins un axe (parmi les axes x , y ou z ici) selon lequel une projection permet de récupérer tous les états de chacun des hypercubes simulés. En effet, dans les deux cas montrés dans cet exemple figure 9.2, ce nœud de contrôle est le nœud z . Pour l'hypercube A, la séquence 0000 appliquée sur z permet la sélection de l'hypercube rouge, tandis que la séquence 1111, toujours appliquée à z , sélectionne le vert. Pour l'hypercube B, ce sont des séquences de contrôle toutes deux appliquées au nœud z qui permettent la sélection d'un hypercube simulé ou de l'autre (0001 dans le cas du rouge, 1110 dans le cas du vert). Une projection selon z permet de récupérer la totalité des états des 2 hypercubes simulés. C'est ce qui est montré figure 9.3 A. En revanche, on voit qu'aucune projection (ni selon x , ni y , ni z) ne permet de récupérer entièrement les états des hypercubes dans le cas de l'hypercube de dimension 3 montré figures 9.2 C et 9.3 B. Les hypercubes de dimension 2 contenus dans cet hypercube de dimension 3 ne sont donc pas une simulation au sens où nous l'avons exprimé.

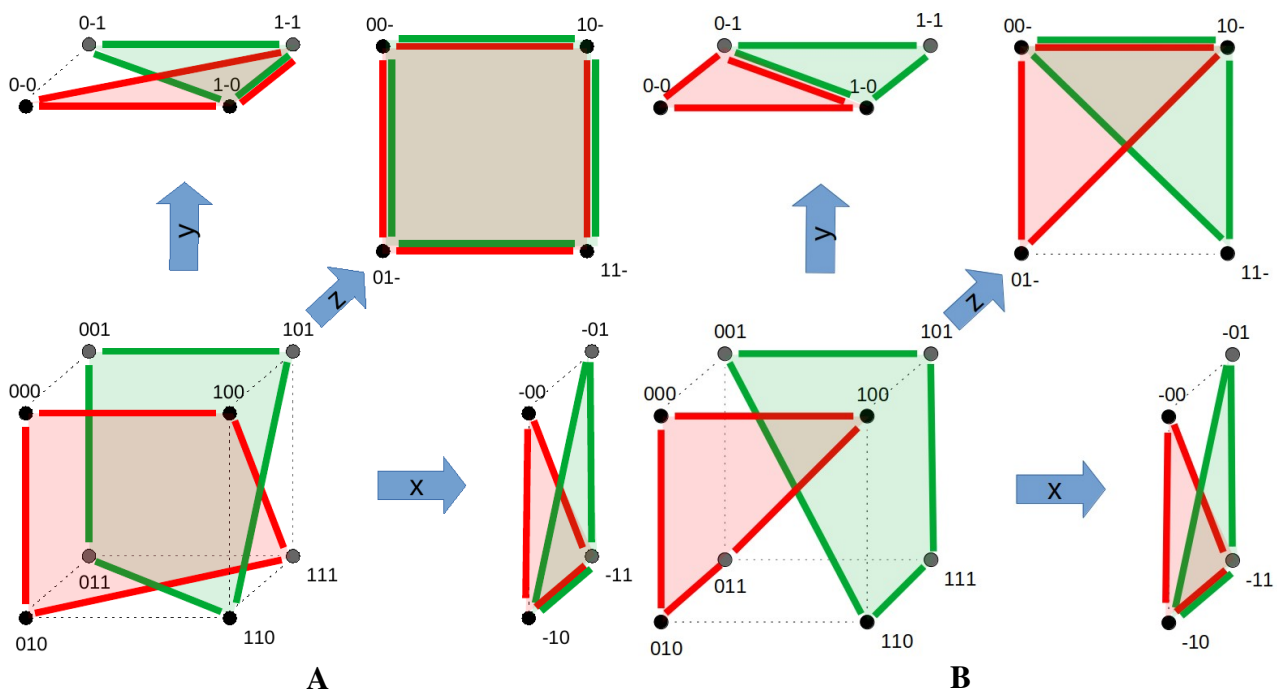


Figure 9.3. Deux hypercubes de dimension 3 contenant chacun 2 hypercubes sous-jacents de dimension 2, et leurs projections selon x , y et z . (A) On voit que s'il n'est pas possible d'extraire entièrement les hypercubes de dimension 2 (récupérer tous leurs états (4 états)) par une projection selon x ou y , cela est en revanche possible selon z . Le nœud z est donc un nœud de contrôle et les 2 hypergraphes, vert et rouge, sont donc simulés. (B) On voit que toutes les projections (x , y et z) échouent à extraire les 2 hypercubes sous-jacents. Il n'y a pas de nœud de contrôle et aucun des hypercubes sous-jacents n'est simulé.

Pour que des hypercubes sous-jacents soient considérés comme simulés, il faut qu'il existe une direction, dans l'hypercube simulant, selon laquelle la projection permet l'extraction de ceux de

dimensions inférieures. Il faut donc qu'il existe au moins un bit de contrôle (impliquant 1 nœud). Si le changement d'un seul bit à la fois permet de sélectionner un hypercube de dimension plus petite (hypercube sous-jacent) que la dimension de l'hypercube simulant, alors il est simulé. Dit autrement, si un hypercube sous-jacent présente l'absence d'au moins une transition n'impliquant qu'un seul nœud, alors il est simulé car cette transition à 1 seul nœud manquante est forcément un bit de contrôle.

Élargissons aux plus grandes dimensions !

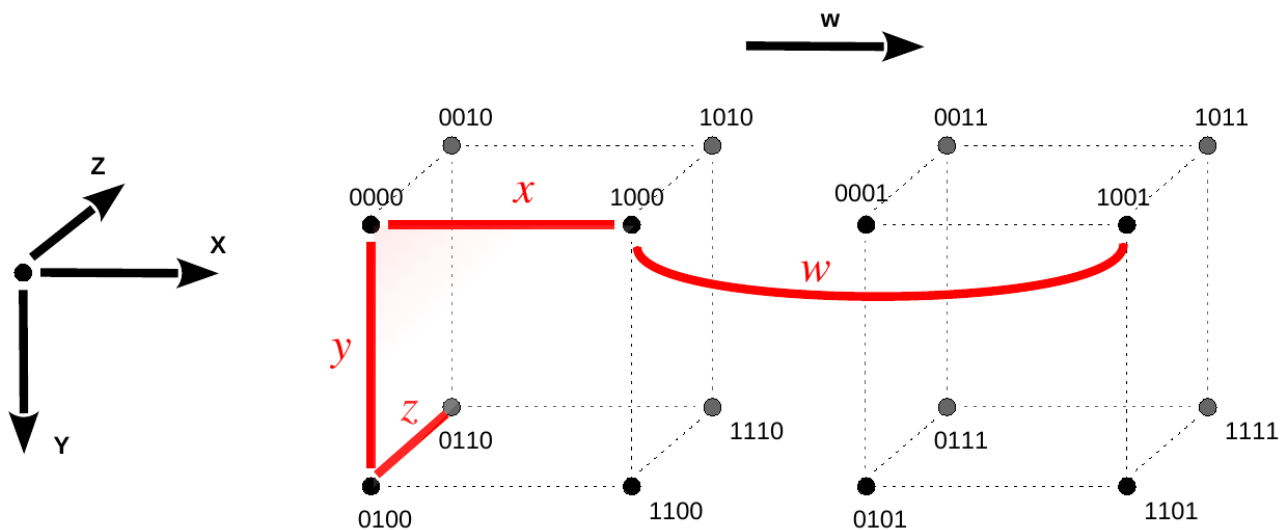
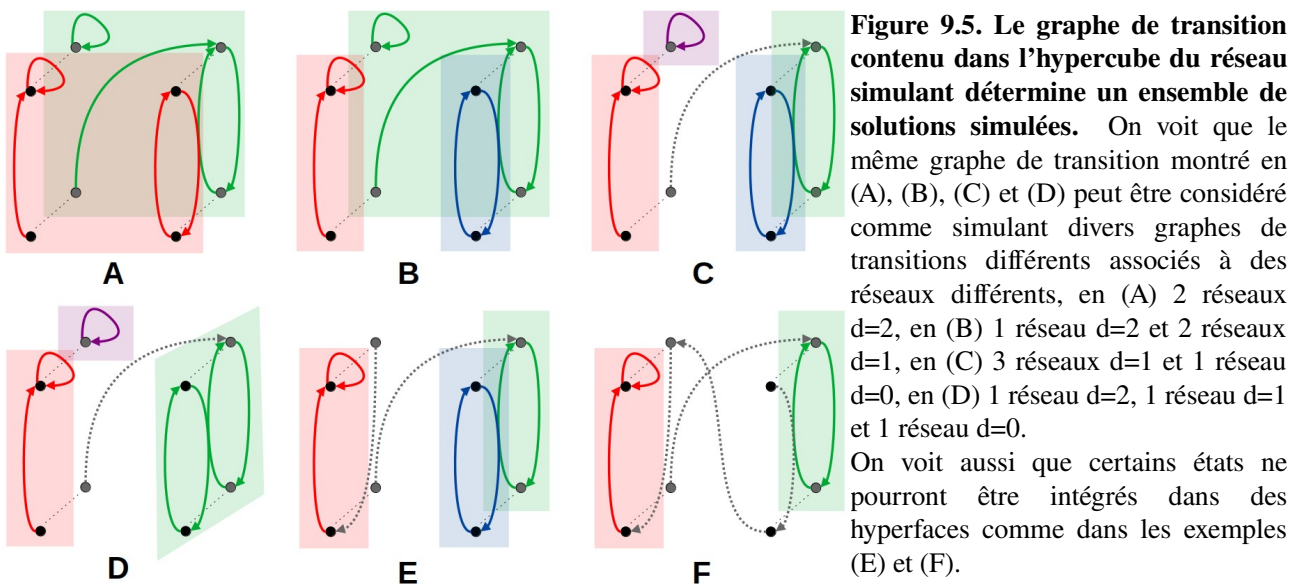


Figure 9.4. Un hypercube de dimension 4 et un graphe contenant les 4 directions disponibles x , y , z et w . Cette figure illustre le fait qu'il n'est pas possible de construire un hypercube de dimension 2 (4 transitions fermées, c'est-à-dire bouclant sur 4 états) avec les 4 transitions pures (x , y , z , w) que contient un hypercube de dimension 4. Tous les hypercubes de dimension 2 présentent l'absence d'au moins 1 direction pure ; ils sont donc tous simulés.

Comme en 4 dimensions il existe 4 transitions n'impliquant qu'un seul nœud (selon x , y , z ou w) et qu'un hypercube de dimension 2 comprend 4 états, il n'est pas possible d'exprimer ces 4 directions pures (x , y , z ou w) dans un seul hypercube de dimension 2 contenu dans le graphe de dimension 4 (voir figure 9.4). Cela signifie qu'il existe toujours, pour tout hypercube de dimension 2 plongé dans un hypercube de dimension 4, au moins 1 direction (x , y , z ou w) correspondant au contrôle. Tout hypercube de dimension 2 plongé dans un hypercube de dimension 4 est simulé !

Le corollaire, c'est que tous les hypercubes de dimension d plongés dans un hypercube de dimension d' avec $d' \geq 2^d$ (d directions pures disponibles et 2^d états pour l'hypercube sous-jacent) sont simulés. Par exemple, un hypercube de dimension 3 plongé dans un hypercube de dimension 4 n'est pas nécessairement simulé car $4 < 2^3$; il est possible de trouver les 4 transitions pures (x , y , z et w) de l'hypercube de dimension 4 dans l'hypercube de dimension 3 sous-jacent. En revanche, un hypercube de dimension 3 plongé dans un hypercube de dimension 8 est nécessairement simulé car $8 \geq 2^3$ (8 directions pures disponibles et 2^3 états pour l'hypercube sous-jacent). Pour simuler de façon certaine des hypercubes de dimensions d inférieures, avec d croissant il faudra très vite des

réseaux simulants de très grande dimension d' , car d' doit augmenter comme la puissance d de 2. Pour simuler de façon certaine un réseau Booléen de seulement 40 nœuds, il faudra plus de 100 milliards de dimensions (près de 110 milliards) ... le nombre de neurones du cerveau ! Tout graphe de transition généré par un réseau à 40 nœuds trouvera au moins un réseau simulant de près de 110 milliards de dimensions. Mais il est tout aussi possible de trouver des réseaux simulés de plus grande taille, si son graphe de transition (son hypercube) n'utilise que peu de dimensions du réseau simulant ... s'il est mieux agencé, mieux rangé, laissant plus de place à d'autres hypercubes et libérant des bits de contrôle.



Un réseau de dimension d' grande pourra donc accueillir dans l'hypercube contenant sa dynamique plusieurs simulations de comportements de réseaux de dimensions inférieures, mais aussi des dynamiques qui ne correspondent pas à des comportements simulés.

De façon générale, on pourra se poser la question, dans le cas d'un mélange d'hypercubes de dimensions d_j homogènes ou non (mélange de dimensions), de la probabilité que ceux-ci soient assimilables à des simulations ou non. On pourra étudier l'ensemble des mélanges possibles d'hypercubes sous-jacents et déterminer la proportion de simulés et de non simulés. On pourra étudier les caractéristiques des réseaux simulants en fonction des proportions d'hypercubes qu'ils simulent. Les aspects techniques de cette recherche de réseaux simulants à partir de réseaux de plus petites dimensions ou, inversement, la recherche dans des graphes de réseaux de grande dimension d'hyperfaces contenant chacune des graphes de transitions indépendants (puis la recherche des ensembles de réseaux satisfiables pour ces dynamiques) constitue un challenge. En particulier car il n'y a pas nécessairement unicité des solutions concernant ne serait-ce que la détermination des hyperfaces à partir d'un grand graphe de transition, et encore moins des réseaux à l'origine de tels graphes de transition simulés (contenus dans les hyperfaces) (figure 9.5).

Si l'on essaie de ramener cela à un sens biologique, c'est dans le contexte évolutif notamment qu'on verra un intérêt à ces considérations théoriques. Un peu plus loin j'aborde la question du contrôle, du développement des grands réseaux d'interaction biologiques à partir de petits réseaux initiaux. En cherchant à utiliser cette idée hypercubes simulés par des réseaux de plus grande dimension, on comprend que ce problème évolutif a un lien avec le peuplement d'espaces de comportements (les hypercubes de grande dimension) par des sous-espaces de comportements disjoints, reliés uniquement par le contrôle assuré par certains nœuds particuliers. Une autre façon d'augmenter cet espace de comportements est de passer par leur duplication puis leur mutation. Des systèmes (réseaux) dupliqués puis divergeant progressivement constituent une source de nouveaux comportements dont on peut espérer qu'il ne diffèrent que progressivement aussi, comme le suggèrent les premières observations que nous avons faites concernant les distances entre réseaux et entre comportements (voir la sous-section 9.2 ci-dessous) : il ressort que, pour les petits réseaux sur lesquels nous avons travaillé (dimensions 2 et 3), de petites distances (de Hamming) entre réseaux sont effectivement très fortement corrélées à de petites distances (de Hamming également) entre graphes de transitions.

Dans tous les cas, dans le cadre évolutif, on peut s'attendre à une forme d'offuscation progressive des réseaux simulants, autrement dit à ce que des nœuds initialement dédiés à certains comportements, différents des nœuds de contrôle, puissent peu à peu se confondre avec des éléments du contrôle. L'hypercube sous-jacent initialement simulé (ayant au moins une direction pure manquante) pourrait apparaître comme non-simulé (pas de direction pure manquante), voire particulièrement offusqué avec le contrôle si chaque direction pure est impliquée plusieurs fois dans l'hypercube sous-jacent.

9.2. Retour sur les distances et la structure de l'espace des réseaux

J'ai déjà évoqué cet espace dans lequel les réseaux, selon un ensemble de relations de distance, sont plus ou moins voisins les uns des autres. J'y reviendrai d'ailleurs plusieurs fois dans ce chapitre. Je vais cependant introduire quelques précisions sur la manière dont il faut voir cet espace et les distances entre réseaux et graphes de transition.

Commençons par rappeler qu'une distance c'est un modèle : le choix de cette mesure n'est pas objectif car il exprime la manière dont nous souhaitons apparenter ou différencier des objets. Il y a donc une infinité de manières de concevoir une distance entre deux réseaux. De même entre leurs dynamiques. Le choix des distances a des conséquences sur la façon dont l'espace des réseaux peut être compris. Prenons le temps de considérer l'exemple de 2 distances structurelles possibles et de les définir ; c'est un peu mathématique mais, finalement, nul besoin de s'attarder sur les expressions, le texte suffit :

Considérons une distance structurelle de Hamming entre deux réseaux (graphes) g et g' :

$$d_{SH}(g, g') = \sum_{i=1}^d \sum_{j=1}^d |g_{ij} - g'_{ij}|, \quad \forall g, g' \in \mathbb{G} \text{ avec } \mathbb{G} \text{ l'ensemble des réseaux (graphes)}$$

La distance de Hamming qu'on a définie accorde la même importance à chacun des arcs (proportionnellement à son poids), mais on peut chercher à définir une autre distance qui s'appuie sur les chemins entre nœuds qui constituent un réseau en accordant plus d'importance à certains arcs qu'à d'autres. Par exemple, il peut exister des systèmes biologiques où l'importance d'une interaction dans le réseau impacte directement le coût de modification de cette interaction.

En théorie des graphes, définir l'importance d'un arc revient à définir ce que l'on appelle un indicateur de centralité [Nushi, 2021]. Il n'y a pas de définition universelle de la centralité ; il faut poser des critères venant du système à modéliser. Ici, on se fixera deux critères, illustrés avec la figure 9.6:

- à poids équivalent, un arc a plus d'importance s'il est plus impliqué dans la connectivité du réseau.
- à position équivalente, un arc a plus d'importance si son poids (en valeur absolue) est plus élevé.

On veut construire une distance pour laquelle une modification sur un arc a un coût d'autant plus élevé que l'arc a de l'importance. Notre nouvelle distance doit prendre en compte la connectivité du réseau : on devra faire un calcul non pas sur les arcs, mais sur les chemins qu'il contient. Ici, on ne considère que les chemins qui ne constituent pas de cycles, afin de maintenir les calculs sur des ensembles finis. En outre, on doit étendre la notion de poids aux chemins. On introduira pour cela la notion d'influence d'un arc et d'un chemin dans un graphe g .

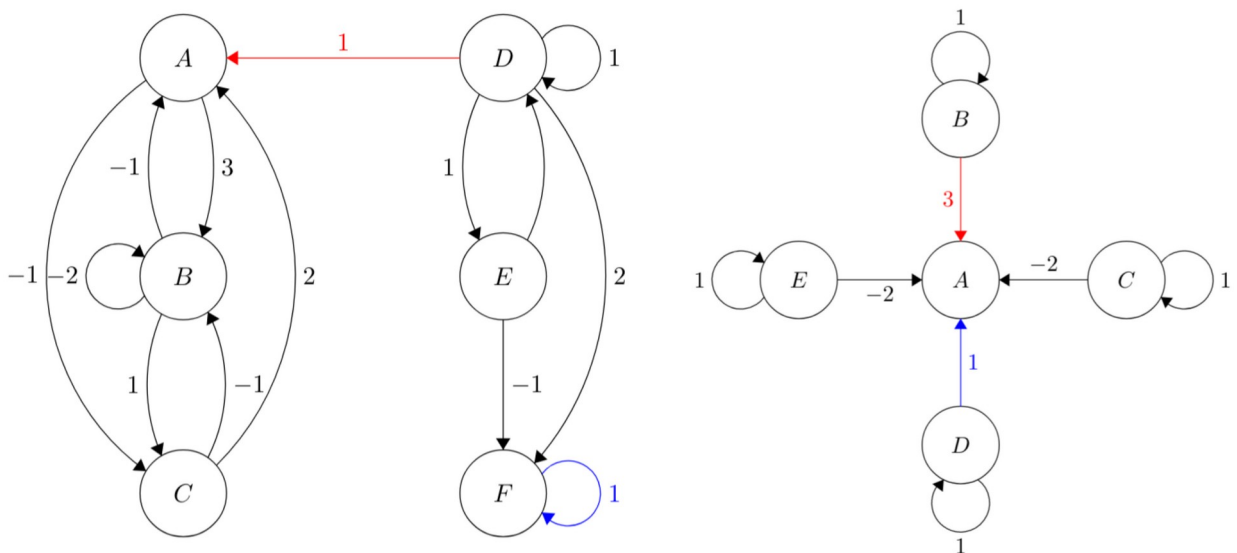


Figure 9.6. Deux illustration de l'importance d'un arc dans un réseau à poids équivalent (à gauche) et à position équivalente (à droite). (Gauche) Tous les arcs sur cette figure n'ont pas la même importance. L'arc DA (en rouge) est le seul à faire communiquer le bloc DEF avec le bloc ABC, tandis que l'arc FF (en bleu) n'influe que sur le sommet F, car celui-ci n'a pas d'arc sortant. (Droite) Ici, l'arc BA (en rouge) et l'arc DA (en bleu) ont une position équivalente, mais ils n'ont pas le même poids. On s'attend à ce que l'arc DA ait moins d'importance que BA car il est de poids plus faible.

On définit l'influence d'un arc a à la position g_{ij} ainsi :

$$I_g(A) = \frac{g_{ij}}{\sum_{k=1}^d |g_{ik}|} \quad \text{où } g_{ij} \text{ est l'arc allant de } i \text{ à } j$$

L'influence d'un arc représente la corrélation que l'on peut espérer entre la valeur des deux sommets liés. Si elle vaut 1, le sommet de départ détermine complètement l'état du sommet d'arrivée. Si elle est proche de zéro, alors la probabilité que la valeur du sommet de départ ait une importance pour déterminer la valeur du sommet d'arrivée est faible.

On définit ensuite l'influence d'un chemin c par la formule : $I_g(c) = \prod_{a \in \text{arcs}(c)} I_g(a)$

Ainsi, si l'influence d'un chemin est proche de 1, cela signifie que l'état du sommet de départ détermine en grande partie l'état du sommet d'arrivée (après un certain nombre d'étapes). Un chemin d'influence nulle signifie qu'un de ses arcs est de poids nul, donc qu'il n'existe pas dans le réseau.

Une fois avoir défini l'influence d'un chemin, on peut alors définir la distance d_{SC} sur les chemins entre deux réseaux g ainsi :

$$d_{SC}(g, g') = \sum_{c \in \text{paths}(g) \cap \text{paths}(g')} |I_g(c) - I_{g'}(c)| + \sum_{c \in \text{paths}(g) \setminus \text{paths}(g')} |I_g(c)| + \sum_{c \in \text{paths}(g') \setminus \text{paths}(g)} |I_{g'}(c)|$$

ou de manière équivalente, en notant K_d le graphe complet à d sommets :

$$d_{SC}(g, g') = \sum_{c \in \text{paths}(K_d g)} |I_g(c) - I_{g'}(c)|$$

Notons que d_{SC} ne vérifie pas l'axiome de séparation : en effet, pour $\lambda \in \mathbb{R}_+^*$, $d_{SC}(g, \lambda g) = 0$. Il s'agit donc formellement d'une *pseudométrie* : une fonction qui vérifie tous les critères des distances sauf l'axiome de séparation. Cette dégénérescence est parfaitement justifiée : si l'on considère que les valeurs des poids n'ont de sens que relativement les unes par rapport aux autres, il est logique de considérer que deux réseaux identiques à multiplication par un réel strictement positif près sont exactement les mêmes réseaux.

Revenons à ce qui nous intéresse ici : la structure de l'espace des réseaux.

La distance de Hamming d_{SH} est à valeurs entières et permet de définir de manière très claire la notion de réseaux voisins : deux réseaux g et g' sont dits voisins si $d_{SH}(g, g') = 1$. Cela nous permet également de définir clairement une mutation : la transformation d'un réseau en un de ses voisins. Ainsi, l'espace des réseaux prend (lui-même) la forme d'un graphe non orienté où chaque réseau est connecté à ses voisins (on a un réseau de SBNs).

La distance sur les chemins d_{SC} prend cependant un très grand nombre de valeurs différentes rationnelles, et il est plus ambigu de définir à partir de quelle distance deux réseaux sont voisins. Il est alors plus cohérent de considérer l'espace des réseaux comme un espace métrique où chaque réseau est bel et bien en relation avec tous les autres et non juste ses voisins comme cela peut-être vu avec une distance de Hamming.

Ainsi, suivant si d_s est à valeurs discrètes ou continues, on se retrouve deux structures bien différentes pour l'espace des réseaux. Les deux structures ont leurs propres avantages et inconvénients. La structure de graphe est plus simple et bien adaptée à une première approche des SBN, mais la distance utilisée ne contient en réalité que peu d'informations et peut ne pas traduire une complexité existante. La structure d'espace métrique permet de poser un formalisme qui conserve une certaine finesse, mais on perd la relation de voisinage, sauf à introduire des seuils de distance.

En outre, à nouveau, certains systèmes se prêtent mieux à l'étude avec une certaine structure que d'autres. On peut comprendre que des réseaux génétiques tendent à évoluer par petites mutations que l'on peut énumérer (comparaisons de séquences génétiques) tandis que les écosystèmes peuvent faire de grands pas d'évolution, difficiles à évaluer, chaque espèce évoluant séparément.

Notons que la frontière entre graphe et espace métrique est perméable. D'une part, un graphe peut aussi être vu comme un espace métrique (en ne se limitant pas au sauts de 1 mutation) et, d'autre part, si d_s est à valeurs continues, on peut fixer un seuil en-deçà duquel deux réseaux sont considérés comme voisins et obtenir ainsi une structure de graphe.

Du fait de cette distinction, il sera nécessaire pour définir certaines notions, comme celles de trajectoires dans l'espace des réseaux, mais aussi la robustesse et l'évolutivité, de prêter attention à la structure de l'espace dans laquelle on travaille.

Pour ces notions, il faudra aussi faire appel à ce qui permet de mesurer des différences entre comportements, des distances dynamiques. Là aussi, le paysage des dynamiques (des graphes de transition) peut être vu soit comme un graphe (et dans ce cas on parlera de mutations élémentaires du comportement et on pourra définir une notion de voisinage des comportements), soit de façon plus continue, dans un espace métrique (et dans ce cas la notion de voisinage n'existera qu'au seuil près mais on disposera de distances pouvant prendre en compte finement plusieurs aspects des dynamiques). On pourra alors définir des distances adaptées à ce que l'on souhaite tester (figure 9.7). Je vais ici en préciser deux.

De même que pour les structures, une distance de Hamming peut être utilisée pour différencier les dynamiques des réseaux. Pour cela, on définit d'abord une distance de Hamming entre deux états (ensemble des états des nœuds) s et s' du réseau :

$$\forall s, s' \in \mathbb{S}, \quad H(s, s') = \sum_{i=1}^d |s_i - s'_i|$$

et on définit la distance dynamique de Hamming entre deux réseaux :

$$d_{DH}(g, g') = \sum_{s \in \mathbb{S}} H(\text{succ}_g(s), \text{succ}_{g'}(s)) \quad \text{où } \text{succ}_g \text{ désigne la fonction de succession d'un état } s$$

$$\text{pour un graphe } g : \text{succ}_g : \begin{array}{l} \mathbb{S} \rightarrow \mathbb{S} \\ s \mapsto (g_i(s))_{1 \leq i \leq d} \end{array}$$

Cette distance représente donc comme proches les états de réseaux où la fonction succ ne change

que sur peu d'entrées, ce qui est une bonne première approche de la distance entre comportements.

Suivant le système étudié, comparer directement les fonctions de succession n'est pas forcément la meilleure représentation. Généralement, l'intégralité de la fonction $succ$ n'est pas nécessaire à l'étude du système, on a un surplus d'information. Il suffit souvent de ne connaître que les attracteurs et les bassins d'attraction. Définissons ces deux notions formellement :

On appelle attracteur d'un réseau g une suite d'états (s_0, s_1, \dots, s_k) telle que $\forall (0 \leq i \leq k-1)$, $succ_g(s_i) = s_{i+1}$ et $succ_g(s_k) = s_0$, avec k la taille de l'attracteur.

On appelle bassin d'attraction d'un attracteur $a = (s_0, \dots, s_k)$ (dans un réseau g) l'ensemble des états qui mènent à cet attracteur par applications répétées de la fonction $succ_g$:

$$BA(a) = \{s \in \mathbb{S}; \exists n \in \mathbb{N}; succ_g^n(s) = s_0\}$$

De ces définitions découlent les propriétés suivantes pour tout SBN g :

- L'état nul est un point fixe de g . En effet, la définition du comportement des SBNs induit $succ_g(0)=0$
- Tout état s appartient à un bassin d'attraction. En effet, la suite $(succ_g^n(s))_{n>0}$ est à valeurs dans un ensemble fini, donc est périodique à partir d'un certain rang. La période correspond à l'attracteur.
- L'ensemble des bassins d'attraction de g forme une partition de \mathbb{S} . Ce fait découle directement de la propriété précédente.

Bien souvent, il est important de considérer les attracteurs dans l'étude des systèmes car ils représentent leurs états asymptotiques. L'étude des bassins d'attraction est importante également dans une moindre mesure car ils témoignent de nombreuses propriétés de l'état asymptotique correspondant dans le système biologique. Suivant les modèles, un attracteur avec un plus grand bassin d'attraction aura une plus grande stabilité face aux changements d'états ou une plus grande probabilité d'apparaître.

On cherche donc à construire une distance pour l'espace métrique des dynamiques selon ces contraintes :

- Modifier (ou supprimer) un attracteur à petit bassin d'attraction coûte moins cher que de modifier (ou supprimer) un attracteur avec un grand bassin d'attraction. Ceci pourrait se justifier d'un point de vue évolutionnaire : les comportements les plus complexes ont été *a priori* plus difficiles à atteindre que les plus simples.
- Modifier un bassin d'attraction (hors attracteur) coûte moins cher que de modifier son attracteur. Le comportement est plus précieux que la façon d'y parvenir.

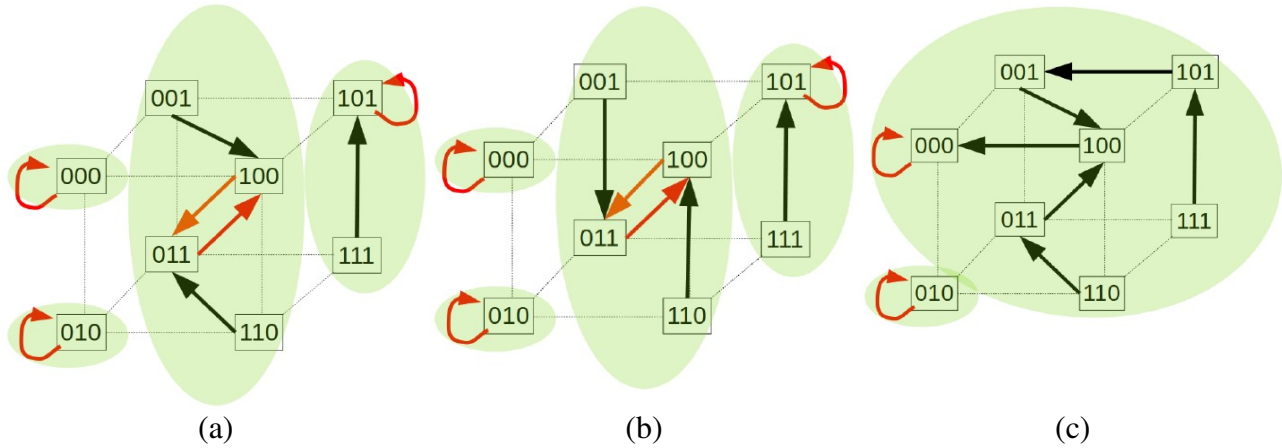


Figure 9.7. Trois graphes de transition avec différents bassins d'attraction. Selon la distance de Hamming, (a) est plus proche de (c) que de (b), mais les attracteurs et les bassins d'attraction sont plus ressemblants entre (a) et (b) qu'entre (a) et (c). Une distance alternative serait utile pour mesurer ces différences.

Si nous décidons de prendre en compte la taille des bassins d'attraction, il faut effectuer un calcul sur chaque état du bassin d'attraction. L'ensemble des bassins d'attraction étant une partition de \mathbb{S} , il faut finalement faire un calcul sur \mathbb{S} .

Introduisons une distance entre les attracteurs pour les différencier. Il n'est pas évident de mesurer une distance entre deux attracteurs de longueurs différentes car une boucle n'a pas de début ni de fin et qu'on compare des objets de tailles différentes. On va devoir tester tous les débuts de boucle possibles et prendre le minimum, ce qui donne pour deux attracteurs de même taille :

$$d_{cycle,1}(C_1, C_2) = \min_{1 \leq j \leq |C_1|} \left(\sum_{i=1}^{|C_1|} H((C_1)_{i+j}, (C_2)_i) \right)$$

où l'on a $(C_1)_1 = (C_1)_{|C_1|+1}$.

Concernant la longueur, on choisit ici d'allonger le cycle le plus court en ajoutant l'état nul autant de fois que nécessaire, ce qui nous donne pour un début de boucle arbitraire si $|C_1| \leq |C_2|$:

$$d_{cycle,2}(C_1, C_2) = \sum_{i=1}^{|C_2|} H((C_1)_i, (C_2)_i) + \sum_{i=|C_2|+1}^{|C_1|} H((C_1)_i, 0)$$

et en combinant les deux formules :

$$d_{cycle,1}(C_1, C_2) = \begin{cases} \min_{1 \leq j \leq |C_1|} \left(\sum_{i=1}^{|C_2|} H((C_1)_{i+j}, (C_2)_i) + \sum_{i=|C_2|+1}^{|C_1|} H((C_1)_{i+j}, 0) \right) & \text{si } |C_1| \leq |C_2| \\ d_{cycle}(C_2, C_1) & \text{sinon} \end{cases}$$

Avec ces deux mesures, on peut enfin définir une distance d_{DA} sur les attracteurs des réseaux, en notant $lim_g(s)$ l'attracteur dans lequel un état s dans g aboutit :

$$d_{DA}(g, g') = \sum_{s \in \mathbb{S}} d_{cycle}(lim_g(s), lim_{g'}(s))$$

Cette distance remplit bien nos deux critères de construction : si seul le bassin d'attraction (hors attracteur) est modifié, seuls les états déplacés produiront une différence, tandis que si l'attracteur change, tous les états du bassin produiront une différence. En outre, si un attracteur est modifié, le coût de la modification est proportionnel à la taille du bassin d'attraction.

Comme on le voit en figure 9.7, deux comportements (a) et (b) peuvent avoir les mêmes attracteurs et les mêmes bassins d'attraction. Ainsi, de même que la distance sur les chemins vue plus tôt, la distance sur les attracteurs ne respecte pas non plus l'axiome de séparation : c'est une pseudométrique. Cela ne cause pas de problème non plus : travailler avec d_{DA} suppose que seuls les attracteurs comptent dans l'étude. Il est donc normal de ne pas différencier les comportements sur d'autres critères.

Voilà, c'était un passage un peu formel (ce qui est rare finalement dans ce document), mais il a le mérite de faire ressortir 2 points importants : d'abord qu'on peut définir des distances sur les réseaux et sur les dynamiques, ce qui nous permet d'expérimenter sur ces systèmes, mais surtout que ces distances dépendent de ce que l'on veut tester, ceci requérant une connaissance perçue du système naturel que l'on modélise (réseau génétique, réseau métabolique, réseau de neurones, écosystème, *etc*) et de ce que l'on perçoit comme différences, impliquant que de telles distances sont nécessairement subjectives. Ces outils vont néanmoins nous permettre d'avancer dans une définition formelle de la robustesse, de l'évolutivité et de la viabilité des systèmes vivants.

9.3. Formaliser l'évolution - Quelques mots sur la Robustesse, l'Évolutivité et la Viabilité

Ce n'est pas parce qu'un comportement peut exister qu'il est viable. Wagner dans son article traitant de l'évolutivité [Wagner, 2008] montre qu'un système vivant doit résoudre le paradoxe d'être à la fois robuste aux changements (son génotype ne doit pas changer trop vite) et à la fois évolutif (le génotype doit être en mesure d'évoluer et d'explorer un paysage morphogénétique au gré des changements de l'environnement). Wagner inclut implicitement dans son analyse la notion de viabilité : la robustesse et l'évolutivité du génotype sont conditionnées par la viabilité du phénotype. Nous pensons qu'il est nécessaire de découpler le concept de viabilité de ceux de robustesse et d'évolutivité. La robustesse peut être vue comme inhérente à l'organisme ; elle provient de multiples mécanismes biochimiques et biologiques qui, d'une part freinent la dérive génétique et imposent une redondance (redondance du code génétique, redondance fonctionnelle des acides aminés, redondance des voies métaboliques, de contrôle, *etc*), d'autre part vient du fait que toute variation doit tout de même permettre le fonctionnement général de l'organisme, indépendamment des contraintes que lui impose l'environnement. La robustesse conditionne l'existence d'une fonction biologique. L'évolutivité n'est pas en reste : elle peut être aussi formulée de manière indépendante, comme une capacité à muter et à autoriser un certain fonctionnement des systèmes de l'organisme. La viabilité, elle, est la contrainte environnementale (environnement physico-chimique, radiatif,

cellulaire, tissulaire, mécanique, social, *etc*) testant le phénotype. Elle est extrinsèque.

La trajectoire d'un système au cours de l'évolution peut être donc vue comme dépendant de ces 3 caractéristiques, 2 intrinsèques, 1 extrinsèque, respectivement sa robustesse, son évolutivité et sa viabilité. Un système quand il évolue, voyage dans le paysage morphogénétique, lui même modifié dans le temps par un environnement changeant (en partie parce que les organismes vivants le modifient). Vu *a posteriori*, on pourra étudier ces trajectoires de systèmes vivants dans le cadre de problèmes d'atteignabilité (voir section suivante).

J'ai présenté la viabilité comme un critère s'imposant à la dynamique des systèmes (l'expression de leur phénotype). Cependant la viabilité est dynamique mais aussi structurelle ! Les gènes sont faiblement connectés au reste du génome (la régulation de chaque gène n'implique l'activité que d'un nombre très limité de gènes), tandis que les neurones sont fortement connectés à de nombreux neurones dont ils captent l'état (chaque neurone peut recevoir plusieurs milliers d'afférences). Partir d'un réseau faiblement connecté comme un réseau génétique, pour en atteindre un autre également faiblement connecté, ne peut se faire qu'en passant par des réseaux dont la dynamique est compatible avec la vie (viabilité imposée au phénotype) mais, en plus, cela doit se faire en passant par des réseaux dont la structure aussi est compatible avec le système biologique qu'il décrit. Chaque structure locale s'articule avec un ensemble plus large. Changer une telle structure locale peut casser ses relations avec le reste du système auquel il était relié. Par exemple, en considérant un SBN, si l'une des relations qui unit deux SBFs change de poids (par exemple en passant de la fonction $(-1,-1,2,3)$ à la fonction $(-1,-1,1,3)$), alors sa relation de dominance (fonction maître ou esclave) avec le reste du système va possiblement changer (dans l'exemple donné, la fonction évolue vers une moins grande dépendance vis-à-vis du reste du système). La viabilité s'impose donc aussi au maintien des relations structurelles dans un système.

Encore une fois, les réseaux formels comme les SBNs sont un excellent moyen de réfléchir à ceci. En 2021, nous avons engagé des travaux dans ce sens avec un excellent stagiaire de L3 de l'ENS Paris, Rémy Desfossez. Je le cite car je trouve son travail remarquable ; c'est en partie son texte que j'utilise dans cette section. Rémy s'est attelé à définir plus précisément qu'Urvan Christen (étudiant en M1 à l'X qui avait magistralement initié ce travail) des distances possibles entre SBNs et entre leurs dynamiques. Il a aussi introduit cette distinction entre structure de graphe et structure métrique pour l'espace des réseaux (section 9.2). Mais Rémy s'est très vite intéressé aux concepts de robustesse, d'évolutivité et de viabilité qu'il a commencé à formaliser en fonction des distances entre réseaux et entre dynamiques. Je vais présenter une partie de son travail (dont je reprends une partie de ses mots dans les sections 9.2 et 9.3), encore très perfectible mais qui constitue pour nous un point de départ attendant un bon étudiant.

Dans la suite de cette section, je présenterai le formalisme que Rémy a construit pour permettre de définir les notions relatives à l'évolution, tout en garantissant une grande flexibilité par rapport au modèle choisi. Comme tous les concepts ici développés, de la distance à la viabilité, sont subjectifs, il apparaît nécessaire de pouvoir modifier une définition du système formel sans avoir à retravailler

tout le système. La solution que Rémy a trouvée est une architecture modulaire, dont le schéma est présenté en figure 9.8. L'idée est de séparer l'existence d'une notion de sa définition. Par exemple, la définition formelle de la robustesse fait intervenir la notion de distance, mais est totalement indépendante de la manière dont sont définies ces distances. On peut comparer cette architecture avec un programme informatique bien organisé où l'on peut ajuster une variable sur la ligne où elle est définie sans avoir à reporter la modification sur l'ensemble du programme.

Donnons une première approche informelle de ce schéma. Notre réflexion est fondée essentiellement sur la notion de distance (ce qu'on a vu dans la section 9.2). On utilise une distance structurelle et une distance dynamique. La distance structurelle permet de comparer la structure de deux systèmes et de formaliser la notion de mutation sur la structure. La distance dynamique permet de comparer deux comportements, afin de quantifier les expressions "conserver son comportement" et "changer de comportement". À partir de ces deux notions, on peut quantifier la robustesse.

Pour définir l'évolutivité, il faut d'abord comprendre ce que signifie "évoluer" dans l'espace des réseaux. Pour cela, il faut déterminer la structure de l'espace des réseaux induite par les distances pour savoir comment les sauts de réseau en réseau se font. On distingue donc deux cas possibles de structure de l'espace des réseaux : une forme de graphe et une forme d'espace métrique (voir section 9.2). On définit ensuite deux notions : le chemin d'évolution, qui est une trajectoire dans l'espace des réseaux, ainsi que le coût de cette trajectoire, qui représente la difficulté d'obtenir une telle trajectoire (notion variable qui dépendra du modèle). À partir de l'ensemble des trajectoires d'un réseau g vers un réseau g' , on peut attribuer une valeur à l'évolution $g \rightarrow g'$ indépendante de la trajectoire traduisant la difficulté d'atteindre g' depuis g .

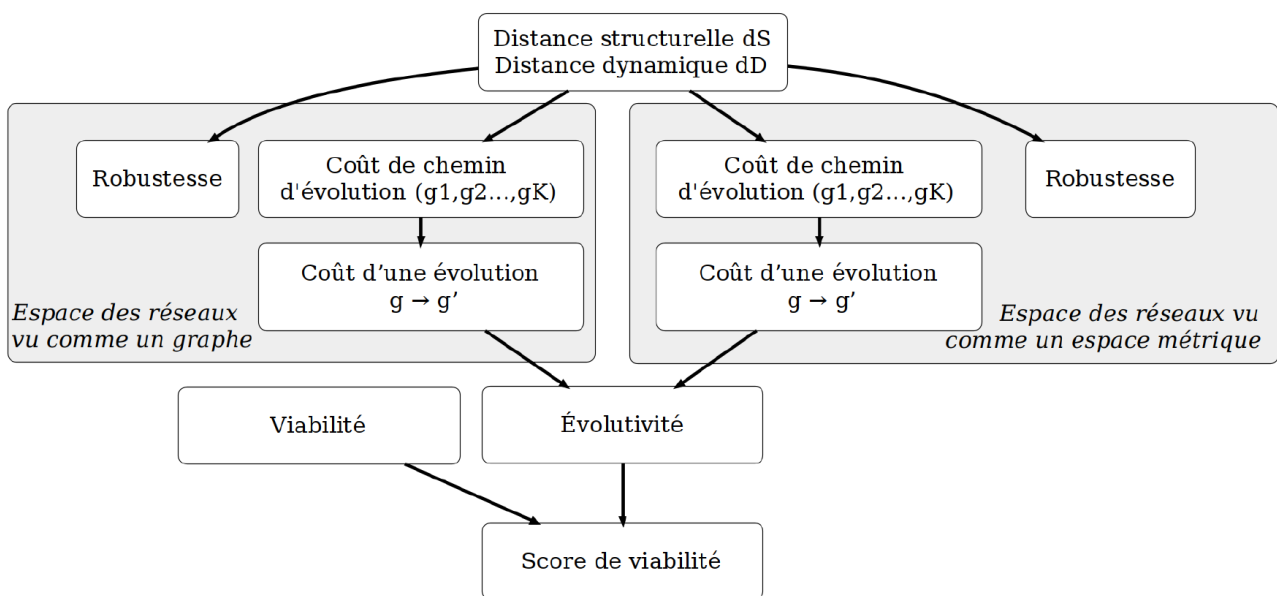


Figure 9.8. Organisation des définitions formelles autour de l'évolution. Les flèches entrantes d'un bloc indiquent quelles sont les notions supposées déjà définies utilisées dans la définition du bloc.

9.3.1. Robustesse

La robustesse d'un système peut être définie comme sa capacité à conserver son comportement lorsqu'il subit des mutations. On peut donc chercher à traduire cette notion par une fonction $R : \mathbb{G} \rightarrow [0; 1]$ avec $R(g) = 1$ dans le cas où toutes les mutations conservent le comportement de g . Dans le cas où l'espace des réseaux adopte une structure de graphe, une mutation se traduit formellement par la transformation d'un réseau en l'un de ses voisins. Pour la conservation du comportement (on note d_D la distance dynamique, donc sur le comportement), en l'absence de données supplémentaires sur le système biologique étudié, on fixe arbitrairement un seuil ε en-dessous duquel le comportement est considéré conservé. On a alors :

$$R_\varepsilon(g) = \frac{|\{g' \in \text{Voisins}(g); d_D(g, g') < \varepsilon\}|}{|\text{Voisins}(g)|}$$

Dans le cas où l'espace des réseaux a une structure d'espace métrique, il est nécessaire de considérer l'intégralité des autres réseaux comme résultant potentiellement de mutations. On pondère chacun des réseaux par sa distance au réseau étudié (d_S) et par la différence de comportement (distance de comportement d_D) pour obtenir la formule suivante :

$$R(g) = \frac{\sum_{g' \in \mathbb{G}} \left(e^{-d_S(g, g')} e^{-d_D(g, g')} \right)}{\sum_{g' \in \mathbb{G}} \left(e^{-d_S(g, g')} \right)}$$

On a bien ici $R(g) = 1$ dans le cas limite où tous les autres réseaux ont le même comportement que g .

9.3.2. Coût d'un chemin d'évolution, coût d'une évolution

Avant de définir l'évolutivité, il est nécessaire de définir deux notions intermédiaires : le coût d'un chemin d'évolution (dans l'espace des réseaux) et le coût d'une évolution (d'un réseau à un autre). On définit formellement un chemin d'évolution comme une suite finie de réseaux $(g_0; \dots; g_k)$, avec le réseau de départ g_0 et le réseau d'arrivée g_k . On définit une évolution comme un couple de réseaux $(g \rightarrow g')$, où g est le réseau de départ et g' le réseau d'arrivée. On désigne donc ici par un chemin d'évolution une trajectoire dans l'espace des réseaux, tandis qu'une évolution désigne une transformation d'un réseau en un autre, indépendamment du chemin choisi. Les notions de coût de chemin d'évolution et de coût d'évolution ont pour objet de traduire la difficulté ou la rareté d'une certaine évolution d'un système biologique.

On cherche donc à définir deux fonctions CC (coût d'un chemin) et CE (coût d'une évolution), l'une sur les chemins dans \mathbb{G} , l'autre sur \mathbb{G}^2 , à valeurs dans $\bar{\mathbb{R}}^*$. On doit avoir en outre $CE(g \rightarrow g') = \infty$ s'il n'y pas de chemin de g à g' .

À nouveau, ces notions ont des définitions très différentes suivant si l'espace des réseaux a une

structure de graphe ou d'espace métrique et il convient de distinguer les deux cas.

Structure de graphe : La notion de coût d'une transformation est complexe et intimement liée au système étudié. C'est pourquoi il vaut mieux sans doute faire le choix de la définition lors du choix du système étudié. Je donnerai cependant ici une définition simple à titre d'exemple du coût d'un chemin d'évolution et d'une évolution.

Dans une structure de graphe, on se restreint aux chemins de voisins en voisins. Un chemin d'évolution est donc une suite finie de réseaux telle que deux réseaux consécutifs sont voisins. Définissons par exemple le coût d'un chemin d'évolution comme étant la longueur du chemin :

$$CC(g_0; \dots; g_k) = k$$

On peut alors définir le coût d'une évolution comme étant le coût minimal d'un chemin allant du réseau de départ au réseau d'arrivée :

$$CE(g \rightarrow g') = \begin{cases} \min_{c \in \text{paths}(g \rightarrow g')} CC(c) & \text{si } \text{paths}(g \rightarrow g') \neq \emptyset \\ +\infty & \text{sinon} \end{cases}$$

Il est important de noter que CC et CE sont deux notions bien distinctes à ne pas fusionner pour garantir la flexibilité du système formel élaboré. La définition du coût de chemin d'évolution est sujette à beaucoup de variations suivant le système étudié comme expliqué plus haut, tandis que la définition du coût d'évolution fait déjà intervenir le coût des chemins et donc peut être commune à de nombreux systèmes biologiques.

Structure d'espace métrique : Comme dit précédemment, la structure d'espace métrique est bien plus compliquée à appréhender que celle de graphe. Pour l'instant, ce travail de formalisation n'est pas encore fait, mais une piste de travail consisterait à étudier des marches aléatoires sur l'espace des réseaux, en étudiant d'abord la structure de graphe et en adaptant le modèle obtenu sur la structure d'espace métrique.

9.3.3. Évolutivité

L'évolutivité est définie comme la capacité d'un système biologique à acquérir de nouveaux comportements avec de petites évolutions. Pour calculer l'évolutivité d'un système, on fera donc un calcul sur l'ensemble des réseaux. Les évolutions faisant varier le comportement doivent être mises en valeur, tandis que les évolutions coûteuses seront dévalorisées. Rémy est arrivé ainsi à la formule suivante :

$$E(g) = \text{moyenne}_{g' \in \mathbb{G}} \left(\frac{d_D(g, g')}{CE(g, g')} \right)$$

où *moyenne* désigne ici (et par la suite également) la moyenne arithmétique. On note ici

l'importance d'avoir un coût d'évolution de g vers g' non nul ($CE > 0$). Notons que sont prises en compte les évolutions impossibles dans le calcul. Comme on a $CE(g \rightarrow g') = \infty$, le terme correspondant dans le calcul de la moyenne est nul.

9.3.4. Viabilité

La contrainte de viabilité de l'environnement détermine la survie des systèmes biologiques. Il y a deux manières de présenter cette contrainte. On peut choisir une viabilité binaire, pour laquelle un système survit ou ne survit pas (voir figure 9.9). On peut choisir d'être plus graduel avec une viabilité pour laquelle un système a plus ou moins de chances de survie. Ce cas correspond à un animal normal et un autre amputé d'une patte. Les deux animaux peuvent vivre et se reproduire, mais le deuxième a toutefois moins de chances de survie car il y a une probabilité importante pour qu'il finisse par se retrouver dans une situation où sa patte en moins lui sera fatale. Pour des gènes, on peut visualiser des différences d'accès au gène pour les polymérases, de niveau de régulation, *etc.* Ici à nouveau, on va donner des définitions dans les deux cas, car le premier cas présente l'avantage d'être plus simple à appréhender pour une première étude, mais le deuxième cas permet une plus grande complexité pour représenter plus fidèlement certains modèles de systèmes naturels.

On appellera viabilité binaire toute fonction de \mathbb{G} dans $\{0, 1\}$ et viabilité graduelle toute fonction de \mathbb{G} dans $[0; 1]$. On appellera viabilité parfaite la fonction V_P constante égale à 1 et viabilité nulle la fonction V_N constante égale à 0.

Rajouter une contrainte de viabilité modifie les coûts des chemins d'évolution CC , les coûts d'évolution CE et l'évolutivité E . On notera respectivement CC_V , CE_V et E_V ces quantités modifiées où V est la viabilité.

Cette modification est bien entendu différente suivant le modèle. Or, les définitions données jusqu'ici ne supposent pas l'existence d'une contrainte de viabilité, ce qui revient à se placer dans le cas d'une viabilité parfaite. On voudrait donc avoir les propriétés suivantes :

- Les définitions précédentes correspondent à une viabilité parfaite, donc $CC_{V_P} = CC$, $CE_{V_P} = CE$, $E_{V_P} = E$.
- Rajouter une contrainte de viabilité ne peut qu'augmenter les coûts et entraver l'évolution, donc on a pour toute viabilité V : $CC \leq CC_V$, $CE \leq CE_V$ et $E \geq E_V$.
- Dans le cas d'une viabilité nulle, il n'y a aucune évolution possible, donc : $CC_{V_N} = CE_{V_N} = \infty$ et $E_{V_N} = 0$

Nos nouvelles définitions devront donc remplir ces critères. Je donne ici un exemple d'une modification prenant en compte la viabilité (en se plaçant dans une structure de graphe). Les définitions qui suivent sont valables pour une viabilité binaire aussi bien qu'une viabilité graduelle :

$$CC_V(g_0; \dots; g_k) = \frac{k}{\prod_{i=0}^{k-1} V(g_i)}$$

$$CE_V(g \rightarrow g') = \min_{c \in \text{paths}(g \rightarrow g')} CC_V(c)$$

$$E_V(g) = \text{moyenne}_{g' \in \mathbb{G}} \left(\frac{d_D(g, g') \times V(g')}{CE_V(g \rightarrow g')} \right)$$

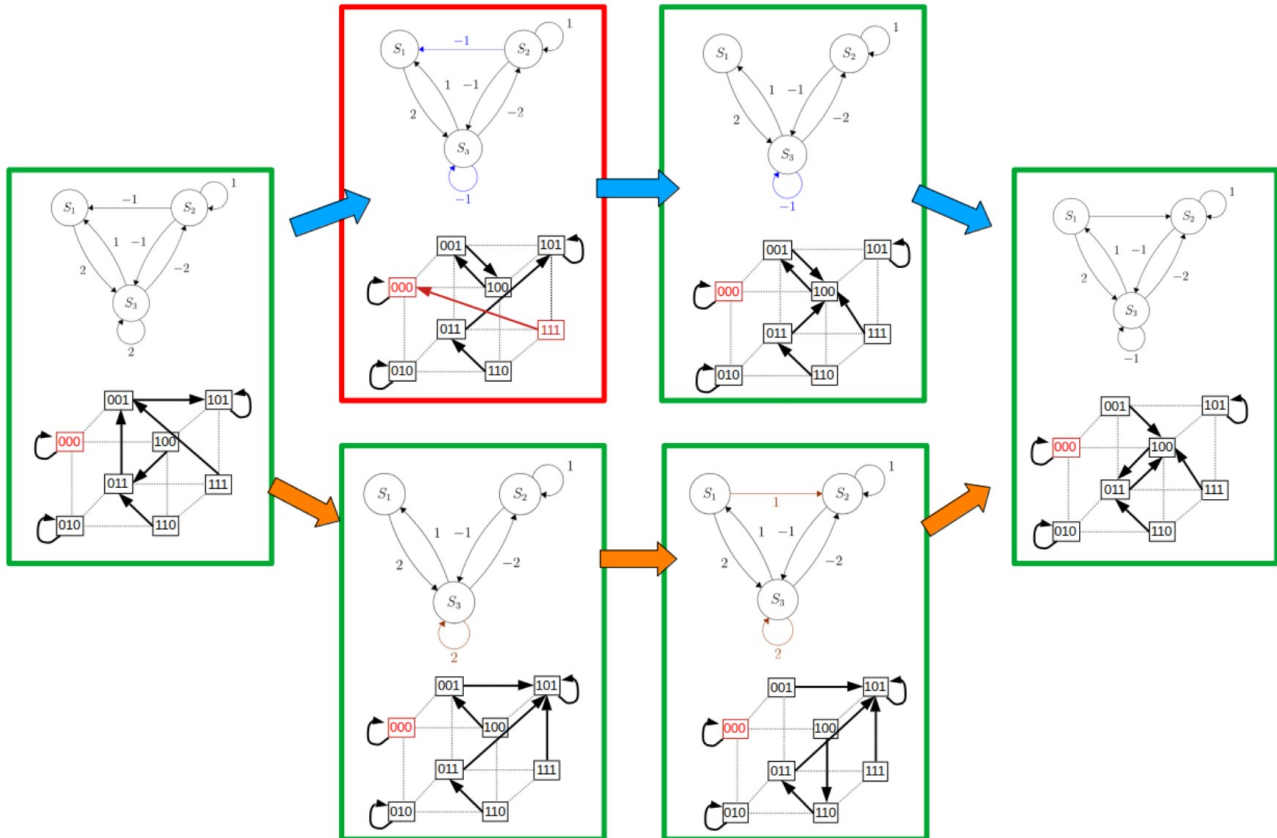


Figure 9.9. Deux exemples de chemins dans l'espace des réseaux. On considère ici comme voisins deux SBNs qui ne diffèrent que d'un arc et comme viable un SBN dont aucun état non nul n'aboutit à l'état nul. Le chemin orange est viable, le chemin bleu ne l'est pas.

Elles remplissent bien les contraintes demandées plus haut. En outre, l'esprit de la définition de l'évolutivité est préservé : un système est évolutif si des mutations dans sa structure peuvent induire un caractère héritable [Wagner, 2008]. Un trait "héritable" est un trait qui fait augmenter les chances de survie du système pour une certaine contrainte de viabilité. On ne peut le définir simplement dans le système formel mais on constate que nos définitions gardent cet esprit : les évolutions vers des systèmes plus viables, donc qui offrent des caractères héritaibles, sont valorisées.

On peut aussi définir le score de viabilité Sc . On cherche ici à définir une fonction sur l'ensemble des viabilités qui attribue un score élevé à une viabilité binaire si elle ne contraint que peu l'évolution des systèmes, et un score faible si de nombreuses évolutions sont impossibles. Dans le cas d'une viabilité graduelle, on aimerait que les chemins maximisant la viabilité de la trajectoire ne fassent que peu de détours par rapport à la viabilité parfaite (voir figure 9.10). Les contraintes à

remplir sont donc les suivantes :

- S_c est une fonction de l'ensemble des viabilités dans $[0; 1]$.
- La viabilité parfaite a un score de 1 et la viabilité nulle a un score de 0.

On a déjà défini avec E_V comment une viabilité contraignait l'évolution ; il ne nous reste plus qu'à trouver une formule synthétisant ces valeurs. Dans notre description du score de viabilité, on compare les viabilités au cas idéal. Notre formule sera probablement un quotient avec le cas réel au numérateur et le cas idéal au dénominateur. Enfin, pour intégrer sur l'espace des réseaux, nous choisirons une moyenne. Cela nous donne la formule suivante :

$$S_c(V) = \frac{\text{moyenne}_{g \in \mathbb{G}} E_V(g)}{\text{moyenne}_{g \in \mathbb{G}} E(g)}$$

De la viabilité et de l'évolutivité (la robustesse constituant la force de rappel s'opposant à l'évolutivité), découlent un problème d'atteignabilité. C'est ce dont nous allons commencer à discuter.

9.4. Contrôle des réseaux

Nous allons revenir un instant sur les réseaux pluripotents et parler du contrôle externe qui permet, par des variations d'états de bits de contrôle, de passer d'une hyperface à une autre, autrement dit de changer de comportement comme on changerait de réseau dans l'espace des réseaux précité (voir les deux sections précédentes 9.2 et 9.3). Je vais donc élargir la discussion sur ce contrôle à l'espace des réseaux et introduire le contrôle comme ce qui permet d'éventuellement guider une trajectoire dans cet espace.

Ce dont je vais discuter ici, c'est d'une part la possible nature de tels systèmes de contrôle et, d'autre part, de leur émergence.

On cherche dans le vivant des éléments qui, agissant sur un système assimilable à un réseau, le font basculer vers un tout autre fonctionnement et comportement associé. Dit ainsi, on voit bien que la plupart des systèmes biologiques répondent à cette présentation. C'est lorsqu'on les assimile à des réseaux formels et qu'on en fait la modélisation que nous faisons revêtir à de tels réseaux l'habillage pluripotent.

Je pense donc qu'il y a de nombreuses façons de concevoir de tels systèmes de contrôle dans les systèmes naturels, mais il y en a une que je trouve particulièrement intéressante car se comportant intrinsèquement comme un *switch* : les IDPs ! Nous avons déjà parlé de ces *intrinsically disordered proteins*. Ce que je vais écrire rejoint parfaitement l'excellent article éditorial de P. Kulkarni, intitulé *Intrinsically Disordered Proteins : Insights from Poincaré, Waddington and Lamarck*, publié dans *Biomolécules* en 2020 [Kulkarni, 2020], article présentant une rétrospective des avancées faites sur ces molécules biologiques qui constituent notamment 80 % des protéines associées au cancer et de nombreux facteurs de transcription et dont P. Kulkarni fut un des pionniers. Ces protéines intrinsèquement désordonnées sont capables d'adopter dynamiquement de nombreuses

conformations, des centaines parfois comme c'est le cas pour la fameuse *p53*. Elles adoptent dynamiquement des conformations à mesure que se trouvent dans leur voisinage immédiat telle ou telle autre biomolécule, en particulier leurs ligands potentiels. Le schéma d'interaction classique est brisé : ni la clef ni la serrure n'existent préalablement ; la conformation de l'IDP (et même éventuellement de son ligand) est induite par la proximité des molécules biologiques entrant en interaction.

Les IDPs se comportent donc comme des hubs moléculaires adoptant à la volée telle ou telle conformation éphémère selon l'environnement moléculaire qui sollicite leurs structures primaire et secondaire.

Premier point important : la fonction n'est plus portée par la protéine ! Elle n'est qu'une potentialité. La fonction émerge du couple IDP-ligand.

Deuxièmement, un paysage de ligands (en quantités variables dans le temps selon la vie cellulaire) tirera (au sens des créodes de Waddington) la conformation d'une IDP dans des directions particulières. C'est une configuration composition-dépendante.

Les IDP jouant pour beaucoup le rôle de facteurs de transcription, on comprend alors que, dans le modèle des réseaux pluripotents que je donne, ces switches peuvent jouer le rôle assuré par les bits de contrôle. P. Kulkarni ne s'y est pas trompé car il a effectivement proposé un modèle de réseau dans lequel les IDP jouent un tel rôle !

Question : dans le vivant, lorsque la composition du milieu cellulaire pousse l'IDP dans certaines conformations qui provoquent un changement d'activité des réseaux associés, est ce que les transitions entre les comportements sont proches ? Existe-t-il une gradation des changements de comportements associée aux séquences classiques de changements de milieux cellulaires (un petit changement de la composition du milieu induit un changement d'activité graduel) ?

Un ingénieur, s'il devait concevoir un tel système, ferait en sorte que les configurations du contrôle soient ordonnées d'une manière « intelligente », c'est-à-dire telles qu'un changement d'un bit de contrôle ne provoque pas une catastrophe, un changement drastique d'activité. En tout cas, il essaierait d'exploiter la fréquence (connue) de telle ou telle action extérieure devant affecter le système, dans cet encodage du contrôle pour y introduire une forme de robustesse. Traduit dans nos réseaux, ceci devient : on cherche un encodage « intelligent » des bits de contrôle tel que la ou les succession(s) « attendue(s) » de ces bits de contrôle, en s'éloignant « peu à peu » de leur configuration initiale, sélectionne des graphes de transition changeant progressivement (et non brutalement). Si j'ai mis entre guillemets les termes « intelligent », « attendu » et « peu à peu », c'est qu'ils sont importants. L'intelligence d'encodage en particulier est liée au système qui l'exploite. S'il s'agit d'un ingénieur, on pourrait imaginer qu'un encodage particulier permette par petits changements des bits de contrôle (là, on parle du « peu à peu ») de changer d'hyperface de telle manière que les graphes de transition simulés changent aussi « peu à peu ». Un tel encodage, choisi par un ingénieur, serait par exemple binaire (ex : binaire avec 3 bits de contrôle : 000→001→010→011→100→101→110→111) ou, mieux, suivrait un code de Gray (ex : code de Gray avec 3 bits de contrôle : 000→001→011→010→110→111→101→100). Étant donné une distance de

Hamming, le code de Gray est plus progressif car un seul bit est changé à chaque fois (ce n'est pas le cas pour le code binaire qui demande souvent plusieurs changements de bits d'un coup). Mais cette progressivité ne nous paraît intéressante que parce que notre façon de voir les choses implique plus facilement de ne considérer qu'un seul changement à la fois. Ces considérations sur l'encodage n'ont pas pour but de déterminer lequel est le meilleur ; il s'agit plutôt de signifier que ce problème, hors contexte, n'a aucun sens. Le contexte extérieur, c'est ce qui provoque les changements de configuration du contrôle (vu autrement, le contexte extérieur est le contrôle externe). Si nous choisissons intuitivement un encodage avec des changements progressifs de bits (suivant par exemple un code de Gray), c'est parce qu'il nous paraît (subjectivement) plus simple de passer de 00000 à 00001 que directement à 10110. Le vivant ou tout autre système (ex : un rover martien) rencontre des situations successives, voisines dans un espace de probabilités d'apparition. Elles sont voisines dans le vivant car la cellule ne peut que rarement changer brutalement ses concentrations internes, bien que cela puisse tout de même se produire, par exemple des entrées ou sorties massives d'ions dans les cellules neurales. Si l'on fait l'hypothèse que les variations de quantités de composants cellulaires (ions, facteurs de transcription ...) sont graduelles dans une direction ou dans une autre, autrement dit si l'on parcourt un paysage morphogénétique (paysage des facteurs de transcription, des ligands des IDPs, *etc*) dont peu de frontières entre les vallées sont abruptes, alors il envisageable de considérer ce paysage comme un encodage particulier de telle manière que se déplacer dedans revient le plus souvent à faire changer les comportements associés de façon progressive pour la cellule (c'est-à-dire que ses états morphologique, métabolique, *etc*, se ressemblent immédiatement avant et après). Le terme « attendu » a sa place dans le texte maintenant ! Par « attendu » j'entends « qui ne provoque pas de catastrophe ne permettant pas la survie (non viable) ».

La ramification des successions de changements de contrôle possibles (en supposant les changements trop importants impossibles ou en tout cas durs à obtenir) doit permettre d'atteindre les comportements souhaitables.

9.5. Atteignabilité – Rôle de l'évolution

J'ai principalement présenté la viabilité dans le contexte du paysage des réseaux, mais on peut tout aussi bien la considérer dans le cadre plus restreint des réseaux pluripotents. Je vais commencer cette section sur l'atteignabilité des comportements par ceux-ci.

Dans ces réseaux pluripotents, le passage d'un comportement à un autre est contrôlé par un système externe qui configure les bits de contrôle pour passer d'une hyperface à l'autre. Ce contrôle permet éventuellement, en ressentant les changements de l'environnement, d'adapter les comportements du système pluripotent. Maintenant que nous avons vu que la viabilité imposée au paysage des réseaux peut interdire certaines trajectoires, les comportements associés ne permettant pas au système qui en dépend de vivre, nous pouvons sans peine transposer cela de façon équivalente aux réseaux pluripotents et aux comportements qu'ils simulent. Pour atteindre un graphe de transition en partant

d'un autre graphe de transition, tout en conservant le système viable en n'autorisant que de petites transitions (de petits changements entre les comportements) par un contrôle externe faisant progressivement changer les bits de contrôle selon une séquence particulière, il est possible que le système doive traverser une région non viable (ou de faible viabilité), une région dans laquelle l'environnement n'est pas favorable à la survie du système qui adopte un tel comportement transitoire. On est là dans un problème d'atteignabilité : le nouveau comportement n'est pas atteignable selon la séquence de contrôle utilisée (mais il pourrait éventuellement l'être selon une autre séquence de contrôle nécessitant éventuellement des changements plus drastiques en termes de comportements transitoires).

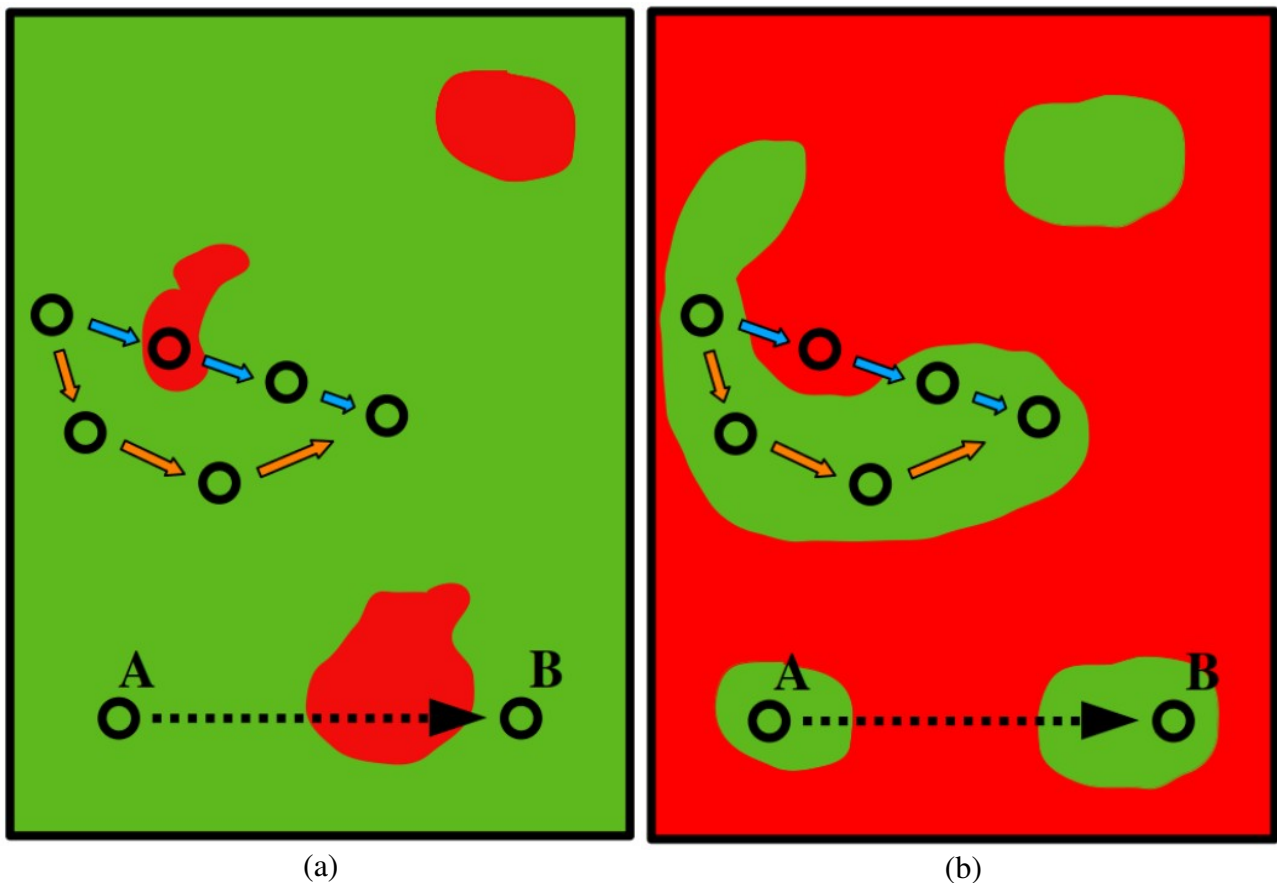


Figure 9.10. Représentation de l'espace des réseaux avec deux viabilités différentes. On cherche à construire un score de viabilité qui attribue une valeur plus élevée à la viabilité de gauche. (a) Environnement peu contraignant : score de viabilité fort ; tous les réseaux viables sont atteignables. (b) Environnement très contraignant : score de viabilité faible ; tous les réseaux viables ne sont pas atteignables.

Ce dont nous venons de discuter, plaçons le maintenant dans un contexte évolutif !

La mise en correspondance, « intelligente » au sens « d'adaptée à sa vie menée au gré des variations environnementales qu'elle rencontre », des changements des états cellulaires (quantités de substances, morphologie, *etc*) avec les changements de mise en œuvre de ses systèmes est le résultat

d'une « optimisation » faite par l'évolution, au même titre qu'elle a œuvré dans l'optimisation d'un code génétique (la mise en correspondance entre des triplets de bases azotées et des acides aminés aux propriétés particulières) qui présente de nombreuses robustesses. L'évolution a œuvré durant les âges changeants de la vie biologique sur Terre. Elle a sélectionné les meilleures séquences de contrôle, les plus adaptées à l'ensemble des variations physiologiques comme pathologiques ou accidentelles qu'elle a rencontré. C'est en soit une forme d'apprentissage et il ne m'apparaît donc pas insensé de penser que l'évolution puisse avoir inscrit l'histoire des changements conjoints de contrôle externe et de fonctionnement dans un tel encodage, et qu'une partie au moins puisse avoir été conservée. Nous en reparlerons quand nous aborderons l'atavisme, puisque c'en est une forme.

La question de l'extensibilité des réseaux, traitée en section 8.3.5, soulève un point intéressant dont je n'ai pas parlé jusque là (voir aussi figure 9.11) : est ce que tous les réseaux R de dimension $d+1$ peuvent être atteints par simple extension, c'est-à-dire en ajoutant un nœud au réseau, en conservant la structure de R_l et en conservant M_l dans M , une musique étendue ? Autrement dit, existe-t-il des réseaux de dimension $d+1$ ne pouvant être atteints par extension des réseaux de dimension d , avec conservation de leur comportement ? Il est envisageable que les réseaux R atteints par extension ne soient qu'un sous-ensemble des réseaux de dimension $d+1$. La réponse ne nous apparaît pas encore clairement au moment où j'écris ces lignes ; il nous appartient de déterminer cela dans les années à venir et, si les réseaux R obtenus par extension sont une inclusion stricte dans l'ensemble des réseaux de dimension $d+1$, alors on pourra chercher aussi à évaluer la proportion de tels sous-ensembles dans les réseaux d'une dimension donnée lors d'une extension.

Si la réponse s'avérait négative, ceci aurait des conséquences concernant l'atteignabilité de réseaux et de comportements associés lors d'une évolution par croissance : certaines solutions (des réseaux réalisant une fonction biologique) ne pourraient être jamais atteintes sans l'aide de mutations ponctuelles. Ces solutions constitueraient des jardins d'Eden⁵⁰ (configurations ne possédant aucun prédécesseur dans le cadre d'une évolution déterministe, donc ne pouvant être atteintes de façon déterministe). Partant d'un réseau R_l initial, l'ensemble des extensions vers des réseaux $d+k$ successifs, croissant progressivement par extension, dessine une ramification dans l'espace des réseaux. Les mutations ponctuelles à partir d'un réseau R de dimension $d+k$ obtenu par extension permettraient par contre de s'éloigner de la ramification atteignable à partir des extensions successives possibles partant d'un réseau R_l donné, et d'atteindre d'autres réseaux R' qui ne font pas partie de cette ramification. Certains de ces réseaux R' pourraient appartenir à d'autres ramifications produites par l'extension d'un réseau R_2 initial différent de R_l , pouvant éventuellement afficher le même comportement M_l et même M , ou un comportement différent. Certains pourraient par contre appartenir au sous-ensemble des réseaux de dimension $d+1$ ne pouvant jamais être atteints par

⁵⁰ Les *Jardins d'Eden* font référence au lieu biblique (non atteignable par les humains) et ont été introduits dans le cadre des automates cellulaires par J. Tukey dans les années 50. Ce sont, dans les automates cellulaires déterministes, des motifs finis qui ne possèdent aucun prédécesseur. Leur configuration ne peut être atteinte de quelque configuration antérieure que ce soit. La seule façon d'atteindre un jardin d'Eden est qu'un système extérieur impose des changements de configuration qui ne sont pas intrinsèques au système, que ces changements imposés soient fondés sur des comportements déterministes (contrôle déterministe) ou sur des apports aléatoires comme des fluctuations thermiques.

extension. Dans les deux cas, dans un espace de *proximité structure-fonction* des réseaux, ces réseaux obtenus par mutation ponctuelle pourraient effectivement être proches y compris en comportement.

L'atteignabilité d'un comportement est donc limitée par la capacité exploratoire des réseaux (en termes d'évolution ou de changements induits par un contrôle) et par la viabilité (pour atteindre il faut parfois traverser des zones non viables). Là, on a vu que l'évolution a pu jouer un rôle en sélectionnant non seulement les configurations permettant les changements les plus progressifs, mais aussi ceux qui maximisent l'évolutivité et la fonctionnalité en s'adaptant à la viabilité imposée.

Connaître la façon dont on évolue dans le paysage morphogénétique (réseaux et graphes de transition) en fonction des variations de l'environnement (contrôle externe) et la dynamique décrivant ces changements du contrôle (donc connaître l'encodage) présente un intérêt majeur. En effet, si en tant qu'observateur extérieur, je sais dans quel état se trouve un système et la configuration de ses bits de contrôle, connaître l'encodage permet de diriger le système dans un nouvel état à atteindre. Avec des réseaux, si je sais que le système de contrôle est ordonné d'une certaine manière, telle que les graphes de transition simulés sont ordonnés par proximité étant donné une distance connue (autrement dit, le système de contrôle permet de naviguer de 'proche' en 'proche' dans un paysage des graphes de transitions ordonné) alors, si je suis dans un graphe de transition donné, je dispose d'informations sur comment me diriger vers un autre graphe de transition particulier ; je n'y vais plus au hasard.

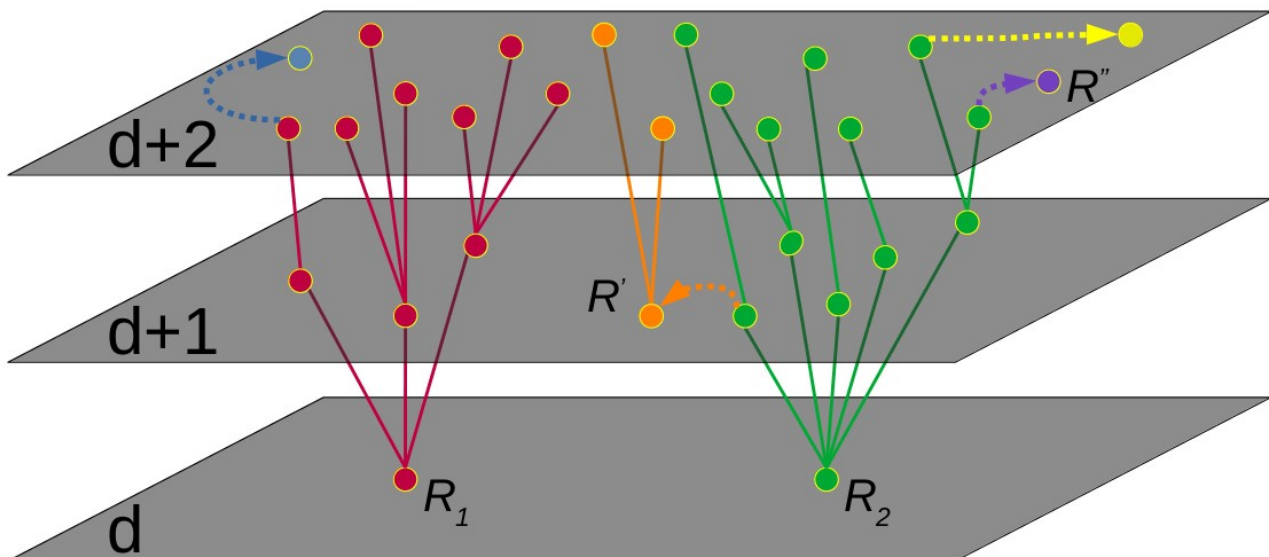


Figure 9.11. Représentation des extensions de réseaux dans l'espace des réseaux de dimension croissante. Partant de réseaux R_1 ou R_2 de dimension d , on peut atteindre par extension (au sens défini en section 8.3.5 notamment, donc avec conservation de la structure et d'une partie de la dynamique au moins) toute une arborescence correspondante (en rouge ou en vert) de réseaux étendus de dimensions supérieures $d+k$. Certains réseaux comme R' ou R'' identifiés en orange et violet dans la figure n'ont pas de prédécesseurs par extension dans la dimension inférieure. Du point de vue de la dimension qui leur est inférieure, ils sont des jardins d'Eden qui ne peuvent être atteints que par mutation depuis un réseau n'ayant pas la même structure ancestrale (flèches pointillées) mais de structure et de graphe de transition voisins.

Peut-être que réfléchir à ceci pourrait aider à faire sortir des systèmes d'états dits « pathologiques » pour les aider à se diriger progressivement vers des états souhaités. C'est peut-être un vœu pieux avec une idée devant rester à un stade théorique. Dans ce cas, son application peut probablement être considérée pour la conception de systèmes artificiels fondés sur des réseaux et auto-adaptables, par exemple dans le cadre des réseaux de neurones.

C'est en tout cas un travail exploratoire théorique possible, travail qui met en œuvre de nombreux aspects que nous avons évoqués dans le cadre des réseaux Booléens.

9.6. Evolution : héritage, offuscation et atavisme

9.6.1. La vie atavique

La cellule actuelle, quel que soit son type, et même son règne, est un long héritage. Les conditions environnementales ont permis de sélectionner des stratégies de survie différentes selon les règnes du vivant, les types cellulaires et, plus finement, selon les espèces. Les procaryotes misent beaucoup sur la pluralité et la diversification. Les eucaryotes, eux, sont davantage des êtres intrinsèquement adaptables (leurs systèmes internes, leur métabolisme, leur physiologie s'adaptent à de nombreuses conditions environnementales) et misent sur plusieurs niveaux d'organisation (de la cellule au tissu, du tissu à l'organe, de l'organe à l'individu, de l'individu à la société).

Quand on regarde de près, on peut facilement voir les eucaryotes pluricellulaires comme de véritables collections de systèmes génétiques et métaboliques, de couches anatomiques de structures organiques (ex : structure du cerveau), certains complexes, d'autres très simples, certains localisés, d'autres distribués, certains à petite et d'autre à grande échelle, mis en œuvre dans de mêmes finalités biologiques, sans qu'il s'agisse de redondance.

Certains systèmes et organes se sont fortement interconnectés jusqu'à confondre leurs parties au niveau génétique ; d'autres ont pu s'accroître par modularité avec une conservation assez forte des modules initiaux. Les façons dont a pu se faire l'intégration de ces systèmes est variée (sous forme de modules bien disjoints, au contraire sous forme de système intriqués, *etc*) et répond vraisemblablement à un impératif de conservation de fonction (robustesse évolutive) modulée par la nécessité d'adaptation aux changements environnementaux (environnement physico-chimique d'une part, écosystémique d'autre part).

Tous ces systèmes (génétiques, métaboliques, organiques) ne sont pas toujours actifs en même temps, certains étant réprimés en situation physiologique, d'autres l'étant en situation de stress, ceci impliquant une intégration à haut niveau dans les systèmes de régulation de l'organisme. Si ces systèmes se sont maintenus malgré la pression évolutive, c'est que leur présence, même peu sollicitée, procure un avantage évolutif à l'espèce qui en dispose.

Certains systèmes sont vraisemblablement des systèmes apparus il y a longtemps ; d'autres sont bien plus récents. Les reliques du passé dont le vivant hérite ont cependant évolué, comme tous les systèmes progressivement apparus. En cela, les différents systèmes et organes d'un organisme sont

tous modernes (au sens de S.G. Gould), de même que le sont les espèces vivantes dans un écosystème. Il n'y a pas d'espèce fossile ; il n'y a que des espèces présentant ou non certains caractères identifiés comme archaïques par leur simplicité et par phylogénie. C'est la même chose pour les systèmes biologiques. Dans les cellules actuelles, ils ont tous le même niveau d'évolution. En revanche, il est exact de dire que certains ont été acquis (ou sont apparus) plus tôt que d'autres dans une lignée évolutive. Cette distinction est importante car il ne faut pas voir les systèmes dits « anciens » comme des reliques qu'un système vivant se traîne depuis la nuit des temps comme un boulet. Non, si de tels systèmes *sont présents* (et non « *persistent* » comme j'ai failli l'écrire) c'est parce que, comme je l'ai écrit au dessus, ils procurent un avantage évolutif, et ils sont bien intégrés à l'ensemble des systèmes d'un organisme grâce à cette co-évolution qui a finement co-ajusté ces systèmes et ajusté leur dialogue.

Quand j'étais à l'université, à la faculté de biochimie de Marseille, j'avais appris que si je courais ventre à terre, je disposais d'environ 4 secondes d'ATP en réserve dans mes muscles. Je n'ai jamais vérifié expérimentalement cet enseignement. Étant toutefois très mauvais en course de vitesse, je dois pouvoir frôler les 14 secondes aux 100 m, soit 10 secondes de trop pour mes maigres réserves d'énergie chimique directement exploitables, dans le meilleur des cas. Même Usain Bolt ne s'en sortira pas beaucoup mieux. Nous savons tous que chez les eucaryotes, l'ATP, adénosine triphosphate, est le produit principalement issu de la respiration cellulaire oxygénée qui se fait dans les mitochondries. L'oxydation complète d'une molécule de glucose produit un total de 38 molécules d'ATP dont 2 proviennent de la première étape, la glycolyse, 2 proviennent de l'étape intermédiaire, le cycle de Krebs, et 34 sont produites par la réoxydation des transporteurs mitochondriaux réduits, c'est-à-dire en bout de chaîne respiratoire par les sphères pédonculées, les ATP synthases. Le rendement de cette conversion est très important, de l'ordre de 40 %⁵¹. Mais pour que cette chaîne respiratoire puisse fonctionner, il faut que les cellules disposent en quantités suffisantes de ce poison cellulaire que les algues et les cyanobactéries (puis les végétaux) ont produit massivement, comme un déchet, il y a 2.4 milliards d'années et que d'autres organismes, dits aérobies, ont pu traiter avec *l'invention* de la respiration : l'oxygène moléculaire O₂. Si cet accepteur final de la chaîne d'oxydoreduction respiratoire n'est pas disponible en quantités suffisantes, ce qui est le cas lors d'un effort violent et prolongé, la chaîne respiratoire n'est pas assez efficace pour produire de l'ATP. Fort heureusement, nous pouvons toujours avoir recours à la fermentation lactique qui fait le bonheur des masseurs puisque sa production, l'acide lactique, accompagne et maintient la formation des crampes (une tétanie des fibres musculaires, donc des moteurs actine-myosine, due à une énergie chimique, l'ATP, en quantités insuffisantes). La fermentation lactique, mécanisme infiniment plus simple que la respiration aérobie, est aussi très peu efficace (rendement de 2%). Pourquoi donc, alors que tout eucaryotes que nous sommes, nous disposons depuis longtemps de mitochondries, avons nous conservé ce mécanisme simple et si peu efficace, la fermentation lactique, apparu il y a longtemps chez des micro-organismes bactériens ? Pourquoi l'évolution l'a-t-elle retenu, y compris chez les animaux, organismes gourmands en énergie ? Il y a

51 Rendement = 100x(Énergie récupérée sous forme utile)/(Énergie disponible au départ) = 100 x (31 x 38) / 2860, avec Énergie potentielle d'un ATP = 31kJ et Énergie potentielle du glucose = 2860 kJ.

certainement plusieurs raisons, mais en voici une qui me paraît évidente : maintenir un effort soutenu et prolongé (comme fuir longtemps devant un prédateur) est un avantage évolutif qui, à lui seul, justifie la conservation d'une telle fonctionnalité. Il y en a d'autres : les cellules peuvent dans certaines situations vivre des périodes hypoxiques, notamment en cas de mauvaise perfusion suite à un accident vasculaire, par exemple dans le cerveau. L'augmentation de temps de survie par un tel mécanisme est bien entendue vitale. C'est aussi le cas dans les tumeurs dans lesquelles certaines régions sont mal ou non perfusées ; dans ce cas, les cellules hypoxiques entrent en survie via ces voies ataviques [Jacquet, 2021]. Évidemment, avec ce dernier exemple, on se demande s'il est judicieux de parler d'avantage évolutif et de tumeurs dans la même phrase. Nous verrons peu après que la frontière entre ces deux états, l'état physiologique et l'état cancéreux, est finalement d'une part progressive (on ne passe pas brutalement d'un état dit « physiologique » à un état dit « cancéreux » ; il vaut mieux dire que des cellules peuvent s'éloigner d'un état physiologique et rentrer peu à peu dans une voie différente, une voie de développement tumoral), d'autre part très poreuse car faisant intervenir tout un gradient de mécanismes ataviques profitables à ces deux états que peuvent finalement adopter les cellules qui nous constituent.

Cet exemple très imagé de la fermentation lactique, illustre la notion d'atavisme : en temps normal la production d'ATP par la chaîne respiratoire est prépondérante, mais en cas de disette (en oxygène), nous avons toujours la possibilité d'utiliser un mécanisme plus rustique, hérité il y a plus longtemps au cours de l'évolution. Ce mécanisme est parfaitement intégré à nos systèmes actuels, subissant ses régulations, communiquant avec le reste du métabolisme, et il a coévolué avec le reste. On peut le qualifier d'ataviste : c'est un système d'origine ancienne qui se manifeste quand des systèmes plus récents sont inopérants.

Je ferais aussi remarquer que si l'oxygène est un poison pour les organismes anaérobies, il reste aussi un poison pour les cellules eucaryotes qui disposent cependant de bon nombre de systèmes biochimiques capables de le prendre en charge : chaîne respiratoire, balance glutathion, métabolisme du fer : tous ces mécanismes sont efficaces pour éliminer toutes les formes que l'oxygène prend dans l'organisme, les ROS⁵². Certains sont très complexes, d'autres finalement simples (ex : oxydation du fer), parfois déplacent le problème (le tampon que constitue l'oxydation du fer passe par une accumulation de fer dans les cellules, accumulation qui mène à leur vieillissement). Les cellules cancéreuses, elles, sont particulièrement empoisonnées par la présence d'oxygène et en conséquence régulent négativement la vascularisation dans les tumeurs [Muz, 2015]. Ce phénomène a de multiples conséquences, comprenant une résistance aux thérapies, les cellules entrant en quiescence, et des changements phénotypiques avec des transitions de types épithéliaux à des types mésenchymateux conduisant à une augmentation de la motilité cellulaire et, à terme, à des comportements métastatiques.

Il y a beaucoup d'autres systèmes qu'on pourrait qualifier d'atavistes.

Le système immunitaire par exemple est composé de nombreux systèmes procurant leur niveau de défense propre et partiellement interconnectés entre eux : de façon extrêmement simplifiée, il y a

52 Reactive Oxygen Species

ainsi une défense innée, qu'on retrouve dans tout le règne végétal et animal, et une défense adaptative qu'on ne retrouve que dans le monde animal chez les vertébrés à mâchoire (branche des gnathostomes ; 500 millions d'années). La défense innée est faite d'une part d'une défense chimique dont l'effet est d'interagir directement et très en amont de la réponse immunitaire spécifique avec les pathogènes (ex : peptides anti-microbiens, dont il existe plusieurs dizaines de milliers de types différents), de développer et contrôler l'inflammation (ensemble des protéines de la phase aiguë et cytokines), et de coordonner les activités des cellules immunitaire (cytokines), d'autre part d'une réponse cellulaire innée faisant intervenir de nombreux types cellulaires permettant la phagocytose (macrophages, cellules NK, cellules dendritiques) ou la sécrétion de molécules de l'immunité telles des peroxydases émises lors de la dégranularisation des polynucléaires, des substances vasodilatatrices ou des cytokines. La défense adaptative, elle, apparue plus tard dans l'évolution, fait intervenir les lymphocytes B (défense humorale reposant sur la production d'immunoglobulines reconnaissant spécifiquement les antigènes) et les lymphocytes T (défense cellulaire permettant la synchronisation de plusieurs cellules immunitaires, l'attaque de cellules infectées, la reconnaissance du non-soi). Cette défense adaptative repose sur un système de recombinaison somatique et aléatoire de l'ADN (recombinaison aléatoire d'une série de gènes V, D et J permettant d'atteindre plusieurs millions de combinaisons de protéines) qui permet la génération de chaînes protéiques variables qu'on retrouve dans les parties variables des immunoglobulines et dans le complexe majeur d'histocompatibilité permettant respectivement la reconnaissance spécifique des antigènes et la reconnaissance des cellules du soi. J'aime cet exemple car on voit bien que ces systèmes de réponse immunitaire sont liés mais on conçoit aussi qu'ils ont des niveaux de sophistication différents. Chacun de ces systèmes et sous-systèmes, ainsi que leurs raffinements sont apparus à différents moments de l'évolution, mais ils se sont progressivement combinés pour former un tout indissociable et globalement très efficace.

La vie sociale des cellules est encore un autre exemple assez évident : elle fait intervenir la communication chimique par diffusion de molécules dans l'environnement (ex : AMP cyclique produit par les amibes sociales *Dictyostelium discoideum*), le chimiotactisme qui permet leur regroupement en colonies, les différents niveaux d'adhérence incluant la formation d'un glycocalyx (manteau cellulaire de proteoglycanes) collant de façon peu spécifique aux surfaces, la production de matrice extracellulaire et de molécules capables de coller à ses constituants (couples intégrines – fibronectine / laminine), les jonctions cellule-cellule comme les jonction Gap, *etc.* Tout ceci est plus ou moins ancien et, dans certaines situations (assez nombreuses somme toutes), en particulier lors de phénomènes de cicatrisation ou lors de l'embryogenèse, mais aussi lors du développement tumoral, différents stades de cette vie sociale sont évoqués.

Le vivant est ainsi construit, par couches successives mises en place par divers mécanismes évolutifs

(copies, symbioses, plasmides, intégration de génomes, etc.)

9.6.2. Cancer

J'ai présenté quelques exemples qui permettent de se faire une idée de ce qu'est l'atavisme dans les systèmes biologiques. Mais s'il y a bien un endroit où ceci est bien présent et qui suscite un intérêt croissant, c'est dans les tumeurs !

Que le vivant comporte des atavismes est une idée déjà relativement ancienne, idée d'ailleurs associée au cancer, que M. Vincent [Vincent, 2011] fait remonter à deux textes de 1893⁵³ et 1926⁵⁴ (que je ne citerai pas dans la bibliographie, ne les ayant pas lus moi même ; je reporte leurs références en notes de bas de page). J'ai fait un peu de bibliographie moi aussi et je dois dire qu'il est difficile de trouver quelque chose à ce sujet jusqu'aux années 2000. Je soupçonne que les progrès de la biologie et de la médecine, la découverte de très nombreux mécanismes de régulation, en particulier les gènes contrôlant la prolifération cellulaire cycline dépendants, les gènes suppresseurs de tumeurs avec le rôle majeur dans ce contexte de certaines protéines clef comme *p53* ou *Rb*, les gènes de contrôle de l'intégrité tissulaire faisant intervenir des signaux provenant des cellules environnantes (voir [Israel, 1996]), mais aussi les thérapies qui devenaient de plus en plus efficaces (ce qui est toujours vrai), donc renforçant la confiance dans l'approche que l'on a du cancer et la perception nécessaire que l'on en a dans cette approche, ne sont pas pour rien si cette théorie a mis du temps à susciter de l'intérêt. C'est juste une hypothèse mal étayée, mais je dirais qu'il s'agit là encore du biais du vivant-machine : comprendre le cancer comme une machine moderne qui déraile est plus évident que d'évoquer et d'aller chercher son origine dans des reliquats datant d'avant même la grande oxydation d'il y a 2.4 milliards d'années ! Une autre raison de l'absence de visibilité de cette thématique est qu'elle ne devait probablement pas être identifiée comme thématique de recherche bien formulée. Il est cependant fort possible que des éléments de ce problème aient été identifiés séparément et aient suscité des interrogations, par exemple l'importance des protéines intrinsèquement désordonnées (IDPs) dans les facteurs de transcription et autres protéines impliqués dans le cancer, le développement cancéreux relevant davantage du désordre fonctionnel progressivement induit (via ces IDPs et leurs cibles, notamment de nombreuses cibles contextuelles comme les miRs) que d'un désordre génétique [Kulkarni, 2020]. La découverte de la régulation de contexte faite par les microARNs (miRs) au début des années 2000 a également changé la façon de voir la régulation cellulaire [Berezikov, 2005] et a certainement apporté de l'eau au moulin de cette théorie ataviste. Ce n'est qu'après que l'ensemble, suffisamment étayé, a pu constituer une théorie. Quoi qu'il en soit, il semble aussi qu'effectivement, toujours d'après M. Vincent, ce sujet refasse surface en 1996 avec un article de L. Israel (oncologue) donc le titre est on ne peut plus clair : *Tumour progression : random mutations or an integrated survival response to cellular stress conserved from unicellular organisms* [Israel, 1996]. Je pense que tout est dit dans ce titre ! Le

53 Référence non lue donnée par M. Vincent dans [Vincent, 2011] : Snow H. 1893. *Cancers and the Cancer Process*, London : J & A Churchill Publishers.

54 Référence non lue donnée par M. Vincent dans [Vincent, 2011] : Roberts M. 1926. *Malignancy and Evolution*, London : Grayson and Grayson Publishers.

cancer ? Il s'agit d'une réponse de survie des cellules à un stress accumulé, réponse de survie naturelle pour des cellules dont L. Israel soutient l'héritage de mécanismes de survie anciens. Israel lui-même se réfère à des articles des années 70 à 80 dans lesquels les auteurs montrent que certains gènes chez les organismes complexes ont été hérités des unicellulaires (procaryotes et eucaryotes), par exemple l'activation du régulon SOS activé chez les bactéries en cas de dommages à l'ADN⁵⁵. Israel, à juste titre, soutient que les systèmes de contrôle cellulaire (intégrité tissulaire, suppresseurs de tumeurs, prolifération) qui, en temps normal, permettent d'éviter le devenir aberrant de certaines cellules, ont des failles nécessaires, des failles qui permettent un fonctionnement hors norme comme la réparation des tissus (cicatrisation), l'embryogenèse et la croissance des organismes.

M. Vincent, mais également ses collaborateurs incluant l'astrophysicien P. Davies, ont aussi proposé plus récemment une compréhension détaillée de la formation ataviste des cancers ; la particularité de leur travail est d'avoir soigneusement identifié de nombreux systèmes affectés par la dérégulation progressive et d'avoir au moins partiellement documenté les stades de la dérégulation et de la dérégulation de plusieurs sous-systèmes [Vincent, 2011 ; Lineweaver, 2014 ; Bussey, 2017].

Tout est perturbation dans la vie cellulaire : l'assimilation de métabolites, des variations hydriques, des variations de température, d'acidité, du stress mécanique, *etc.* Nos cellules ont acquis au cours de l'évolution les dispositifs qui leur permettent d'absorber chacune des perturbations qui se produisent dans leur environnement physiologique, environnement qu'elles contribuent elle-mêmes à former et maintenir. Le maintien de cellules dans des conditions peu physiologiques, autrement dit un ensemble de perturbations soutenues dans le temps, contribue à produire la dérégulation de ces mécanismes ancestraux de survie, autrement dit à faire rentrer les cellules dans des phases ataviques de plus en plus éloignée de leur vie physiologique dans les organismes. En conditions normales, la dérégulation n'est pas un processus qui dure ; elle intervient dans certains processus inflammatoires relativement courts, dans les processus de cicatrisation, de croissance. Dans ces situations, les cellules peuvent, un temps, abandonner leur vie sociale, le confort de leur homéostasie, bien installées dans leurs tissus cellulaires, pour se diviser, migrer, exercer des tractions inhabituelles, *etc.*, bref, adopter des comportements qu'elles n'ont pas en temps normal. Au cours de la vie de l'individu, ceci est ponctuel et souvent progressif. Lorsque les conditions sont particulièrement délétères, ce sont plusieurs stades de régression atavique que la cellule franchit. En temps normal, les processus de contrôle peuvent encore déclencher son apoptose ; ou bien au contraire, si les conditions de vie cellulaire deviennent à nouveau acceptables (diminution d'une inflammation, ré-oxygénation, élimination de l'acidité excédentaire, *etc.*) la cellule peut à nouveau réprimer progressivement ses processus ancestraux et revenir progressivement à son état de fonctionnement normal dans l'organisme.

Je vais tout de suite préciser quelque chose que je pense important pour parvenir à adopter la compréhension de la théorie ataviste du cancer : du point de vue cellulaire individuel, la

55 Références non lues citées dans [Israel, 1996] :

Margulis L. (1970). Origin of eukaryote cell. Vol. 1 Yale: University Press.

Schwartz R. M. & Dayhoff M. O. (1978). Origin of prokaryotes, eukaryotes mitochondria and chloroplasts. Science 199, 395–403

Sedwick S. G. & Yarranton G. T. (1982). How cells in distress use SOS. Nature 296, 606–607.

dérépression et la dérégulation progressive jusqu'à l'état dit « cancéreux » de la cellule est ... tout à fait physiologique. La cellule s'étant adaptée à la nouvelle situation dans laquelle elle est confrontée vit une vie tout à fait acceptable pour elle. Tant qu'elle vit, les systèmes et éléments qui la composent ont joué leur rôle. Ce qui n'est pas physiologique c'est la présence de telles cellules et milieux cellulaires modifiés dans les tissus d'un organisme. Le cancer n'est pas physiologique du point de vue de l'organisme. Kulkarni aborde cet aspect quand, dans son article [Kulkarni, 2021], il distingue la multi-cellularité facultative ou obligatoire (*obligate multicellularity*). La multi-cellularité obligatoire (celle non-réversible des organismes multi-cellulaires) impose une coopération entre cellules au coût d'une moins grande adaptabilité cellulaire individuelle. Les cellules ont néanmoins conservé une certaine capacité à s'éloigner au moins temporairement de ce mode coopératif, en particulier pour permettre un retour à une vie autonome le temps d'une division cellulaire, mais aussi durant des phases de cicatrisation ou le développement embryonnaire.

Si l'on adopte le point de vue des systèmes dynamiques, l'état dit « physiologique » est le voisinage d'un attracteur, les conditions environnementales sont autant de perturbations qui éloignent la cellule de l'attracteur physiologique, et ... il est possible de basculer vers un autre bassin d'attraction, un autre état cellulaire, en cas de perturbations soutenues dans une mauvaise direction. La bascule d'un bassin d'attraction à un autre est quelque chose de classique lorsqu'on envisage la vie cellulaire dans l'espace des phases. Par exemple, les processus de différenciation peuvent être compris comme des voies choisies par un système dynamique dans un espace des phases contenant plusieurs attracteurs, chaque attracteur correspondant à un phénotype que peut exprimer la cellule ... autrement dit un type cellulaire différencié ou tout autre état, y compris un état pathologique ou un état ataviste.

Parmi les systèmes acquis relativement tardivement dans l'évolution, il y a les systèmes de contrôle des espèces oxygénées réactives (ROS) et les systèmes de réparation de l'ADN. Leur désactivation en cas de régression atavique favorise une impossibilité de réversion de l'état atavique : des mutations non réparées, provoquées entre autres par des ROS, frappant au hasard dans le génome, peuvent inactiver définitivement des gènes dont ceux qui seraient requis pour revenir à un fonctionnement physiologique. Cela fonctionne en quelque sorte comme un loquet anti-retour.

Tout indique que, du point de

vue des réseaux de régulation, ceux-ci sont organisés en couches en oignon successives, les plus anciens réseaux encapsulés dans les plus récents et ainsi de suite. L'architecture des systèmes ataviques au fonctionnement moderne est en tout cas une question ouverte à laquelle il faut s'intéresser. Parmi les approches possibles, la plus évidente est de s'intéresser aux systèmes impliqués connus et établir peu à peu une compréhension de leur structure à partir de données biologiques. Des données massives (données omiques : transcriptome et protéome notamment) analysées dans le cas du développement tumoral, puis intégrées dans des réseaux, pourraient bien permettre une telle compréhension. Il s'agit cependant d'un travail difficile pour lequel je n'ai pas la compétence et pour lequel il faudrait recruter quelqu'un ayant ces connaissances et ce savoir faire.

C'est aussi un projet d'envergure, collaboratif de préférence et je ne suis plus bien certain, surtout dans le contexte de recherche actuel, d'avoir la force de m'atteler à ce genre d'opérations. Une autre approche est théorique et, celle que j'ai choisie pour l'instant, bien que ne donnant pas de pistes sur comment agir sur le développement cancéreux et encore moins sur les systèmes réellement impliqués, permet tout de même de réfléchir à comment des systèmes anciens, dotés d'une fonction, peuvent être peu à peu intégrés au sein d'un système toujours plus grand et pluripotent. C'est ce dont je vais parler maintenant puisque c'est au cœur de mon activité de recherche actuelle et que cela va certainement m'occuper un moment.

Avant de passer au chapitre suivant, j'aimerais tout de même revenir sur un point que j'ai soulevé au début de cette section, la question de l'avantage évolutif que donnent aux individus d'une part, aux espèces d'autre part, les cancers et leur développement. Y en a-t-il seulement un ? Je ne sais pas bien répondre à cette question et je dois dire qu'elle m'intrigue. Une première réponse qui paraît évidente est l'élimination des individus susceptibles d'avoir un terrain génétique/métabolique trop fragile ou instable dans un environnement donné ; j'inclus ici le rôle de l'environnement car je ne pense pas qu'on puisse parler de « mauvais terrain génétique » sans parler de l'environnement qui teste le phénotype associé. Une autre raison assez évidente est que, sans atteindre le stade irréversible du développement tumoral, la capacité de rentrer dans un mode dégradé par rapport au fonctionnement dit « physiologique » permet d'augmenter au moins temporairement la survie dans certains tissus. Plus osé encore (et je ne suis pas bien certain de la validité de ce que j'écris là) : voir le processus de carcinogenèse précoce comme nécessaire dans les organismes pour entraîner le système immunitaire ... à y répondre ! Le système immunitaire, inné comme acquis, répond à la présence de cellules transformées, cancéreuses [Adam, 2003]. C'est d'ailleurs une approche importante pour la lutte anti-tumorale : entraîner, stimuler le système immunitaire pour une telle réponse. Un organisme qui générerait régulièrement des cellules anormales, déviantes, entraînerait intensément son système immunitaire et limiterait considérablement la probabilité d'apparition de tumeurs. A l'inverse, un organisme ne voyant aucune de ses cellules se diriger vers une voie tumorale, ne permettrait pas l'apprentissage de son système immunitaire pour des cellules cancéreuses et présenterait alors un risque plus important de développer des tumeurs.

Au delà de ces avantages l'un pour l'espèce, les 2 autres pour l'individu, il ne me paraît pas évident de suggérer d'autres hypothèses sur le rôle des cancers. Muller [Muller, 2017] y voit des organes transitoires chargés de dissiper l'énergie libre excédentaire d'un organisme. Je peine à être convaincu mais je trouve néanmoins l'idée intéressante (idée assez étayée tout de même) ; en tout cas, je constate que cette question, d'autres se la posent. Si une réponse, même partielle, y est associée, autrement dit si les tumeurs ont un rôle qu'elles jouent en tant qu'organe transitoire, la plupart du temps éliminée par l'organisme, alors comprendre ce rôle pourrait donner des indices sur comment éviter que le cancer s'installe.

Je ferais remarquer que l'on parle jusque là de l'atavisme cellulaire, mais on pourrait probablement décrire plusieurs échelles auxquelles des systèmes ancestraux se manifestent en cas de crise

soutenue, systèmes provoquant la disruption de l'ordre sus-jacent pour assurer la survie sous-jacente. Les organismes (animaux, plantes, champignons) sociaux gagnent individuellement à vivre en sociétés, mais lorsque la pression environnementale devient trop forte et trop soutenue et que la survie de l'ensemble des individus ne pourra vraisemblablement plus être assurée, les individus réactivent un fonctionnement davantage autonome améliorant la survie individuelle ou de petits groupes (NB : je ne m'étendrai pas là dessus ; il faudrait longuement étayer ces affirmations sur la sociologie, l'émergence et la disruption de groupes, sur le rôle de l'altruisme, sur sa modulation, sur l'ensemble des mécanismes que l'évolution a sélectionnés pour chaque espèce dans chaque écosystème et biome, permettant d'assurer au mieux la survie de ses gènes). Un autre exemple qui me vient en tête, sont les Troubles Dissociatifs (TD), comme le Trouble Dissociatif de l'Identité mieux connu sous le nom de trouble de la personnalité multiple [Damsa, 2006 ; Matulewicz, 2017]. Ces troubles correspondent à la perte d'une certaine cohésion de fonctions cognitives normalement intégrées : l'identité, la mémoire, la conscience et la perception de l'environnement. Ces troubles peuvent être provoqués par la prise de substances neurotoxiques mais également résulter d'un stress traumatique intense durant l'enfance. Les approches neurobiologiques montrent par des mesures sériques (ex : rapport déhydroépiandrostérone sulfaté/cortisol sérique ou diminution du cholestérol et triglycérides) que le métabolisme est affecté. Dans la dissociation d'identité, l'individu héberge plusieurs alters (personnalités) bien distinct (au moins deux, parfois une dizaine), certains étant des déclinaisons d'autres. Ces alters prennent littéralement tour à tour le contrôle de l'individu. Les changements d'un alter à l'autre font adopter des comportements différents, avec des façons différentes et indépendantes (chaque alter assume sa position indépendamment de celle des autres) de traiter les problèmes, des interactions avec autrui différentes, *etc.* Pour l'adulte et plus encore pour l'enfant qui n'a que peu d'atouts en main pour se défendre, c'est un mécanisme de défense (appelé *coping*). Dissocier ainsi ses personnalités, c'est un moyen de « survivre » parmi les autres dans des situations différentes. L'enfant de moins de 6 ans d'ailleurs, en situation physiologique, a une tendance à la dissociation, en expérimentant des personnalités et des attitudes à adopter dans la vie, phase qui s'atténue puis disparaît lors d'un processus d'intégration qui assoit une future personnalité. Ce trouble (ces troubles) est « vu comme » un trouble notamment parce que ces personnes ont du mal à être acceptées comme telles, avec des attitudes changeantes, et à conserver une place en société. Cependant, si l'on analyse cela avec un regard plus éloigné, davantage centré sur la survie d'une espèce de façon générale, ce n'est plus nécessairement un trouble, mais une capacité. La possible origine acquise (traumatique), voire dynamique donc possiblement et au moins partiellement réversible, d'un désordre fonctionnel générant les TD est encore débattue (mais bien étayée tout de même [Chambers, 1999]). Je trouve intéressant et plausible le fait que l'évolution ait retenu cette possibilité dissociative chez tout un chacun comme mécanisme de défense. Après tout, il y a bien d'autres troubles de la cognition induits par un traumatisme de stress tels l'amnésie, ou à moindre effet les réactions induites par la peur, l'ensemble pouvant être compris comme un phénomène dynamique d'ajustement du fonctionnement du cerveau [Sel, 1997 ; Bracha, 2006]. Cela constituerait en tout cas un atavisme à l'échelle de l'organisme.

9.6.3. L'atavisme dans les réseaux – trajectoires ataviques

Le vivant ne rebat pas systématiquement les cartes ; il hérite, puis il intègre cet héritage, éventuellement jusqu'à l'offusquer. Par extensions successives (telles que définies précédemment) ou par tout autre mécanisme évolutif de croissance de réseaux (par exemple l'ajout de modules et de systèmes de contrôle de l'expression comme dans les réseaux pluripotents), les réseaux des organismes peuvent se complexifier tout en conservant une trace ré-activable des systèmes ancestraux.

Les travaux que nous menons sur l'extension des réseaux Booléens et de leurs dynamiques associées, de même que ceux qui concernent les réseaux pluripotents, s'inscrivent évidemment dans cette démarche ; inutile d'y revenir, j'en ai suffisamment parlé. Une démarche inverse consiste en revanche à identifier dans de larges réseaux des sous-réseaux (modules) et les sous-parties de la dynamique globale issues de ces modules. Cette démarche n'est pas évidente non plus : elle consiste à déterminer si des sous-parties d'une dynamique peuvent provenir spécifiquement de sous-partie du réseau étudié. Cela demande dès lors de spécifier un degré de spécificité réciproque du sous-réseau et de la sous-dynamique : on peut vouloir rechercher un sous-réseau dont l'activité propre, indépendante du réseau auquel il est connecté, décrit complètement la sous-dynamique, mais il est fort probable qu'une telle dynamique sera partiellement perturbée par l'activité du reste du réseau. De même, on peut vouloir identifier un sous-réseau à partir d'une sous-dynamique dont on suppose qu'elle est le résultat de l'activité d'un module, mais il est probable que, le temps de l'évolution aidant, le réseau à l'origine de cette activité se trouve peu à peu offusqué, ses fonctions (interacteurs biologiques : gènes, IDP, ...) se trouvant factorisées. Le réseau initial a perdu son intégrité. Le retrouver devient difficile.

Cependant, ce qui à mon sens est plus intéressant que le « produit atavique » de l'évolution (un réseau de régulation construit comme un héritage de réseaux peu à peu intégrés), c'est la « trajectoire atavique » que suit l'évolution d'un organisme conservant certains systèmes dits ancestraux. Les « choix » que l'évolution aura faits de connecter tel et tel système ensemble, d'en offusquer une partie, *etc.*, sont une contrainte pour l'avenir des lignées qui hériteront de ces choix.

Si les systèmes vivants se complexifient depuis les origines du vivant, tout en héritant des précédents, ils gagnent d'une part en capacité de réponse cellulaire aux nombreux stimuli extérieurs et croissent en complexité organisationnelle (tissulaire, organique, sociétale, *etc.*), ceci leur donnant également des capacités accrues pour appréhender leur environnement et y faire face. Il me semble que les organismes ainsi développés ont cependant une spécialisation plus importante pour l'environnement dans lequel il se sont développés, une spécialisation qui les rend très efficaces dans ces environnements mais qui réduit leur capacité exploratoire génotypique et phénotypique. Cela pose problème lors de changements environnementaux (dérive climatique, pollution croissante de l'air, des eaux, variation des niveaux de radiation solaires, *etc.*). Là où des organismes peu spécialisés pourront très vite produire une très grande variabilité génotypique et afficher la variabilité phénotypique induite permettant à certains organismes issus de ces larges lignées de rester viables

malgré ces variations environnementales, ce ne sera pas aussi facilement le cas de systèmes très complexes comme les organismes supérieurs actuels.

Parce que les organismes, aux grandes échelles temporelles de l'évolution, n'ont pas traversé une variabilité importante d'environnements, ils n'ont accumulé que les choix évolutifs qui ont été autorisés par les contraintes de viabilité imposées par ces environnements changeants. Je dis là qu'un organisme est une forme d'enregistrement des différentes conditions que la lignée dont il est issu a traversées. L'évolution a inscrit en dur, dans les systèmes du vivant, leur réponse aux environnements qu'ils ont rencontrés et ses variations. *A priori*, plus un organisme contient de systèmes d'interaction, plus il est susceptible d'avoir accumulé l'histoire de sa trajectoire évolutive.

De tels organismes complexes n'ont pas si facilement que ça la possibilité de s'adapter aux variations environnementales importantes en profitant de leur nombre comme le font les bactéries ; ils ne sont clairement pas assez nombreux. Il ne peuvent le faire que lentement. Les humains se sont adaptés à tous les milieux sur la planète parce qu'ils disposaient de temps, de beaucoup de temps pour le faire. Les espèces vivantes mettent du temps à s'adapter aux changements climatiques globaux ; elles le font rapidement par des déplacements quand c'est possible, génétiquement sinon mais cela prend du temps. Leur complexité est en jeu. Comme nous l'avons vu quand nous avons parlé de la complexité des réseaux et du problème de l'extension, les réseaux *a priori* les plus évolutifs ne sont pas les plus complexes. Le « vivant qui anime ces organismes complexes »⁵⁶ a cependant conservé une stratégie d'échappement pour sa survie : la dissociation, le changement de fonctionnement. La dégradation des conditions environnementales déplace les équilibres dynamiques dans lequel ce vivant évolue et le pousse à sortir de la multi-cellularité, de la vie collective, à changer de relations écosystémiques, *etc.*, à la survie quoi qu'il en coûte ; c'est en tout cas une des solutions qui peut être retenue par un système vivant. Néanmoins, les systèmes vivants, dans un environnement changeant trop vite, selon leur complexité, pourront montrer leur limite d'adaptation et s'avérer finalement être des impasses évolutives. Leur génétique et plus encore les réseaux que ces systèmes expriment deviennent une fin de trajectoire. Ceci n'est vrai que si l'on considère de tels réseaux comme peu malléables, trop rigides du point de vue évolutif. La souplesse introduite par la régulation de contexte sur laquelle nous avons encore beaucoup à apprendre, en particulier celle permise d'une part par les microARNs, d'autre part par les IDP, est un moyen de s'échapper de telles trajectoires figées.

J'ai introduit l'idée de trajectoires évolutives dans l'atavisme, de voir les êtres vivants comme des lignées accumulant un historique de mécanismes de survie, mais gagnant souvent en complexité, l'ensemble résultant en un potentiel évolutif et une certaine robustesse. Ceci rejoint ce dont nous avons parlé sur l'évolutivité, la robustesse et la viabilité. Mais peut-être est-il plus judicieux pour parler de cela de commencer par le commencement de ces trajectoires : les origines du vivant.

⁵⁶ je formule cette phrase ainsi pour faire ressortir que l'organisme multicellulaire est composé de parties vivantes (tissus, organes) et à la plus petite échelle d'êtres vivants (les cellules) dont le fonctionnement les contraint normalement à la vie coopérative et collective.

9.7. *Le rare, le faible l'amorphe et la vie ... artificielle*

Nous allons maintenant boucler ce document (il est grand temps !) avec un retour sur les origines du vivant ! C'est ce sujet qui me passionne le plus. Forcément, c'est le plus intrigant !

Le long chapitre 8.2 s'est intéressé à ce qui constitue probablement la question la plus complexe de la biologie, à savoir l'origine de la réplication et de la traduction (génome nucléique → protéome) [Koonin, 2007]. Nous avons approfondi la réflexion sur ce point en proposant d'abord de tester expérimentalement la théorie stéréochimique du code génétique et, d'une façon plus générale, l'affinité spécifique d'acides aminés pour certaines séquences nucléotidiques courtes, mécanisme fondamental qui permettrait de soutenir certaines hypothèses sur l'apparition du mécanisme primordial de traduction. Ensuite, j'ai proposé d'étudier par simulation l'émergence primordiale des premiers systèmes faisant à la fois émerger génomes et protéomes mutuellement liés. J'ai aussi souligné à quel point ceci n'était pas simple à modéliser, ni correctement ni efficacement. Maintenant que j'ai aussi longuement parlé des réseaux et de ce qu'on pouvait faire avec, il est temps de voir que ces deux thématiques peuvent ne constituer qu'un tout.

Avant de voir comment, permettez moi de discuter un peu, dans un premier paragraphe, des effecteurs de la vie. Il existe en effet plusieurs manières de représenter la dualité génotype-phénotype dans les systèmes de vie artificielle, mais il n'est pas très aisé, dans un modèle, de rendre effectrice une structure (ex : protéine) produite par un génome, de telle manière que ce génome, s'il évolue, puisse générer une structure dont la fonction évolue aussi. Autrement dit, on souhaite un encodage (génétique, fait de nucléotides) qui produit un autre encodage (protéique, fait d'acides aminés) duquel émergent une à plusieurs structures tridimensionnelles dont la forme et la chimie émergente permettent des interactions avec d'autres structures tridimensionnelles. Si la traduction, sous certaines hypothèses (hypothèse stéréochimique du code génétique par exemple) peut être admise, il reste, en modélisation, à parvenir à simuler l'émergence de la fonction en fonction de la structure, et de l'auto-génération de systèmes autopoïétiques fondés sur ces structures fonctionnelles. On parle là de simuler, dans un ordinateur, l'émergence d'une chimie prébiotique, c'est-à-dire de structures fonctionnelles générées de façon abiotique puis, de là, de systèmes biochimiques primordiaux. Dans cette section, je commencerai par traiter cette problématique qui s'avère constituer un joli sujet de M2 dans le contexte des origines du vivant.

Si l'on veut créer le vivant artificiel, il faut que, d'une part, celui-ci intègre des fonctionnalités qui le rendent vivant dans ses éléments (molécules support d'information, molécules effectrices, dissipation d'énergie, *etc*, toute une biochimie et biologie moléculaire compatible avec le fonctionnement de la vie) mais également que les systèmes qui émergent et qui sont composés de ces éléments biochimiques et biomoléculaires en interaction portent en eux la capacité d'évoluer, donnant à l'organisme qui en dépend et aux lignées qu'il initie une possible échappatoire lui assurant sa survie dans un environnement changeant. Par « porter en eux la capacité d'évoluer », j'intègre bien sûr la vision classique de l'évolution des molécules et des systèmes qui en découlent par

l'extraordinaire propriété de l'ADN à évoluer, mais j'entends également une capacité émergente des systèmes à évoluer, être combinés, *etc*, en maintenant leur viabilité. C'est une triple contrainte incluant robustesse face à l'évolution, évolutivité et viabilité, un triplet *a priori* paradoxal, qui peut être étudié à l'aide des réseaux Booléens. C'est également ce que je montrerai ici.

9.7.1. Émergence de la fonction biologique et de l'organe – les réseaux-molécules

Les réseaux Booléens dont nous avons discutés longuement sont des entités animées d'une activité. En cela, il s'apparentent un peu aux objets fonctionnels du vivant comme les neurones dans le cerveau, les systèmes génétiques ou métaboliques, ou encore les atomes, groupes d'atomes et niveaux de structuration dans une molécule. Ils en constituent une abstraction dont on ne retiendrait que les changements d'état et la rythmicité : succession ordonnée des états électro-chimiques d'activité neuronale dans le cerveau, succession des états d'expression génétique (il y a synthèse ou non) dans les systèmes génétiques, succession ordonnée (régulée) des états locaux de conformation moléculaire ou atomique au niveau des sites actifs ou des régions d'interaction, ou de tout autre région des molécules biologiques dont la conformation est susceptible de changer et d'affecter l'activité.

Mais pour revendiquer plus pleinement leur ressemblance avec les objets du vivant, il leur manque au moins le mouvement.

Le monde est état et mouvement. Les molécules ont des conformations que l'on traduit en états et qui passent d'état en état grâce au mouvement, ce même mouvement des groupes atomes qui provoque leur activité : la catalyse, l'inhibition, *etc*. Les cellules, les tissus et les organes sont dans ce que l'on appelle un état physiologique ou pathologique ou au repos, état que l'on peut préciser par des niveaux d'activité (ex : niveau de consommation d'oxygène dans telle ou telle partie du cerveau, niveau de production d'insuline par le pancréas), des concentrations en molécules particulières, *etc*, et en changeant grâce aux mouvements de matière comme le transport des molécules qui le perfusent ou les activités biomécaniques telles l'activité pulsatile du rein ou du cœur, ou au niveau cellulaire le changement d'activité des plaquettes sanguines du à la restructuration de leur cytosquelette. Le cerveau n'est pas une « simple » machine à états dont la complexité est liée à l'organisation complexe du réseau neuronal ; sa complexité provient aussi de ce qui se trame en son sein à l'échelle moléculaire, de la diffusion des neurotransmetteurs dans l'espace synaptique et perisynaptique, à la traversée des membranes axonales par les ions et à la perfusion hétérogène dans le tissu neuronal à partir des vaisseaux sanguins.

Le mouvement est nécessaire ; il est l'effecteur du vivant. Les états n'en sont que la conséquence perçue, mesurable.

La fonction biologique, l'organe, émergent donc de la matière en mouvement et de la synchronie et prennent sens dans un système. L'organe ou la fonction biologique (pour généraliser parlons d'*organe-fonction*) c'est la matière structurée en action. C'est le résultat collectif de l'action synchronisée de composants matériels. Je cours parce que mes fibres musculaires sont synchrones

entre elles, que les milliards de molécules qui créent le mouvement (actine, myosine ...) sont synchrones, que les neurotransmetteurs traversent les nombreuses synapses de façon synchrone, que l'activité des différents muscles est permise par une synchronie encore plus large impliquant le système cardiorespiratoire... Le maintien homéostatique comme le contrôle nécessaire pour assurer la synchronie des systèmes au cours du temps se font dans un référentiel qui est le système dans lequel cet organe-fonction s'intègre.

Donc, ce qui évolue, finalement, c'est la fonction biologique qui contribue à la vie de l'organisme, dans un contexte environnemental (environnement cellulaire, tissulaire, organique, écosystémique) particulier, à cause de changements dans la structure matérielle de l'organisme, dans l'ADN directement lié à la fonction évidemment, mais aussi dans les configurations physico-chimiques de l'organisme ou du système dans lequel cet organe-fonction opère (forme anatomique, modifications post-traductionnelles de protéines ...).

Je m'intéresse beaucoup à ce moment où l'activité chimique se met à « jouer un rôle » et devient partie intégrante et nécessaire d'un système plus vaste qui se maintient, c'est-à-dire où elle commence à devenir fonction, à jouer un rôle vital, nécessaire à la survie de quelque chose (un cycle métabolique, un répliqueur (gène au sens de Dawkins), un organisme, *etc*). J'en ai déjà parlé dans le chapitre 8.2 consacré aux origines du vivant ; je vais en parler à nouveau en m'appuyant sur les travaux et propositions que j'ai présentés dans le chapitre 8.

Dans les recherches que nous faisons avec mes étudiants et collègues, nous utilisons comme support de réflexion les réseaux Booléens à signe. Ces objets théoriques représentent bien les réseaux d'interaction naturels. Ils présentent aussi l'avantage d'être dénombrables (leurs états et paramètres sont discrets) ce qui facilite les calculs que l'on peut vouloir effectuer dessus : complexités, distances, classification subséquente. Je vais montrer ici que l'on peut utiliser de tels réseaux, en les augmentant d'une dynamique physique pour modéliser et étudier les phénomènes liés à l'émergence d'une chimie prébiotique artificielle, en particulier l'héritage de fonctions dû à l'héritage des structures, et les brisures de symétrie liées notamment à la composition initiale du milieu et à la chimie élémentaire des éléments en interaction.

L'idée ne vient pas de moi mais de Rémi Segretain, étudiant en thèse encore sous ma direction alors que j'écris ces lignes. Cette idée est née d'une longue discussion que j'ai initiée sur les systèmes de vie artificielle dans lesquels nous souhaitions forcer des réseaux à évoluer (par mutations, copies, *etc*) et dans lesquels nous souhaitions mettre en évidence la formation de lignées de créatures héritant de réseaux leur donnant une fonctionnalité intéressante pour la survie. La question s'est posée sur comment encoder la fonction (au sens de système effecteur) de façon à ce qu'une gradation de sa complexité (mélange de fonctionnalités) ou de ses performances (fonction plus ou moins efficace) puisse se faire. Il fallait pour cela que la fonction émerge d'un encodage sous-jacent et non qu'elle soit pré-établie.

Avant d'expliquer comment faire, je vais prendre quelques lignes (nous ne sommes plus à ça près je crois !) pour parler des quelques systèmes de vie artificielle, en particulier ceux que nous avons

envisagés pour étudier l'évolution des réseaux. Cela permettra de bien comprendre ce qui est en jeu.

Pour modéliser l'effet de fonctions biologiques ou d'organes dans un système de vie artificielle on peut procéder, et c'est souvent le cas, par une mise en correspondance entre la présence d'une caractéristique associée à un système ou organisme (ex : propriété ou particule) avec une fonction préprogrammée. Par exemple, si un organisme artificiel dispose d'une nageoire (encodée par exemple la forme d'un champ Booléen avec une valeur particulière, disons 10), alors le programme qui simule le comportement de cet organisme contient aussi une description de l'effet attendu sur cet organisme par la présence d'une nageoire. Dans un tel système sans émergence, difficile de concevoir une quelconque évolution de la fonction, sauf à programmer autant de variations de l'organe et leurs comportements associés.

Il y a cependant des exemples de programmes dans lesquels la fonction émerge de l'arrangement d'éléments simples, par exemple les créatures de Karl Sims [Sims, 1994 a et b], des blocs géométriques animés, reliés entre eux par des effecteurs (des « muscles ») et dont l'évolution se faisait sur la structure de contrôle (un réseau) coordonnant leur activité motrice, ou les Core Wars proposés par D.G. Jones et A.K. Dewdney [Jones, 1984 ; Dewdney, 1984 & 1985 ; Rennard, 2002], des environnements mémoire dans lesquels des créatures formées d'instruction en pseudo-langage assembleur⁵⁷ se combattent et évoluent, *Venus* de S. Rasmussen [Rasmussen, 1989 & 1990] ou *Tierra* de T. Ray [Ray, 1992 & 1994], puis *Avida* [Adami, 2000] en étant des raffinements particulièrement étudiés. Ces modèles historiques en ont suscité de nombreux autres dont je ne vais pas faire l'inventaire ici.

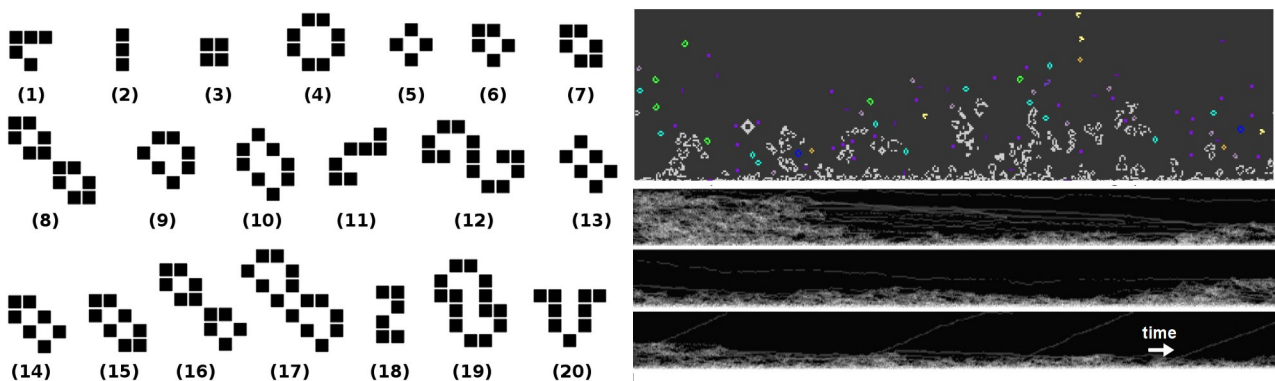


Figure 9.12. Créatures du jeu de la vie et simulation de vie artificielle. Des simulations (droite) utilisant les règles du jeu de la vie ont permis, en suivant une vingtaine de créatures périodiques (gauche), de s'intéresser à l'effet de la gravité et de l'agitation du milieu (diffusion) sur la survie des créatures et leur diversité [Glade, 2017]

En 2007, avec Pascal Ballet, nous avons commencé à nous intéresser à la vie artificielle pour étudier l'effet du rôle de l'environnement physique, en particulier l'agitation thermique et la gravité, sur un écosystème pouvant éventuellement représenter un système prébiotique [Glade, 2017] (figure 9.12). Les éléments de ce système étaient les créatures du jeu de la vie. Un tel système permet

⁵⁷ Le langage *Redcode*

l'émergence de structures, dont certaines présentent une activité (une fonction dans l'écosystème?) comme se déplacer (gliders), s'agiter (blinkers), l'auto-régénération (square). Cependant, ces créatures sont très fragiles (un contact avec une autre cellule extérieure et c'est la mort) et ne se reproduisent pas (d'une créature, on ne peut obtenir en aucun cas un autre exemplaire semblable) ni n'évoluent. En revanche, les créatures ne sont pas formées complètement au hasard : les 2 états possibles de cette matière, les règles de voisinage et d'évolution du système, et la densité locale en cellules, densité auto-régulée par les règles d'évolution, font que la distribution des configurations atteintes est biaisée. De ce point de vue, ce système n'est pas un système de vie artificielle, mais de prébiotique artificielle.

Nous avons peu après envisagé et programmé un autre système de vie artificielle avec des créatures plus durables et dotées d'un phénotype variable (voir figure 8.23). L'objectif était toujours d'étudier l'effet de la stratification de l'environnement par la sédimentation sur la diversité et la durée de vie de formes de vie artificielles, cette fois avec des créatures plus « solides ». Dans ce système, les créatures disposent d'organes dont la présence est assurée par la présence de pseudo-gènes dont l'encodage spécifie notamment le type d'organe (2 bits), sa position sur la créature (3 bits). Il ne s'agit pas d'un génotype au sens classique du terme, aucun système de traduction n'étant requis, la mise en correspondance étant directe (ex : 11 = nageoire, 10 = bouche), ceci ne laissant aucune place à une quelconque possibilité d'évolution génétique. Un certain héritage génétique se produit toutefois, accompagné possiblement de variations phénotypiques : les mutations (mutations ponctuelles, duplications ou délétions de gènes) permettent le changement de composition et de localisation des gènes. Cela permet l'exploration de la composition idéale en organes pour que l'ensemble des créatures en disposant vivent en écosystème. En revanche, rien ne les synchronise. Dans ce système, les créatures qui croissent, à force d'accumulation de masse par assimilation d'autres créatures, finissent par se diviser, maintenant ainsi un rapport surface/volume c'est-à-dire assimilation/consommation raisonnablement constant. Les créatures filles sont alors une copie potentiellement imparfaite des créatures mères. En cela, c'est une simulation de vie artificielle.

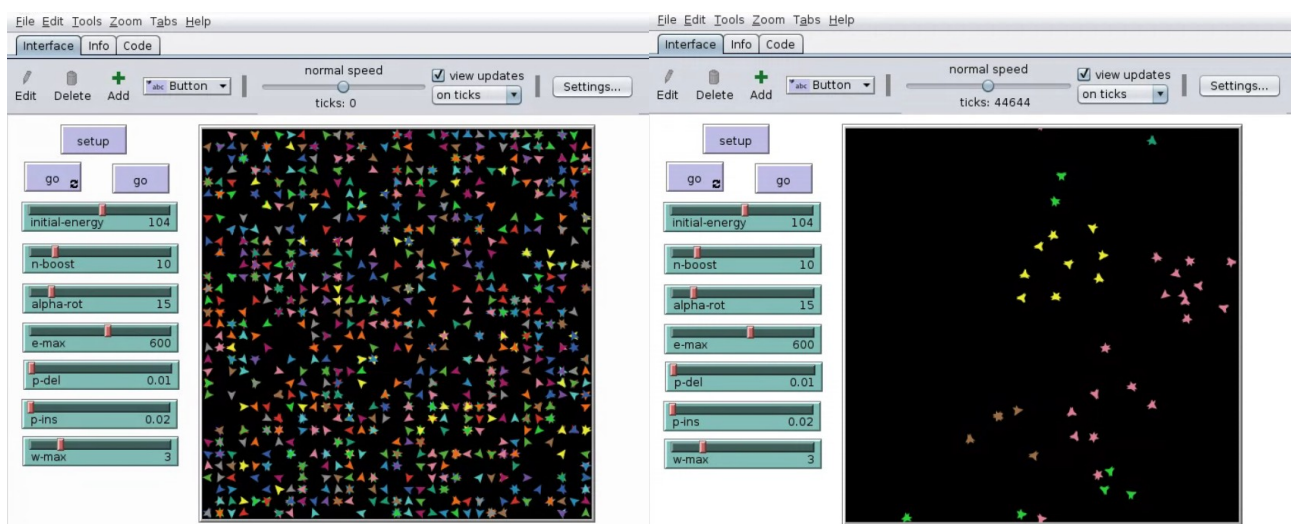


Figure 9.13. Simulation de créatures-réseaux. Simulation en NetLogo de créatures-réseaux (modèle de S. Ivanov ; voir texte) à l'état initial et après quelques milliers d'itérations. Certaines lignées de créatures-réseaux se sont particulièrement développées, formant ensemble un écosystème [Larchères, 2020].

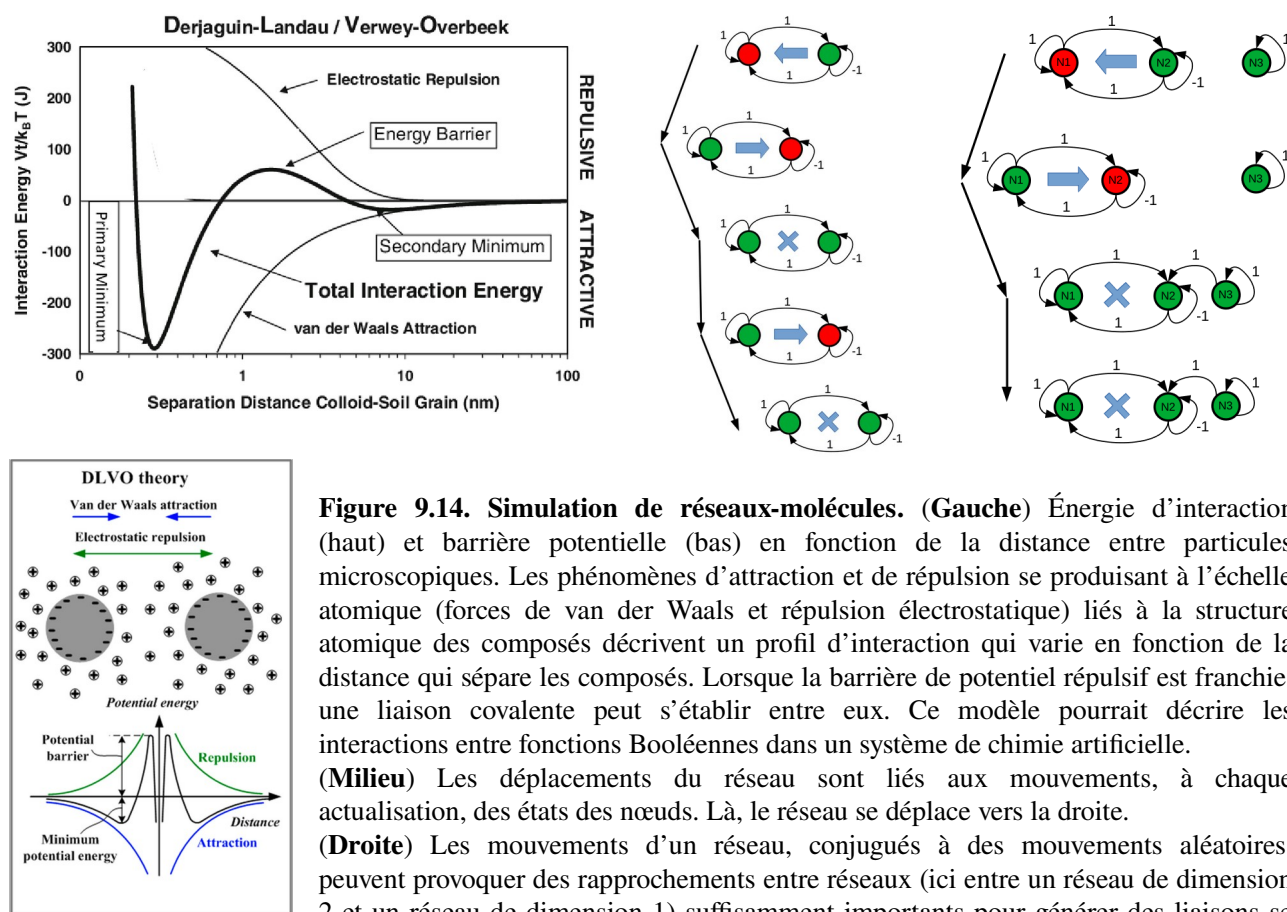
Depuis, notre objectif a changé. Nous souhaitons exploiter les réseaux Booléens à signe, en particulier les liens structure/fonction, complexité structurelle/caractéristiques comportementales (complexité dynamique, multi-attractivité, robustesse dynamique), qui ressortent dans nos travaux, pour se pencher, à l'aide d'un système artificiel évolutif, sur des questions d'évolution des réseaux et des comportements associés. Comme je vais développer ce point dans la section suivante (section 9.7.2), je ne m'y attarde pas ici. L'idée générale est d'en tirer des enseignements sur les types de réseaux les plus à même d'évoluer facilement sans casser les comportements associés et la survie subséquente. Il en ressort des réflexions supplémentaires sur comment modéliser un génotype, son phénotype associé, émergent (non préprogrammé) et évolutif (nouveaux comportements), c'est-à-dire sur la naissance de la fonction-organe biologique.

C'est mon collègue d'Evry, Sergiu Ivanov, qui a proposé de mettre à profit nos connaissances sur les réseaux Booléens à signe et d'utiliser ceux-ci dans un nouveau système de vie artificielle, encadrant à l'occasion deux étudiants en M1 (pas très impliqués malheureusement) dont le travail nous a permis d'évaluer l'intérêt de cette approche (voir figure 9.13). Son système est une simple variation du système présenté précédemment : les organes (nageoires, bouches) préexistent mais ne sont pas nécessairement actifs. Chaque organe est au moins contrôlé par un nœud du réseau, réseau qui peut posséder plus de nœuds que d'organes. L'activité successive des nœuds du réseau contrôle l'activité successive des organes. Ce réseau peut muter par changement des valeurs des poids associés aux arcs. L'évolution explore donc les systèmes de contrôle de synchronie des activités organiques. L'écosystème est donc une composition harmonieuse de créatures disposant de systèmes de contrôles particuliers. Comme pour le système que nous avons implémenté avec P. Ballet, la division en quasi-copies se fait lorsque la créature mère atteint un rapport surface sur volume trop faible (surface et volume étant simulés par une masse de matière accumulée par ingestion et des masses critiques, inférieure et supérieure, définissant des seuils respectivement en dessous duquel l'organisme tend à se dégrader en simple matière inerte ou au dessus duquel l'organisme tend à se diviser). C'est aussi un système de vie artificielle.

J'ai alors cherché à lier, dans une même créature artificielle, l'activité d'un réseau à une fonction émergente. La solution qui m'est apparue alors est un mapping entre les nœuds du réseau et une grille de type automate cellulaire assurant une action physique sur la matière environnante plongée dans l'automate. L'idée était que le réseau servait d'initiateur, de modificateur, de régulateur de l'activité d'un automate cellulaire (ou de plusieurs : 1 par organe) et que cet automate cellulaire, par son comportement, produirait une action. Les règles des automates cellulaires de la simulation correspondent à la physique injectée dans ce système. Ce n'est pas un système de vie artificielle simple à implémenter ; je ne l'ai d'ailleurs pas fait. Il s'agirait toutefois d'une simulation de vie artificielle car reposant sur l'existence d'organismes ayant un génome, capables de se reproduire et d'évoluer (à la manière du modèle précédemment décrit) mais disposant en plus de la notion de fonction émergente. Remi a proposé ce qui, finalement, est bien plus qu'une variation. C'est ce

système qui va nous intéresser maintenant.

Cette fois, l'« organisme » est un réseau. Comme nous allons le voir, il sera difficile de parler d'organisme maintenant vu que ces particules ne disposent pas de fonctionnalités implicites comme la reproduction. Par analogie, il s'agit davantage de *réseaux-molécules*. Appelons les ainsi.



L'idée, qui est finalement davantage à considérer comme une expérience de pensée plus qu'une réalisation à faire, est la suivante : on simule un milieu dans lequel des réseaux-molécules peuvent se former en suspension à partir de fonctions Booléennes, ont leur activité propre générant une chimie physique et peuvent interagir entre eux formant une chimie complexe.

Un réseau-molécule est représenté par un réseau Booléen à signe dans lequel les nœuds et les arcs ont une consistance physique : les nœuds sont des masses, les arcs des ressorts modélisant des liaisons moléculaires ; on a donc un système masse-ressorts qui peut adopter des conformations spatiales et peut donc représenter des molécules. Ces réseaux-molécules sont considérés comme des entités dispersées dans un environnement spatial. Si les ressorts qui modélisent les liaisons

moléculaires simulent un potentiel DLVO⁵⁸ (voir figure 9.14), alors des collisions suffisamment importantes entre réseaux-molécules peuvent rapprocher suffisamment 2 fonctions Booléennes et, étant donné des règles déterministes ou probabilistes qui restent à définir pour décrire de telles interactions, provoquer la formation de nouvelles liaisons (arcs) et l'obtention de réseaux-molécules plus grands. De même, au début de la simulation, le milieu n'estensemencé d'aucun réseau mais de fonctions Booléennes à signe de dimensions et de paramètres variés. La proportion des différentes fonctions Booléennes est fixée au départ de la simulation, par exemple sur la base de la rareté des fonctions : les fonctions de plus basse dimension sont les plus fréquentes et le volume de leur classe d'équivalence dans une dimension donnée ajuste leur distribution. Les collisions entre fonctions Booléennes ou entre fonctions Booléennes et réseaux peut amener à la formation ou à l'accroissement de réseaux-molécules.

Ces réseaux-molécule ont une activité qui dépend de leurs paramètres propres (poids sur les arcs) et de leur initialisation. Des 0 et des 1 vont donc circuler, durant la simulation, dans les nœuds du réseau. Là, Rémi a eu l'idée de transduire cette circulation des bits dans le réseau en mouvement : le déplacement d'un bit d'un nœud à l'autre génère une force de réaction. De telles forces déforment le système masse ressort qui, en se déformant puis en se relaxant, peut vibrer, se déplacer, *etc* au gré de l'activité particulière du réseau. À une structure et des états des nœuds correspondent une activité particulière du réseau-molécule. La physico-chimie associée est donc directement liée à l'activité des réseaux-molécules et à l'agitation du milieu à laquelle ils sont soumis.

Nul besoin de réaliser une telle simulation pour comprendre ce qu'il va se produire : dans le texte que j'écris ci-dessous, il suffit de remplacer « atomes » par « fonctions Booléennes » et « molécules » par « réseaux-molécules » et on comprend que tous les réseaux-molécules ne pourront pas se former.

La distribution des atomes dans l'univers n'est pas aléatoire : il y a beaucoup d'hydrogène, puis d'Hélium, puis ... (cf. tableau Mendeliev). La distribution non aléatoire des atomes et leurs particularités physico-chimiques respectives (nombre de valences, électronégativité, ...), et leurs densités locales dans l'univers (les nuages moléculaires sont riches en certains composés, la composition atomique n'est pas la même en fonction de la distance à l'étoile ...) font que certaines combinaisons atomiques (molécules) se font plus que d'autres dans l'univers et que cela dépend des conditions physico-chimiques locales. Ces molécules nouvellement formées constituent des briques dont la distribution de composition et de forme spatiale elle-même n'est ainsi pas aléatoire. La conséquence est que la combinaison de ces molécules entre elles (dictée par leurs proportions respectives, leurs caractéristiques physico-chimiques et leurs propriétés émergentes, l'activité catalytique due à l'arrangement particulier des atomes) génère là aussi des profils non aléatoires de molécules de plus en plus complexes. Ainsi, on comprend que certaines molécules ne pourront se former, ou seront rares car difficiles à former, et que les lignées qui pourraient en résulter ne verront pas le jour. La seule possibilité de les atteindre est l'apport de mutations, d'aléatoire en général

58 La théorie DLVO établie par Derjaguin, Landau, Verwey et Overbeek explique la stabilité des suspensions colloïdales. Elle décrit l'équilibre entre 2 forces, les répulsions électrostatiques et les forces d'attraction de van der Waals.

(notamment grâce à l'agitation thermique apportant un mouvement aléatoire aux réseaux-molécules), qui permettent de rejoindre des trajectoires autrement non atteignables comme le montre la figure 9.11.

Sous réserve de résolution du problème de l'origine de la réplication et de la traduction, la vie n'est probablement pas rare dans l'univers. Elle dépend alors probablement davantage de l'obtention de conditions favorables et stables, de planètes telluriques placées à bonne distance de leur étoile, d'une étoile stable, ... que d'autre chose. Les biais de composition et de règles (physico-chimiques) amènent vraisemblablement inexorablement à explorer un arbre biaisé d'évolution moléculaire (prébiotique) et probablement l'obtention de cycles métaboliques auto-entretenus et de systèmes mémoires pouvant être traduits en systèmes effecteurs tels l'ADN et les protéines. Sur le même principe, si la stabilité des conditions physico-chimiques est assurée, rien ne s'oppose à la complexification des proto-systèmes vivants en Vie, puis en Vie de plus en plus complexe (la capacité d'être à la fois stable et partiellement versatile des chaînes mémoire étant une conséquence de la physico-chimie des systèmes qui les lisent et traduisent).

9.7.2. La vie depuis ses origines, des trajectoires biaisées dans un mille-feuille évolutif

La vie peut-être vue comme un ensemble de trajectoires biaisées à travers un mille-feuille dont chaque couche représente un accroissement temporel, un accroissement de complexité, une adaptation, une évolution par rapport à la précédente...

Ces trajectoires qui symbolisent les lignées moléculaires, systémiques, organiques, sont contraintes par de nombreux facteurs. Tout ce dont nous avons discuté depuis le début de ce document, notamment ce qui concerne la complexité, l'évolutivité, la robustesse, la viabilité, *etc*, et tout ce que nous avons dit sur les réseaux, l'émergence biaisée de réseaux-molécules, leur extensibilité, leur composition, leur niveau de description même, est particulièrement adapté à réfléchir aux origines du vivant. Je taquinerais volontiers le lecteur arrivé à ce stade du document en lui suggérant de relire le document avec ce prisme des origines du vivant ; ce serait probablement un nouveau chemin de croix mais ce serait pourtant une bonne idée.

J'ai donc écrit beaucoup de choses à ce propos. Je voudrais seulement ajouter quelques dernières réflexions sur ce qu'impliquent les trajectoires évolutives, sur les contraintes qu'elles subissent, sur ce qu'elles enregistrent. Autrement dit qu'est ce que les cellules ont du mémoriser au cours de l'évolution ? Et comment ? Ce faisant, j'en profiterai pour essayer de raconter une histoire convaincante de la vie à ses origines.

Aux origines du vivant, durant la phase prébiotique, ont dû apparaître des cycles métaboliques, des microcosmes contenant des réactions chimiques en interaction, microcosmes contenus dans des membranes ou toutes autres frontières limitant leur dissolution et leurs échanges avec le milieu extérieur et entre microcosmes. La description du plan ou de la fabrique de certains composés

essentiels, facilitateurs de ces flux chimiques, s'est parfois trouvée encodée soit *a priori* soit *a posteriori* sous un autre format plus durable car reproductible et transmissible de microcosmes en microcosmes, sous un format que l'on désigne principalement maintenant comme génétique. J'ai écrit plus tôt que je pensais que seules les choses compliquées à produire sont inscrites sous la forme d'un plan dans le génome. Ce que l'abiotique, la chimie peuvent régénérer (facilement), le vivant n'a pas eu besoin de les sauvegarder.

Une discussion que j'ai eu avec I. Cheddadi⁵⁹ me suggère d'aller un peu plus loin. Le vivant exploite beaucoup la physique des milieux : tension superficielle, pression osmotique, échanges passifs, diffusion, ... Certaines cellules ou éléments cellulaires adoptent des formes particulières, géométriques et simples souvent car principalement produites par des phénomènes physiques comme c'est le cas pour la forme hexagonale des cellules du méristème des plantes ou celles de la morula lors du développement embryonnaire qui résultent principalement d'équilibres de pressions osmotiques et de tensions de surfaces et, dans ces situations, les tissus que ces cellules composent ont un fonctionnement assez « simple », les cellules se « contentant » d'assurer un métabolisme permettant la croissance et l'assimilation / excrétion de matière. Le vivant laisse au monde physique le soin d'assurer une bonne partie de son fonctionnement. Dans ce contexte, les formes particulières parfois obtenues par l'ajout d'une régulation fine mais peu coûteuse sont exploitées comme des transducteurs ou adaptateurs du monde physique. Cela me permet de reformuler un peu cette proposition que j'avais faite dans la section introductive 3.1.2, « *voir le hasard comme fonctionnel et ce qui est ordonné comme un biais sélectionné par l'évolution qui évite à ce hasard de partir dans toutes les directions* » en une phrase incluant les flux, les équilibres physiques, l'abiotique de façon générale et non seulement le hasard. Dans ma vision des choses, ce qui s'est développé depuis les origines et qu'on appelle « la vie » c'est la matérialisation d'un biais, partiellement ordonné, imparfaitement auto-reproductible se présentant sous la forme de molécules et de leurs interactions particulières, qui s'est progressivement imposé de façon de plus en plus complexe au hasard, aux phénomènes physiques purs, abiotiques, à la diversité des réactions chimiques, pour ne produire qu'un nombre restreint de comportements de plus en plus complexes, de plus en plus adaptés à la survie d'organismes dans un environnement changeant, principalement à cause de la vie elle-même. En revanche, dans d'autres situations, le vivant doit assurer des fonctions très complexes et ne peut plus se reposer sur les comportements émergeant de son interaction avec le monde physique. La cellule devient très active et met en œuvre des actions d'une grande complexité faisant intervenir de très nombreux systèmes, lui permettant d'aller même à l'encontre de la physique ; le biotique prend le dessus sur le passif et l'inerte, le biais devient plus fort et le vivant devient à la fois davantage « *intrinsically purposive* », son contrôle interne se renforçant et l'éloignant de la machine qui est « *externally purposive* » [Nicholson, 2013], mais aussi, du point de vue du biologiste, c'est le moment où il devient subjectivement machine car ses comportements sont plus identifiables et reliables à des processus qui font l'objet de nos recherches. Lors de la gastrulation, les cellules qui forment l'invagination différenciant l'ectoderme de l'endoderme s'opposent à la sphère, structure simple aisément maintenue par une activité cellulaire simple. Les cellules immunitaires telles les

59 ... avec qui je partage très agréablement un bureau depuis quelques années...

lymphocytes ne laissent pas la physique du milieu décider de leur devenir ; elles passent de partout entre cellules et tissus, peuvent remonter des flux, forcent des communications à travers tout l'organisme et donnent des ordres chimiques à l'organisme. Dans les deux cas, les fonctions cellulaires mises en œuvre sont complexes.

Le vivant a mémorisé comment remettre en œuvre de tels processus complexes, mais il a aussi mémorisé ce qui lui était essentiel dès les origines : l'exploitation, par le contrôle membranaire et biochimique d'abord, génétique ensuite, du monde physique.

Par l'intermédiaire des plans de fabrication d'éléments biochimiques (la seule chose qui soit réellement encodée dans le génome), ce qui est sauvegardé c'est l'assurance de pouvoir régénérer des systèmes métaboliques et d'interaction biologiques, ce qui permet le fonctionnement de la cellule. À leur tour de tels réseaux biologiques (réseaux génétiques, communication intercellulaires, inter-tissulaires, réseaux de neurones, réseaux écosystémiques / sociétaux), par leur structure, encodent une partie de la coordination nécessaire à une certaine échelle. L'émergence des réseaux d'interaction (donc de systèmes dans lequel les éléments sont « avertis » de l'activité des autres, ce qui permet une certaine coordination de leurs causalités) permet un certain déport d'une tâche complexe de coordination assurant le fonctionnement d'un système, dans un niveau sus-jacent de structuration émergent (le réseau d'interaction).

Un réseau, la mise en interaction d'éléments disjoints, permet la coordination de ses éléments. Un signal émis par un élément est perçu directement ou indirectement par tout ou partie du réseau. Dans le vivant, selon leur nature, les réseaux biologiques sont plus ou moins connectés. Certains éléments d'un réseau perçoivent les signaux de nombreux autres éléments du réseau, voire tous ; ce sont des intégrateurs. D'autres au contraire n'en perçoivent que très peu, voire aucun. Dans le premier cas, les nœuds du réseau sont fortement connectés, leur activité est donc fortement coordonnée avec le reste du réseau, au contraire des nœuds faiblement connectés. Tous les réseaux n'affichent pas nécessairement (à leur échelle de représentation) une coordination totale des éléments. Par exemple, les gènes sont finalement peu connectés entre eux, chaque gène étant « directement » régulé par 0 à 2 gènes (je fais une approximation un peu douteuse là, en utilisant le nombre de facteurs de transcription). De même, dans les réseaux métaboliques, chaque élément (enzymes notamment) voit son activité affectée « directement » par 0 à 2 interacteurs. Si chaque gène devait être averti à chaque instant de l'activité des 30000 autres gènes (une sorte de système expert génétique), il faudrait que la région de régulation de chaque gène comporte des dizaines de milliers de boîtes de contrôle orchestrant savamment l'influence des interacteurs et ajustant ainsi finement la dépendance de chaque gène vis-à-vis de tous les autres. Dans un tel cas, on peine à concevoir toute évolution possible depuis les premiers réseaux d'interaction. Nous savons bien entendu que ce n'est pas le cas. L'intégration est locale ; les éléments du réseau de gènes ne sont avertis de l'activité du reste du réseau que très indirectement, par l'intermédiaire de cascades de réactions par exemple ; ce sont des mécanismes sous-jacents qui opèrent et réalisent cette intégration. On remarquera là que je triche : bien entendu, tout cela ne constitue qu'un seul gigantesque interactome. Cependant, appliquer un niveau de représentation permet justement d'identifier des sous-mécanismes et éventuellement de les

présenter comme interchangeable, tant qu'ils assurent le job ! Subir des mutations, voire des remplacements d'un sous-mécanisme peut s'avérer sans conséquences pour le fonctionnement habituel d'un réseau tout en le dotant potentiellement de nouvelles capacités. Voici une façon de faire évoluer les réseaux d'interaction biologique. Cela rejoint ce dont nous avons discuté dans la section 8.3.5.3, le niveau d'expressivité utilisé pour représenter les réseaux.

En revanche, les neurones dans le cerveau sont beaucoup plus connectés ; chaque neurone peut être averti de (intégrer) l'activité de nombreux autres, parfois des milliers. La connectivité du réseau encode une grande partie de ce qui produit son activité cohérente là où les réseaux de gènes reportent cette activité sur d'autres systèmes, métaboliques notamment, ne faisant qu'orchestrer leur synchronie.

Les réflexions qui découlent des quelques points que j'ai évoqués ici sont relatives à la complexité / connectivité de tels réseaux, à leur plasticité dans le fonctionnement du vivant et leur évolutivité.

Premièrement, étant donné la complexité d'un système intégrateur / de régulation, comment un tel système s'est-il mis en place ? Concernant les réseaux de régulation génétique, on peut raisonnablement supposer que les premiers systèmes vivants n'en disposaient pas ou très peu, que toute la dynamique cellulaire (ou proto-cellulaire) résultait de l'activité, probablement assez cacophonique, de systèmes métaboliques et de la régulation *a posteriori* des quantités de transcrits (et leurs produits) par des mécanismes que l'on retrouve aujourd'hui, c'est-à-dire par interaction directe, compétition et dégradation (ex : miRs). On peut aussi supposer que le niveau de cette cacophonie devait être finalement plutôt modéré si l'on considère d'une part un nombre très limité de systèmes mis en œuvre et, d'autre part, la faible efficacité (spécificité) qu'ils devaient avoir. Cet ensemble plutôt restreint de systèmes à dynamiques lentes (car peu spécifiques) devait pouvoir trouver une certaine harmonie de fonctionnement qui produisait peu-à-peu une proto-vie lente et fragile, mais de la vie. L'augmentation du nombre de systèmes, leur complexité et leur efficacité dynamique n'a pu se faire que par la mise en place de coordinations assurée probablement par des hubs moléculaires d'abord (comme p53), hubs permettant des bascules (switches) d'activité comme le propose justement Kulkarni [Kulkarni, 2020]. L'évolution a tout doucement remplacé de tels systèmes de synchronisation biochimique par l'apparition de régulations à un plus haut niveau, c'est-à-dire au niveau génétique. Cependant, si la coordination de la vie cellulaire a gagné ainsi en complexité, elle a perdu en souplesse d'évolution : la viabilité des organismes devenait dépendante de ces réseaux naissants. Une façon de faire évoluer ceux-ci fut notamment de les conserver et de les encapsuler comme sous-systèmes de plus grands réseaux d'interaction. De plus, le fait que chaque gène ne dépende que de peu d'interacteurs et que de tels gènes pouvaient se trouver dans des réseaux déjà encapsulés, c'est-à-dire contrôlés par des couches sus-jacentes, permettait éventuellement leur remplacement par d'autres, différents, ou leur connexion à d'autres sous-systèmes, tout en limitant l'effet des changements. Cet accroissement de complexité s'est également fait par l'émergence d'échelles encore supérieures, tissulaire, organique, sociale, chaque échelle sus-jacente assurant la synchronisation des échelles sous-jacentes. Chaque échelle sous-jacente devient un élément fonctionnel et remplaçable (les nouvelles cellules remplacent les anciennes, les tissus

peuvent être remplacés, éventuellement artificiellement, *etc*) dans ce qui est un réseau biologique plus grand pouvant être décrit indépendamment de ses composants. Et à chaque fois que l'évolution a fait avancer d'un cran cette complexité organisationnelle et fonctionnelle, elle n'a fait qu'inscrire dans le génome, là où se trouve *in fine* mémorisée l'information vivante, des configurations macroscopiques capables de maintenir de façon encore plus performante des organismes en vie dans des environnements changeants.

Ce que je viens d'exprimer (avec difficulté) ci-dessus, c'est que les réseaux d'interaction biologiques, à une échelle donnée, à l'échelle au dessus, et à l'échelle au dessus encore, *etc*, ont une structure qui organise spatialement et temporellement les interdépendances entre échelles et à travers une même échelle, et tout ceci découle de mécanismes divers d'auto-organisation (à différentes échelles) dont tout ce qui les génère, les constitue, se ramène en fait à ce qui a été mémorisé génétiquement. Or l'ensemble doit, pour ne pas condamner les descendants lointains d'un organisme, permettre une certaine évolution. Par l'usage de ce « pour », je parais sous-entendre là une intentionnalité. Il n'en n'est rien, mais je me demande toutefois s'il n'est pas envisageable que, parmi l'ensemble des systèmes cellulaires, il n'y ait pas de systèmes de contrôle limitant la complexification, des systèmes contrôlant l'évolutivité elle-même, une forme de Lamarckisme finalement, dans laquelle sont mémorisées dans les espèces les bons processus évolutifs, ceux qui, par leur souplesse, ont été autorisés à continuer leur chemin dans l'évolution. Bien entendu, l'alternative connue, solide, reste celle du test évolutif : un système non évolutif car éventuellement trop complexe ou trop simple pourra être simplement éliminé, maintenant ainsi un niveau de complexité de chaque système biologique à son niveau optimal d'évolutivité, ou bien pourra éventuellement être encapsulé dans des couches protectrices de contrôle permettant notamment de module tout changement de tels systèmes ou même leur complexité.

Les réseaux biologiques sont des structures émergeant des interactions des éléments en présence. L'émergence ou la reconstitution d'un réseau lorsque ces éléments sont mis en présence est une manifestation phénotypique qui garde trace, au moins partiellement, de leur trajectoire dans l'évolution, des conditions qui ont été traversées. Ces réseaux, leur structure, leur fonctionnement, leurs interactions avec le monde plus macroscopique ou au contraire avec les échelles plus microscopiques, résultent de sélections évolutives. Ces structures, ce sont des mémoires biologiques qui s'expriment. Le niveau de coordination des éléments du réseau, lui-même, constitue une forme auto-organisée de mémoire évolutive, mémoire non stockée directement sous forme génétique. La fonction du réseau, inscrite dans un fonctionnement écosystémique plus global, sa nécessité vitale, est une contrainte qui s'applique à tout changement à l'échelle de ses éléments.

Si je fais probablement partie des quelques rares personnes qui se posent (encore) la question de l'émergence des premiers réseaux, de leur rôle, de leur faible efficacité néanmoins avantageuse, de la façon dont ils se sont transmis depuis les origines du vivant, je suis certainement loin d'être le seul à me poser la question du rôle qu'a ensuite joué l'évolution dans le niveau de coordination interne des réseaux biologiques, donc dans la sélection de niveaux de connectivités adaptés à la fonction (procurant un avantage évolutif) qu'ils jouent dans un contexte donné. La sélection de telle

ou telle configuration de réseau et des niveaux de connectivité peuvent être vus du point de vue individuel comme la capacité des éléments du réseau à écouter le reste du réseau (nombre de facteurs de transcription capables d'interagir sur un gène ; nombre, sensibilité et fiabilité des organes des sens pour les espèces vivantes ; *etc.*). Dans ce cas l'émergence fortuite de fonctionnalités à l'échelle des éléments du réseau provoque un éventuel changement de structure et de fonctionnement du réseau qui déplace globalement les équilibres dynamiques de l'écosystème dans lequel il jouait un rôle. Là où, dans la masse inextricable des interactions redondantes, solidement contrôlées par des processus garants d'une robustesse importante, mais aussi bénéficiant de la présence de voies alternatives permises par la souplesse des hub moléculaires, on peut imaginer que des mutations d'un système puissent ne provoquer que des changements mineurs, il est difficile d'imaginer que les déplacements d'équilibre de ces écosystèmes biochimiques n'aient pas subi des bascules majeures, changeant leur fonctionnement du tout au tout, lorsqu'ils étaient rares, faibles, plus amorphes, peu spécifiques, à peine en train de s'établir. Je pense qu'il faut voir là l'établissement des premiers systèmes proto-vivants, dont certains étaient peut-être des systèmes vivants autonomes mais qui n'ont pas survécu jusqu'à pouvoir faire partie du consensus LUCA, comme un processus long qui, protégé par des membranes ou en tout cas contenu dans quelques microcosmes que ce soit, a vu des ensembles de systèmes fragiles, peu spécifiques, se connecter, se déconnecter, se complexifier, jusqu'à peu à peu faire émerger des systèmes de plus en plus robustes, de moins en moins sensibles aux changements de configuration locaux.

Les organismes vivants disposent de nombreuses ressources leur permettant, sans évoluer, de s'adapter à un environnement changeant, en migrant, en adaptant leur alimentation, en se protégeant d'un extérieur nocif, ce à quoi il faut ajouter les formes de robustesse qui se manifestent à l'échelle cellulaire, voire moléculaire et qui sont elles-mêmes le fruit d'une longue évolution. De deux choses l'une :

- soit les organismes mutés, adaptés à un environnement possible à venir existent déjà (et sont moins adaptés à l'environnement actuel ; dans ce cas l'évolution s'est faite *a priori* et ces organismes (ou certains de leurs systèmes) constituent en ce cas une potentialité évolutive qui est révélée *a posteriori* par un changement environnemental. C'est vraisemblablement le cas des micro-organismes (et virus) qui se multiplient très vite et produisent ainsi beaucoup de variations.
- Soit ces mutations n'existent pas encore et dans ce cas les organismes résistent à des changements environnementaux, usant de tout ce que les générations précédentes ont accumulé comme systèmes de résilience et de mécanismes de robustesse retardant ainsi l'échéance d'une évolution génétique nécessaire. L'évolution se fait ainsi *a posteriori*, éventuellement sous pression de l'environnement. En effet, certains mécanismes peuvent favoriser l'augmentation de mutations permettant aux organismes de chercher des solutions génétiques compatibles avec un nouvel environnement, en particulier la répression des systèmes de réparation de l'ADN, mais aussi une augmentation (par dérégulation de systèmes de contrôle) d'inducteurs de mutations comme les radicaux libres. Le développement

cancéreux des cellules en est probablement une manifestation.

Dans cette 2ème situation, on peut concevoir ce qui permet la viabilité d'un organisme, son adaptation à un environnement, comme un *trade-off* entre des mécanismes immédiatement mis en œuvre assurant robustesse et résilience, et une évolution génétique, en quelle que sorte un passage d'une adaptation fonctionnelle immédiate à une adaptation structurelle à long terme. Ces deux processus ne se font pas sur les mêmes échelles de temps ; les changements de conditions environnementales ne sont inscrits en dur que si celles-ci durent.

Je trouve, mais je m'avance peut-être un peu trop, qu'il y a une sorte de parallèle à faire entre le passage aux origines du vivant de l'émergence des premières boucles métaboliques, fonctionnelles, aux premiers génomes et le *trade-off* robustesse et résilience vers évolution génétique que je viens d'évoquer.

Dans l'espace des réseaux dont nous avons plusieurs fois parlé (espace défini par rapport à leurs fonctionnalités et structures respectives, et plongé dans des contraintes de viabilités) on peut voir l'évolution précoce des systèmes-réseaux depuis les origines du vivant comme un parcours de réseau (et dynamiques associées) en réseau en direction des centres de noyaux de viabilité. De là, depuis ces ensembles de réseaux initiaux qui ont permis une viabilité suffisante aux premiers organismes proto-cellulaire, l'évolution se dessine comme un ensemble de trajectoires qui s'apparentent à celles de la figure 9.11. Ces réseaux d'interaction biochimiques puis biologiques sont des structures diffuses qui ont peu à peu biaisé, orienté, guidé la domination physique et chimique à l'œuvre aux origines pour en renforcer l'efficacité, la robustesse aussi, ce sont des biais qui ont été mémorisés pour ces raisons. Ils l'ont également été parce qu'ils portaient en eux la capacité d'évoluer, de s'intégrer ou des communiquer avec d'autres systèmes mieux que d'autres. Les réseaux formels, leurs propriétés (complexité structurelle et dynamique en particulier) et les concepts (robustesse, évolutivité, viabilité, atteignabilité) que nous associons à ces objets théoriques et qui leurs donnent une consistance vivante sont une fois de plus un moyen de tester de tels concepts, par des approches de vie artificielle notamment, au même titre que cela fut fait avec Avida en particulier [Adami, 2000]. Les réseaux, par leur structure très fondamentale, leur expressivité de très bas niveau, permettent cependant de se rapprocher de la simplicité moléculaire et de comprendre ces mêmes concepts dans des systèmes extrêmement simples, représentant numériques d'une chimie prébiotique.

10. CONCLUSION

Pour clore cette discussion sur le document que vous avez pris le temps de lire, document qui livre d'une part une introspection concernant mon rapport à la science et une réflexion personnelle sur la nature de la matière vivante, et qui explique d'autre part la façon dont je compte, en pratique, réaliser des recherches pour me convaincre d'abord moi-même puis éventuellement dresser un tableau peut-être un peu moins classique de certains aspects de la vie, je vais commencer par citer le paragraphe introductif d'un article de Carl R. Woese qui arrive à point nommé puisque que je ne l'ai lu que vers la fin de ma rédaction, *A new biology for a new century*, publié en 2004 dans *Microbiology and Molecular Biology Reviews* [Woese, 2004]. Il écrit :

« Science is an endless search for truth. Any representation of reality we develop can only be partial. There is no finality, sometimes no simple best representation. There is only deeper understanding, more revealing and enveloping representations. Scientific advance, then, is a succession of newer representations superseding older ones, either because an older one has run its course and is no longer a reliable guide for a field or because the newer one is more powerful, encompassing, and productive than its predecessor(s).

Science is impelled by two factors, technological advance and a guiding vision (overview). A properly balanced relationship between the two is key to the successful development of a science : without the proper technological advances the road ahead is blocked. Without a guiding vision there is no road ahead ; the science becomes an engineering discipline, concerned with temporal practical problems. »

Carl Woese mit en évidence l'existence d'un 3ème règne du vivant en plus des eucaryotes et des bactéries, celui des *Archae* ; il fut aussi celui qui proposa l'émergence de la vie dans un monde à ARN avant celui de l'ADN, et contribua en 1966 à la découverte et la caractérisation du code génétique. Il fit ces propositions et les défendit à une époque où la biologie moderne et la phylogénie moléculaires étaient encore à l'état embryonnaire. Cet éminent biologiste pourtant pose la question du bien fondé du modèle biologique actuel en vigueur, s'inquiétant notamment la dérive de la recherche biologique vers l'ingénierie du vivant-machine. Cette même inquiétude, Nicholson la décrit, la détaille, notamment en spécifiant ce en quoi le vivant n'est pas machine ; il identifie les sources de cette dérive et en montre les conséquences [Nicholson, 2013 & 2019]. Il explique le danger qu'il y a à persister dans cette facilité et montre notamment que des idéologies (le créationnisme et les mouvements transhumanistes notamment) se saisissent de cette opportunité d'un vocabulaire tout à fait adapté à soutenir la création de la vie comme celle d'une machine, ... utilisant à cet effet le même vocabulaire et la même rhétorique ... que les biologistes ! Même chose pour le transhumanisme qui, évidemment y voit là une argumentation bien pratique leur permettant de suggérer d'améliorer la vie humaine jusqu'à fusionner le cerveau-machine avec la machine-cerveau artificielle et transcender la mort tout en rejetant le contrôle qu'à la vie elle-même sur son évolution.

Le « vivant machine » ... On peut certes continuer à le montrer du doigt comme je l'ai fait moi-même dans ce document, identifier cette dérive, tenter de (se) convaincre qu'il y a un biais de raisonnement évident. On peut aussi tenter de proposer des modèles alternatifs, des modèles qui, par raisonnement, construisent un fonctionnement que n'aurait pas une machine, comme ceux que j'ai proposés pour les réseaux d'interaction biologiques. Mais il ne s'agit pas là d'une mise en évidence de ce qu'est effectivement la cellule vivante (et par extension le vivant). En réponse à cette inquiétude, la question qui vient immédiatement à l'esprit est « si la cellule ou un organisme n'est pas une machine, alors qu'est-ce donc ? ». Il s'agit là bien sûr d'énoncer des faits observationnels sur le vivant, la cellule vivante, en montrant particulièrement en quoi il ne peut pas s'agir d'une machine. A cela, Woese répond « *Let's stop looking at the organism purely as a molecular machine. The machine metaphor certainly provides insights (NDLR : ce à quoi j'ajouterais « au même titre que n'importe quel modèle sur lequel projeter nos observations et hypothèses »), but these come at the price of overlooking much of what biology is. Machines are not made of parts that continually turn over, renew. The organism is. Machines are stable and accurate because they are designed and built to be so. The stability of an organism lies in resilience, the homeostatic capacity to reestablish itself. While a machine is a mere collection of parts, some sort of "sense of the whole" inheres in the organism, a quality that becomes particularly apparent in phenomena such as regeneration in amphibians and certain invertebrates and in the homeorhesis exhibited by developing embryos.* ».

Dans la même veine, Nicholson [Nicholson, 2013, 2019] développe en plus de ces mêmes arguments (qu'il affine considérablement) la notion d'« *intrinsically purposive* » vs « *extrinsically purposive* ». La machine est *extrinsically purposive* : le contrôle de ses actions ne provient pas de l'intérieur ; les organismes vivants possèdent leur propre « libre arbitre » (si j'ose dire) venant de leur contrôle interne. Je modère cependant cette affirmation pour les organismes car ce serait exclure la pression environnementale (environnement physico-chimique et écosystémique) qu'on peut voir comme un contrôle externe sur une longue période de ce que peut faire le vivant. Le vivant modulant à grande échelle l'environnement planétaire (comme ce fut le cas pour la grande oxydation causé par les cyanobactéries ou actuellement pour l'accélération du réchauffement climatique causé par l'activité humaine), on peut considérer l'environnement planétaire et le vivant comme un tout (un écosystème) *intrinsically purposive*⁶⁰.

Il y a bon nombre d'évidences observationnelles, le fonctionnement des protéines et régions intrinsèquement désordonnées par exemple, et de problèmes que pose l'émergence et l'évolution du vivant, qui devraient nous inciter davantage à lutter contre la construction de la machine vivante à partir du réductionnisme moléculaire.

Nous sommes des êtres vivants qui pensent la nature via des machines, via des modèles, des représentations utiles car confortables pour notre esprit. Cela limite ou peut-être, je l'espère, seulement ralentit notre progression vers une vérité utile à la Science, aux humanités, d'une réalité inatteignable.

60 Une discussion plus approfondie à ce sujet se trouve dans le chapitre 11

Mais je crois, comme j'ai cherché à le montrer dans ce document, qu'il est possible de travailler dans une démarche de recherche effective sur ces conceptions moins mécanistes de la vie, et pas seulement en choisissant la voie de la facilité, autrement dit penser librement et écrire de belles idées dans un gros document long à lire, un agrément que j'ai vraiment apprécié durant la rédaction de ce document mais qu'il va me falloir abandonner bientôt pour retourner me confronter à la dure réalité du travail effectif, de la confrontation bibliographique, de l'écriture d'équations, de l'implémentation informatique, du montage de banc expérimental, *etc*, tous engoncés dans les frustrantes contraintes de leur réalisation effective. Il n'empêche qu'il est utile de garder claire à l'esprit cette description particulière du vivant que je propose dans ce document, comme une force de rappel, une incitation à dévier dès et autant que possible, tant que l'on reste dans la démarche scientifique.

Dois-je convaincre que la recherche en biologie doit peu à peu prendre à bras le corps ces questions et aller dans cette direction en citant à nouveau les 3 points de P. Davies concernant l'intérêt de considérer la mécanique quantique en biologie (voir section 8.3.6.1), qu'il faut investir aussi dans des réflexions profondes sur la nature du vivant, incluant des processus rares ou faibles ou amorphes ? S'ils prennent part à la vie cellulaire (ou tissulaire, organique, écosystémique), s'ils constituent un éventuel léger avantage évolutif, alors ils constituent aussi une voie d'abord pour de futures voies thérapeutiques, mais également techno-biologiques, comme des pistes pour le développement de technologies (informatique molle, matériaux ...) inspirées de cette nouvelle compréhension de la biologie.

Ces processus rares, faibles, amorphes, ainsi que les processus abiotiques, physique sous-jacents, sont aussi probablement, comme je me suis essayé à l'exprimer, l'essentiel de ce qu'est le vivant, de ce qu'il a été dès ses origines, bien que depuis se sont amoncelés des biais successifs, imbriqués, sous la forme de contrôles, de feedbacks, de feed-forwards, biochimiques, enzymatiques, génétiques, des systèmes qui augmentent la diversité des comportements (incluant d'autres contrôles, des systèmes de résilience, *etc*) favorisant la pluralité des adéquations possibles d'un organisme avec des environnements qu'il est susceptible de rencontrer. Ces contrôles sont aussi une fragilité que compense au moins partiellement le fait qu'ils soient composés non pas de pièces mécaniques mais de biomolécules, souples, versatiles, multifonctionnelles, aux interactions imparfaites et pas nécessairement fréquentes.

Adami affirme « *An organism's DNA thus is not only a "book" about the organism, but it is also a book about the environment it lives including the species it co-evolves with.* », rejoignant ce que Deutsch dit de façon plus compacte « *Genes embody knowledge about their niches* », celui-ci allant jusqu'à dire que « *Everything of fundamental significance about the phenomenon of life depends on this property [NDR : the embodying], and not on replication per se* » [Adami, 2000 ; Deutsch, 1997]⁶¹. Au cours de mon document j'ai montré que je soutenais cette idée de mémorisation des environnements traversés au cours de l'évolution par la trajectoire que des espèces ont suivies jusqu'à un certain organisme, plaçant cette mémoire à plusieurs niveaux, du génome même et de sa biochimie, à l'émergence des systèmes biologiques, mais aussi à la partie effectrice et de sa

61 À noter que je n'ai fait que parcourir ce livre de Deutsch, faute de temps vu que je l'ai découvert très tardivement, me contentant de lire en détails les quelques pages d'où j'ai extrait ces citations.

configuration qui est transmise lors de la reproduction. Je pense aussi que c'est cela l'essence du vivant ; ça et la contrainte intrinsèque de rester évolutif « pour » que les générations futures puissent continuer à atteindre les configurations qui lui permettent de continuer à vivre dans un environnement changeant. Je pense que les cellules vivantes contiennent intrinsèquement cette caractéristique d'évolutivité à la fois comme une propriété innée liée aux propriétés de la matière vivante, en particulier à ses imperfections, son caractère versatile, mais aussi comme un caractère héritable. Je propose que plus que la mémorisation des conditions de vie passée, le vivant puisse mémoriser une certaine évolutivité éventuellement sous la forme d'une maintenance de certains niveaux de structuration, de complexité de ses composants ou systèmes. Là où Adami affirme, à juste titre, que « *A true test for whether or not a sequence is information uses the success (fitness) of its bearer in its environment, which implies that a sequence's information content is conditional on the environment it is to be interpreted within.* » je pense cependant que le test environnemental seul échoue à valider complètement l'aspect informationnel d'une séquence génétique car ce test n'inclut pas tout ce qui dans les séquences, dans la structuration des protéines et des réseaux qui émergent de ces effecteurs, constitue un encodage de l'évolutivité. En revanche, c'est au cours du temps, à l'échelle des espèces, de l'évolution, que la capacité à garder atteignables des comportements qui permettent de conserver une viabilité acceptable malgré des changements environnementaux, peut être éprouvée.

J'ai exposé, en m'appuyant sur nos expériences numériques avec les réseaux, comment la complexité croissante des systèmes biologiques auto-contraint leur évolution, restreignant sur les temps « courts » leur capacité exploratoire. J'ai aussi expliqué pourquoi, modulo l'heureux ajout apporté par l'aléatoire (mutations, transpositions, *etc*), les paris évolutifs que constituent les lignées se trouvent conditionnées par l'existant, depuis ses origines. La vie ne pourrait être ce qu'elle est avec sa diversité et la diversité des conditions auxquelles elle s'est confrontée si elle ne se laissait pas autant perfuser par le monde physique qui lui apporte l'aléatoire bien sûr, mais aussi les flux passifs de matière, des potentiels naturels, des directions, des champs, *etc*.

C'est donc ainsi que je vais conclure les réflexions sur la recherche dans ce document : en rappelant que des recherches, potentiellement de bons sujets de thèse, peuvent être menée dans ce cadre conceptuel large, des recherches relatives à l'origine de la vie, à l'évolution, à la balance qui oppose exploration et gradation, aux conséquences de la présence de reliques du passé dans les systèmes actuels, mais aussi à des réflexions sur le calcul non conventionnel. Pour cela, je vais m'appuyer sur les longs développements que j'ai faits, en particulier sur les réseaux, et étendre les réflexions que j'ai entamées dans les chapitres précédents à des réflexions plus générales encore.

Dans « *La Vie et la Matière* », Oliver Lodge [Lodge, 1907] rappelle le danger pour les scientifiques d'aller trop loin dans la démarche philosophique en sciences : « [...] *l'expérience leur [les philosophes] a montré que les hommes de science, après avoir escaladé la barrière, sont quelquefois enclins à perdre toute notion de contrainte et à se conduire comme s'ils avaient enfin pénétré dans une région libre de toute entrave scientifique, où les assertions téméraires peuvent être*

impunément faites, où les hypothèses spéculatives peuvent être appelées des théories et où la vérification est à la fois inutile et impossible. ». Il montre qu'Haeckel (dont nous avons parlé dans la section 8) avec sa théorie des monades s'est ainsi perdu.

Je m'inquiète évidemment de cela pour ceux, moi compris, qui explorent les concepts liés au vivant dans cette approche que l'on désigne, finalement peut-être pompeusement, par « biologie théorique », pompeusement car accoler « biologie » et « théorique » pour ne former qu'un seul mot suggère quelque peu que le théoricien, restant dans le confort théorique, puisse se permettre la dispense de la confrontation à la réalité mesurable du monde pourtant flou, mou, informe, imprécis de la biologie, l'autorisant éventuellement à aller se perdre dans la réflexion théorique pure, dans un monde qui n'est pas vraiment, ou en tout cas dont on ne pourra peut-être pas vérifier la véracité.

Les concepts du vivant (dont la nature réelle est insaisissable) comme la complexité, la viabilité, la robustesse, l'évolutivité, pour ne prendre que les quelques concepts auxquels je me suis intéressé, sont nécessairement subjectifs ; ils obéissent à un besoin du chercheur souhaitant vérifier l'adéquation de ce qu'il mesure du vivant avec les modèles (biaisés par sa nature humaine) qu'il a en tête ; il y a ainsi une inadéquation entre l'outil conceptuel et la réalité objective. La spécification de ces termes est néanmoins nécessaire pour avancer en sciences : l'objectif est de parvenir à définir des cadres conceptuels pour faire des observations et réaliser des expériences pour nous aider à appréhender divers aspects du vivant. Leur définition doit répondre à un impératif de compréhension. De la même manière que l'expérimentateur va concevoir une expérience particulière pour réaliser une coupe, une vue particulière sur le système qu'il veut étudier, lui donnant accès à une part de compréhension, le design de concepts de biologie théorique est une forme de design expérimental : un concept avec sa formalisation associée constitue un outil pour étudier un système. La biologie théorique et ses outils conceptuels évoluent au fil des nouvelles connaissances (d'origine expérimentales, théoriques, philosophiques, ...) qui affinent la compréhension que l'on a du vivant. Il faut néanmoins que les définitions utilisées en biologie théorique restent fixes un certain temps, pour leur permettre d'être testées et produire des connaissances. Je m'aperçois et je crois que cela se voit quelque peu dans mon manuscrit que mon esprit assez prolifique en terme d'idées n'est pas forcément idéal pour maintenir une stabilité. C'est tout l'intérêt de travailler en groupe, entre collègues et étudiants qui constituent des forces de rappels et des impératifs imposant une certaine stabilité.

J'ai montré dans ce document que je m'interroge beaucoup sur la nature du Vivant et ce, par des incursions dans le monde de la philosophie des sciences. J'ai aussi montré mon attirance pour ces objets théoriques fascinants (ex : les réseaux Booléens, les systèmes auto-organisés, *etc*) et ce qu'ils inspirent et permettent d'inférer sur le monde vivant. J'ai montré que j'extirpais de cette fascination ce type d'inférences (ex : les *réseaux pluripotents* ou les canaux de communication de faible entropie) dont on peut largement douter de la possibilité d'une quelconque vérification biologique.

J'en suis conscient.

Je pense déjà qu'il faut, en tant que scientifique, s'autoriser quelques incursions prudentes dans le monde de la philosophie (sinon, à quoi cela sert-il d'accoler à notre nom le terme Ph.D dans notre signature ?). Par « prudentes » j'entends déjà qu'il faut au moins accepter que l'on ne dispose certainement pas systématiquement du bagage philosophique nécessaire pour appréhender de telles incursions. En être conscient invite à la prudence. J'entends aussi qu'il faut être proactif dans ce sens : il faut lire les penseurs et l'histoire des sciences, riches en enseignements ; il faut aussi montrer cette nécessité à nos étudiants et leur transmettre ses fondements. C'est ce que j'ai fait récemment en donnant un cours introductif fort plaisant sur le numérique dans le monde de la biologie et de la santé, cours durant lequel j'ai passé deux heures et demi à parler de la difficulté, si ce n'est l'impossibilité, de percevoir objectivement le monde pour faire de la modélisation, montrant que notre perception est orientée par nos savoirs, par les modèles préconçus (construits peu à peu par le progrès scientifique) que nous avons du monde et sur lesquels nous projetons notre perception. C'est aussi ce que je pousse à faire en transmettant mes pensées à travers ce document, que j'ai presque peur de relire maintenant, aux futurs étudiants.

Je pense aussi que, proposer de nouveaux concepts, des visions alternatives du monde, tout ceci issu de la pensée (qui n'est certes pas vierge, chez le scientifique, de connaissances scientifiques sur le monde) est acceptable. De même, l'utilisation d'objets théoriques, mathématiques ou informatiques, pour tester des mondes ou concepts théoriques que notre esprit associe au vivant a un rôle important : cela crée chez celui qui s'y adonne un éventail de nouveaux modèles, de nouveaux concepts, sur lesquels il pourra, s'il est raisonnable et revient régulièrement au monde réel, à la mesure, et ne brise pas le dialogue indispensable avec ceux qui mesurent, ceux qui expérimentent, testent, *etc*, projeter les perceptions qu'il aura ou qu'on lui proposera sur le vivant.

Pour faire avancer la connaissance scientifique, mis à part espérer mesurer objectivement le monde, c'est finalement ce dialogue interne et inter-individuel entre perception du monde et objets théoriques qui permet aux hommes de se faire une idée de ce qu'est le monde, le monde vivant en particulier.

En ce qui me concerne, je commence à peine à me forger quelques petites idées sur ce qu'il est, sur l'immensité qu'il nous reste à découvrir, me laissant un peu devant un abîme que je peine à combler avec mes propositions. J'aime bien ce que je pense mais, dans ce fouillis que sont mes pensées, je doute de parvenir un jour à délivrer un message qui me convienne. Certes frustré, j'en aurais au moins livré quelques bribes. C'est déjà pas mal, non ?!

11. RAPPORTS DES ÉVALUATEURS – ANALYSE A POSTERIORI

Vous trouverez ci-dessous les rapports des 3 rapporteurs de cette Habilitation à Diriger des Recherches. Je les joins car ils apportent à la réflexion et car j’y apporte des réponses.

11.1. Rapports

Les 3 rapports que vous trouverez dans les pages suivantes sont des captures d’écran des fichiers PDFs qui m’ont été envoyés. J’ai pris le soin d’effacer les signatures.

La présente version de ce manuscrit a été révisée en tenant compte des remarques et corrections proposées par les rapporteurs. Quelques ajustements ont été faits dans le texte initial, cependant l’essentiel se trouve dans la réponse que je fais sections 11.2.

11.1.1. Rapport de Patrick Amar, Université Paris-Saclay**Rapport sur le manuscrit d'HDR présenté par Nicolas Glade intitulé :****«Le Vivant Rare, Faible et Amorphe****Evolution depuis les Origines jusqu'à la Vie telle qu'elle nous Apparaît»**

Le manuscrit d'HDR de Nicolas Glade comporte une introduction générale à ses thèmes de recherche, un CV long, puis la présentation détaillée de ses recherches antérieures et actuelles.

Son CV montre clairement qu'il possède toutes les qualités qu'on attend d'un futur Professeur des Universités. Nicolas a démontré sa capacité à encadrer et co-encadrer plusieurs doctorants. Il a défini le sujet et encadré une dizaine de stagiaires de Master 1 et 2. Il a contribué à obtenir de nombreux financements pour ses recherches. Il s'est aussi très impliqué dans ses enseignements et plus généralement dans l'administration de l'enseignement.

Dans son manuscrit, Nicolas Glade présente tout d'abord ses travaux de thèse ainsi que ses recherches au cours de ses post-doctorats et postes d'ATER. Il a travaillé essentiellement sur les microtubules, leur auto-organisation et notamment sur l'influence des champs faibles tels que la gravitation, les champs magnétiques et électromagnétiques. Lors de son premier post-doctorat, en tant qu'ingénieur de recherches, il a mené des expérimentations en apesanteur réalisées à l'aide d'une fusée.

Durant son deuxième post-doctorat il a travaillé sur l'adhésion cellulaire sur des substrats synthétiques. Il a déposé un brevet sur la localisation d'espèces biochimiques.

Son troisième post-doctorat l'a conduit à aborder des thèmes de modélisation, notamment celle de réseaux métaboliques, de réseaux génétiques et de leurs couplage.

Il s'est intéressé ensuite à la psychomotricité, en participant au développement d'un dispositif de substitution de la vision dans l'espace. Lors de l'encadrement d'un doctorant, Nicolas s'est intéressé à la modélisation multi-échelle en épidémiologie.

Ses travaux plus récents sont ensuite détaillés dans un grand chapitre, puis suit une discussion scientifique et philosophique sur les origines de la vie et l'évolution.

Nicolas commence par poser des questions sur la pertinence des modèles actuels du vivant : le biais de l'ingénierie nous poussant à considérer une cellule comme une *machine* moléculaire. Les observations de plus en plus précises apportées grâce aux progrès de la technologie conduisant généralement à modéliser la cellule de la façon dont on concevrait une machine. Sans nier les progrès évidents qui ont été faits justement grâce à la techno-

logie, il propose une autre approche. Il donne des exemples d'interprétation initialement "mécanistique" qui ont depuis évolué vers des modèles beaucoup plus nuancés.

Il présente ensuite un de ses thèmes de recherche : l'origine du vivant. Il fait un état de l'art bien documenté et très bien écrit, en rappelant les théories et les expérimentations permettant de les justifier ou de les mettre en défaut. Il présente notamment trois théories sur l'origine du code génétique et discute de leurs pertinences. Il montre l'utilité des simulations informatiques pour proposer ou invalider des modèles, notamment sur l'auto-organisation ; ou tout simplement pour tester des idées. Il a proposé un projet de recherche en ce sens, comportant un volet expérimental et un volet plus théorique.

Il s'intéresse dans le chapitre suivant de son manuscrit à l'étude des réseaux biologiques et leur évolution. Pour les étudier d'un point de vue théorique, il présente de façon très détaillée les réseaux booléens à seuils et à signe. Il explique comment il met en relation les réseaux d'interactions biologiques et leur modélisation avec des réseaux booléens (ou des ensembles de réseaux booléens) pour en déterminer les comportements possibles. Le but étant de trouver des modèles qui satisferont les connaissances, même partielles, que l'on a du système biologique étudié.

Nicolas présente ensuite un système de communication à mémoire éphémère, les *trail systems*, procédé où une entité laisse une trace de son passage dans l'environnement. Trace éphémère qui, si elles arrivent à temps, sera *ressentie*, utilisée et renforcée par d'autres entités. Il montre, à différentes échelles, des exemples de tels systèmes. Il fait l'hypothèse de l'existence de trail systems intracellulaires constitués de microtubules, effectifs dans certaines conditions. Il envisage aussi le cytosquelette comme perturbant la diffusion des molécules, induisant des inhomogénéités de concentrations selon la localisation spatiale. Dans cette même veine, il émet l'hypothèse d'une possible propagation rapide d'information le long des fibres du cytosquelette par changement de conformation de la tubuline.

Dans son dernier chapitre, Nicolas revient sur ses réseaux booléens en s'intéressant aux combinaisons de réseaux pouvant selon les cas exhiber les caractéristiques dynamiques de chacun des réseaux constituants. Il introduit une notion de distance entre les réseaux participant à l'ensemble ; distance structurelle et distance entre les dynamiques des réseaux. Dans le contexte de l'évolution, il définit ensuite les notions de robustesse, d'évolutivité et de viabilité étudiées à travers le prisme des réseaux formels. Il introduit enfin l'héritage et l'atavisme comme faisant partie de l'évolution. Les "sous-systèmes" constituant les cellules vivantes ne sont pas apparues en même temps, certains sont beaucoup plus anciens que d'autres. Ils peuvent être inopérants dans les conditions normales, mais être réactivés si l'environnement change. À la fin de ce chapitre, Nicolas revient sur les origines du vivant par le biais de la vie artificielle. Il montre plusieurs exemples de programmes faisant émerger

des fonctions nouvelles à partir d'éléments de base.

Nicolas Glade est l'auteur de nombreuses publications dans des revues et conférences internationales de très bon niveau. Il a apporté un éclairage nouveau sur la modélisation du vivant, d'une part par ses réflexions sur ce qui caractérise le vivant, de l'individu à la lignée et son évolution, et d'autre part par les moyens techniques (réseaux booléens et simulations) qu'il a mis en oeuvre dans ses modélisations.

Son manuscrit fait clairement apparaître sa maîtrise des formalismes utilisés, de la biochimie et de la génétique, ainsi que de l'algorithmique nécessaire pour concevoir et réaliser les divers outils logiciels qu'a contribué à développer. Je considère que la qualité de ses travaux, ainsi que celle de la rédaction du manuscrit, justifient largement sa soutenance pour l'obtention de l'habilitation à diriger des recherches.

Fait à Orsay, le 30 septembre 2022

Patrick AMAR
Maître de Conférences HC, HDR
LISN, CNRS / Université Paris-Saclay

11.1.2. Rapport de Hughes Berry, INRIA



Lyon, 07 octobre 2022

Rapport sur le mémoire et les travaux de Mr Nicolas GLADE en vue de l'obtention de l'Habilitation à Diriger des Recherches de l'Université de Grenoble Alpes.

Rapporteur : Hugues Berry, INRIA, Université Claude-Bernard Lyon 1, France

Dans son mémoire d'HDR intitulé "Le vivant rare, faible et amorphe : évolution depuis les origines jusqu'à la vie telle qu'elle nous apparaît", Nicolas Glade rapporte un travail de recherche multidisciplinaire et courageux, qu'on peut qualifier de biologie théorique, sur la nature du vivant et certains des paradigmes actuels de la biologie (réseaux, code génétique, le vivant comme une machine). Au-delà d'un exposé de sa trajectoire de chercheur, qui embrasse ici la totalité de sa carrière, depuis la thèse jusqu'à ses perspectives de recherche actuelles, il s'agit aussi d'un document de réflexion, empruntant souvent à la philosophie des sciences, dans lequel il présente sa vision des principales questions scientifiques qui l'animent : les origines du vivant et son évolution, en particulier par le prisme du code génétique, de la complexité et de leur modélisation via les réseaux booléens ou les systèmes autoorganisés.

Avant de donner les raisons de mon opinion positive sur le travail présenté, il ne m'est pas possible de taire le principal défaut du document, qui est selon moi son manque de concision. Le texte est trop long (333 pages) et j'ai trouvé de nombreuses sections verbeuses. La description des travaux de recherche *actuels* de Nicolas Glade occupe à elle seule pas moins de 130 pages. La section portant sur les réseaux booléens s'étend sur 50 pages, alors que ce sujet ne correspond finalement qu'à quelques articles de journal et de conférence dans sa liste de publications. Autre exemple, le document utilise 15 pages pour simplement décrire les méthodes expérimentales relativement classiques que Nicolas Glade compte utiliser pour mesurer les interactions codons-acides aminés. Cette verbosité excessive selon moi finit malheureusement par occulter en partie la qualité des propositions qui figurent pourtant dans le document.

Car le document présente indubitablement, selon moi, des réalisations et points de vues originaux et intéressants. Il débute par une introduction qui donne une perspective historique sur le code génétique et l'interprétation du vivant comme une machine. La partie suivante fournit le CV académique de Nicolas Glade (production d'articles, enseignement, encadrement). Ses travaux de recherche passés sont présentés dans la section 7, avec une place importante donnée à l'auto-organisation des microtubules, sujet de sa thèse, et son extension au calcul non conventionnel. J'ai apprécié l'honnêteté scientifique de Nicolas Glade, qui n'a pas hésité à rejeter l'interprétation que son équipe avait initialement faite du phénomène comme résultant d'un processus de réaction-diffusion, lorsqu'il a compris qu'elle n'était pas réaliste.

La section 8 présente le travail de recherche actuel de Nicolas Glade, et les questions qui le motivent, qui portent principalement sur les origines du code génétique. Il y développe en particulier des idées originales et intéressantes sur le rôle des petits métabolites (ATP, GTP) dans la régulation du code génétique. Mais la plus grande partie de cette section concerne les

réseaux booléens (à seuil, principalement). Nicolas Glade présente un travail théorique original dans lequel il propose en particulier de ne pas se focaliser sur un réseau unique pour modéliser un phénomène biologique donné, mais une *population de réseaux* exhibant les propriétés structurales et dynamiques du phénomène modélisé. Il y aborde aussi la complexité de ces réseaux, leur comparaison et leur évolution. Cet ensemble de résultats forme une somme originale et une perspective exaltante pour l'étude théorique des relations structure-fonction des réseaux de régulation car il introduit le concept d'un espace de réseaux, voire de réseaux de réseaux.

Enfin la dernière section présente une série de perspectives, liées principalement à l'étude de l'évolution des réseaux et leur capacité à faire émerger de nouvelles fonctions. Ces perspectives sont présentées sans vraiment de lien apparent entre elles mais j'ai là aussi apprécié l'originalité de ces pistes de recherche. En particulier, tout le travail effectué précédemment sur les réseaux booléens est utilisé ici pour proposer des systèmes de vie artificielle originaux pour étudier l'émergence de nouveaux comportements et fonctions. Cela me semble être une perspective potentiellement fertile pour l'étude théorique de l'origine et de l'évolution du vivant.

Tout au long du manuscrit, j'ai été frappé par le soin apporté par Nicolas Glade à mettre en avant les contributions des nombreux étudiants qu'il a encadrés (dont 13 masters, 4 doctorant.e.s et un postdoc). Cette attitude, ainsi que la qualité du travail fourni par nombre de ces étudiants (attesté par leur production d'articles), me semble un bon indicateur de la qualité de l'encadrement fourni par Nicolas Glade. Sa production scientifique est très notable (60 articles publiés, dont 34 dans des revues) même si sa forte implication pédagogique récente a visiblement ralenti sa productivité en termes de publications. Il a par ailleurs plusieurs responsabilités académiques très prenantes (responsable d'année, codirection de M2).

Mon opinion est donc que Nicolas Glade montre à travers son manuscrit d'HDR un travail de recherche et de réflexion à la fois novateur et multidisciplinaire et qui ouvre la voie pour de nombreuses études à venir. Par conséquent, je considère que Nicolas Glade a largement acquis l'expérience et la maturité exigées pour l'obtention de l'Habilitation à Diriger des Recherches.

Hugues BERRY

Directeur de recherche INRIA

Directeur scientifique adjoint d'Inria pour biologie numérique et santé numérique

Centre de recherche Inria de Lyon

56 Blvd Niels Bohr - CS 52132 , 69603 Villeurbanne cedex France

hugues.berry@inria.fr - www.inrialpes.fr/Berry

INRIA - LYON RESEARCH CENTER

Batiment CEI-2 - 56 Blvd Niels Bohr - CS 52132

69603 Villeurbanne Cedex - France

Phone: +33 (0)4 72 43 74 90

www.inria.fr

2/2

11.1.3. Rapport de *Francesca Merlin*, CNRS

IHPST

institut d'histoire et de philosophie
des sciences et des techniques13, rue du Four
75006 Paris☎ 33 (0)1 43 54 60 36
☎ 33 (0)1 43 54 94 60
☎ 33 (0)1 43 25 29 48✉ ihpst@univ-paris1.fr
www-ihpst.univ-paris1.fr

Paris, le 8 novembre 2022

Rapport sur le mémoire de M. Nicolas GLADE en vue de l'obtention de l'Habilitation à diriger des recherches de l'Université Grenoble AlpesRapportrice : Francesca Merlin, IHPST, CNRS & Université Paris 1 Panthéon-Sorbonne

« Le Vivant Rare, Faible et Amorphe : évolution depuis les origines jusqu'à la vie telle qu'elle nous apparaît » est un mémoire de taille (333 pages) dans lequel Nicolas Glade présente, dans l'ordre : 1) sa vision du vivant, de la biologie, et plus généralement de la science ; 2) son parcours scientifique, en consacrant une partie importante à ses activités d'enseignement et d'encadrement ; 2) ses recherches passées (notamment, mais pas seulement, sur l'organisation des microtubules), suivies par ses travaux actuels et futurs (sur la question de l'origine de la vie et plus précisément sur la mise en place d'une ébauche de système génétique, ou encore sur les réseaux Booléens en tant qu'outils théoriques pertinents permettant de représenter les systèmes biologiques) ; 3) une ouverture vers des réflexions nouvelles au sujet de l'évolution, l'atavisme, et le faible, le rare et l'amorphe comme caractéristiques propres du vivant ; 4) une conclusion qui revient sur les considérations théoriques et méthodologiques faites dans l'introduction du manuscrit.

En tant que philosophe des sciences, ce n'est pas mon rôle d'évaluer les résultats scientifiques produits par Nicolas Glade pendant toutes ces années de recherche, tâche pour laquelle je manque aussi d'un certain nombre de compétences. Je peux néanmoins remarquer deux choses au sujet de son dossier. D'une part, la production scientifique de Nicolas Glade est notable, bien qu'elle ait baissé plus récemment à cause de son implication dans des activités pédagogiques et de la prise en charge de plusieurs responsabilités académiques. D'autre part, Nicolas Glade donne une place très importante, et même centrale, aux étudiants et étudiantes, non seulement par l'importance qu'il donne à leur formation mais aussi et surtout par le rôle qu'il leur attribue dans la

construction de sa propre recherche scientifique. C'est là une caractéristique centrale du dossier de Nicolas Glade, que je tiens à remarquer et à souligner, et qui ne peut qu'appuyer mon avis positif quant à la possibilité qu'il obtienne l'Habilitation à diriger des recherches.

La qualité principale et saillante du parcours scientifique et intellectuel du candidat est la place que ce dernier donne à la réflexion philosophique dans ses travaux. Ces derniers peuvent en effet être caractérisés comme s'insérant dans la biologie théorique, domaine dont les frontières avec la philosophie, notamment la philosophie de la biologie, sont floues et mouvantes. Nicolas Glade montre avoir toujours le (bon) réflexe de prendre du recul par rapport à ses objets d'étude, qu'il s'agisse de l'origine de la vie, de la représentation des systèmes biologiques ou encore de leur simulation. Son objectif est de saisir d'où vient la conception qu'on a du vivant, la façon dont on le représente, les méthodes qu'on utilise pour l'étudier, mais aussi de comprendre quelles sont les conséquences de notre vision du monde biologique, et ses limites aussi. Autrement dit, l'objectif de Nicolas Glade semble être, et cela de manière systématique, de reconstruire une vision d'ensemble (« the big picture ») du monde biologique, ce qui n'est rien d'autre que l'objectif propre à toute recherche philosophique sur le vivant. Plus généralement, comme le disait le philosophe Wilfrid Sellars (1962), "The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term". Nicolas Glade n'est pas philosophe et ne vise pas à une telle compréhension, globale, du monde. Cependant, tout au long de son cheminement scientifique et dans ses projets de recherche pour le futur, il dégage une nécessité irrésistible de questionner continuellement le bien fondé de nos présupposés, de nos conceptions, des paradigmes en vigueur, pour viser *in fine* à une vision plus pertinente et englobante du vivant et de la biologie. C'est donc sur le fait que ses recherches et réflexions sont en partie de nature philosophique, ce qui est relativement rare auprès des biologistes, que porte mon appréciation de son mémoire. En le lisant, j'ai remarqué la qualité de plusieurs questionnements qui l'animent, notamment sur des thématiques comme le rôle et de l'influence du paradigme mécaniciste sur notre compréhension du vivant, sur le rôle des biais en biologie et le caractère toujours partiel et perspectiviste de toute connaissance scientifique, ou encore sur les limites de la vision actuelle, prépondérante, du vivant (plus particulièrement de la cellule en tant que machine) par rapport à certains phénomènes qui s'en écartent, notamment les phénomènes rares, faibles et amorphes. La qualité des réflexions de Nicolas Glade ne serait pas en elle-même une vertu si elle n'était pas accompagnée par un effort constant vers la résolution des questions soulevées. Par exemple, se questionnant sur le problème de l'origine de la vie, Nicolas Glade avance des suggestions originales sur la régulation du code génétique ; ou encore s'interrogeant sur la modélisation des systèmes biologiques, il propose une forme originale pour les représenter en termes de populations ou d'ensembles de réseaux. Son

élan l'amène toujours vers une tentative de solution, vers la mise à l'épreuve d'une nouvelle perspective, qu'il élabore et il essaie de tester expérimentalement.

Je remercie donc très chaleureusement Nicolas Glade de m'avoir invitée à faire partie de son jury d'HDR, ce qui me permet de réfléchir davantage sur toutes les questions qu'il soulève dans son mémoire, qui sont passionnantes à mes yeux car fortement philosophiques.

Mes questions et remarques portent sur les aspects philosophiques du mémoire d'HDR et plus généralement des recherches de Nicolas Glade, et notamment sur les six thématiques suivantes :

1-Ce que Nicolas Glade appelle « subjectivité » (et que j'appellerais plutôt « perspective »).

2-La question du réalisme de nos modèles.

3-La question du pluralisme scientifique (notamment, le positionnement de Nicolas Glade par rapport au paradigme en vigueur et son rapport avec la nouvelle vision (alternative ou additionnelle) qu'il propose).

4-Le statut des simulations numériques (qu'est-ce que c'est et quel est leur rôle).

5-La relation entre mécanisme et holisme et entre mécanisme et hasard.

6-La notion de finalité intrinsèque (à ce sujet, et plus généralement au sujet de la question de ce qu'est le vivant et de comment il se distingue du non-vivant, je suis étonnée de ne voir aucune mention du livre de Jacques Monod « Le hasard et la nécessité », en particulier du premier chapitre de ce livre où Monod se pose cette question et introduit la notion de « projet téléonomique » par laquelle il caractérise la spécificité du vivant).

1-La « subjectivité » en science

Dès l'introduction de son mémoire d'HDR, Nicolas Glade parle de la « subjectivité en sciences » (biologiques en particulier) qui caractérise le paradigme du vivant-machine, comme tout autre paradigme, de la biologie contemporaine. Par « subjectivité » Nicolas Glade fait référence à un biais qui est l'œuvre, de plusieurs biais mêmes, dans le choix à la fois des outils pour observer et mesurer le vivant et des formalismes pour le représenter. Or, « subjectivité » me semble un terme trop fort car il s'oppose à l'objectivité et car il fait le plus souvent référence à une préférence personnelle d'un sujet individuel. C'est pourquoi je parlerais plutôt de « perspective » en faisant référence à la littérature, depuis

quelques années abondantes en philosophie des sciences, sur le perspectivisme en science (Ronald Giere, Baas Van Fraassen, Bill Wimsatt, etc.). De manière générale, c'est l'idée selon laquelle la science ne peut pas dépasser toute perspective humaine (autrement dit, le point de vue de nulle part, comme l'appelait Ernest Nagel, n'est pas accessible, mieux, n'existe pas), et cela a un impact sur tout un ensemble de questions philosophiques que l'on se pose sur la science (notamment, le réalisme des modèles et des théories, ou encore la nature de l'objectivité scientifique). Je préfère le terme de « perspective » à celui de « subjectivité » d'autant plus que le fait d'adopter toujours, nécessairement, une perspective lorsqu'on étudie des phénomènes naturels n'est pas en tension avec l'idée d'objectivité, dans nos pratiques ou procédures scientifiques et dans les résultats que l'on obtient. D'ailleurs, la vision alternative proposée par Nicolas Glade, focalisée sur les phénomènes biologiques rares, faibles, amorphes, caractérisés par peu d'organisation, peu de stabilité, peu d'ordre, est aussi une perspective par laquelle il veut interroger le monde biologique et produire de la science (des modèles, des simulations, des explications, des prédictions) à son sujet.

Une autre façon d'exprimer ce biais dont nous parle Nicolas Glade, mais qui me semble plus éloignée de ce qu'il veut signifier par « subjectivité », est le « point de vue » dont nous parlent les épistémologues féministes. Je fais référence à la théorie du point de vue (Standpoint Theory, cf. Nancy Hartsock, Sandra Harding) et à l'idée de « connaissance située » élaborée par Donna Haraway selon laquelle le sujet connaissant est toujours situé dans des relations particulières par rapport à d'autres connaissances et à d'autres sujets connaissant. Autrement dit, ce qui est connu et la manière dont il est connu reflètent la perspective partielle du sujet connaissant, cette dernière étant due à plusieurs éléments le caractérisant (corps, genre, race, orientation sexuelle, rôle social, etc.). La connaissance est ainsi toujours limitée, locale (par rapport à un certain emplacement culturel, social, politique) et le connaissant est toujours socialement situé. On voit ainsi que, non seulement des éléments épistémiques entrent en jeu lorsqu'on fait de la science, mais aussi des valeurs non épistémiques ou contextuelles, qui influencent la pratique scientifique et l'élaboration des connaissances. Je ne pense pas que Nicolas Glade compte aller jusque-là dans sa réflexion sur les biais en sciences. Je trouve cependant que la réflexion des philosophes féministes des sciences peut permettre d'aller plus loin dans l'analyse des raisons pour lesquelles certains paradigmes plutôt que d'autres s'imposent à certains moments de l'histoire des sciences et de leurs conséquences sur les pratiques scientifiques, théoriques et expérimentales.

2-Le réalisme scientifique

J'aimerais en savoir plus sur le positionnement de Nicolas Glade sur le réalisme (ou pas) des modèles du vivant dans le cadre du paradigme vivant-machine en vigueur.

Malgré leur part de « subjectivité » (comme dirait Nicolas Glade) - autrement dit, malgré le fait qu'il s'agit de modèles toujours partiels car fruits d'abstractions et de simplifications par rapport au monde biologique -, s'agit-il de modèles que l'on peut interpréter de manière réaliste, à savoir comme des descriptions en partie vraies des systèmes vivants, du moins au sujet de certaines de leurs propriétés ? Ou, au contraire, Nicolas Glade considère qu'on ne peut pas être réaliste au sujet de ces modèles à cause de leurs biais et leurs limitations de perspective, et qu'on ne peut donc en avoir qu'une interprétation instrumentaliste, c'est-à-dire les considérer comme des outils empiriquement adéquats nous permettant de faire de bonnes prédictions, et rien de plus. Je rappelle que tout modèle est faux *stricto sensu* car il ne correspond jamais parfaitement à la réalité qu'il vise décrire, représenter. D'ailleurs, si c'était le cas, un modèle ne servirait pas à grand-chose, n'aurait aucun pouvoir, ni prédictif ni explicatif, car il s'agirait d'un parfait miroir de la réalité. C'est pourquoi le caractère perspectiviste de tout modèle n'en met pas nécessairement en cause le caractère, du moins en partie, réaliste. La question du réalisme me renvoie assez naturellement à la suivante qui est celle du pluralisme scientifique.

3-Le pluralisme scientifique

Nicolas Glade, tout en critiquant le paradigme vivant-machine en vigueur, affirme que « Changer de paradigme en vigueur (la théorie cellulaire, le vivant-machine) n'est pas le but. » Il « souhaite davantage considérer un point de vue différent sur la Nature (...) » (p. 30). Ce que je comprends en lisant ce passage est que Nicolas Glade n'est donc pas un train d'essayer de renverser le paradigme vivant-machine pour le substituer par un paradigme alternatif davantage tournée vers les phénomènes rares, faibles, amorphes, phénomènes qui n'ont pas de réalité dans la biologie contemporaine à son avis, car considérés comme étant peu stables, peu ordonnés, peu structurés. D'autant plus qu'il reconnaît à plusieurs reprises les biens faits des modèles en vigueur et les résultats qu'ils ont permis de produire en termes de compréhension du vivant. Nicolas Glade veut plutôt proposer une vision additionnelle du vivant, ce qui l'amène à mon sens vers une situation de pluralisme scientifique caractérisée par une pluralité de modèles différents du vivant, aucun n'étant meilleur que les autres en tant que description vraie du vivant, tous le décrivant d'une certaine perspective, en posant certaines questions, en se focalisant sur certaines de ses propriétés et pas d'autres, etc. Si l'on suit Nicolas Glade (et surtout si je comprends bien son propos), on se retrouve donc avec une pluralité de modèles du vivant, tous empiriquement adéquats, chacun fournissant une certaine vérité au sujet de l'un de ses aspects. D'où ma question : tous ces modèles, et notamment ceux produits au sein du paradigme vivant-machine et ceux que la vision de Nicolas Glade permet de produire,

sont-ils compatibles entre eux ou pas ? Car il y a pluralisme et pluralisme dans les sciences : un pluralisme compatible, qu'il soit dépassable ou pas (notamment par l'intégration de modèles ou d'approches différentes, cf. les travaux de Sandra Mitchell), et un pluralisme incompatible et donc indépassable (cf. les travaux de Helen Longino). C'est ce dernier qui pose le plus de problèmes du point de vue philosophique, notamment en posant plusieurs défis au sujet des questions du réalisme des modèles et de l'objectivité de la connaissance produite. J'aimerais discuter davantage avec Nicolas Glade de son positionnement au sujet de cette question du pluralisme, spécifiquement en ce qui concerne les modèles du vivant qui l'intéresse. Je tiens aussi à remarquer que la proposition de Nicolas Glade d'utiliser non pas un seul réseau mais un ensemble de réseaux pour représenter un système biologique (cf. p. 219 et suivantes) témoigne de son adhésion à un certain pluralisme dans les sciences (en biologie, dans son cas).

4-Le statut des simulations numériques

Nicolas Glade passe une partie de son temps de recherche à simuler le comportement de certains systèmes vivants, ce qui éveille ma curiosité sur ce qu'est une simulation numérique, ce qu'elle permet de faire, autrement dit son statut et son rôle épistémologiques en science. Lorsque Nicolas Glade présente la simulation de proto-métabolisme (p. 138), il parle d'« expérimentation numérique ». Est-ce le statut qu'il attribue aux simulations, à savoir l'analogie des expériences à la pailleasse mais faites avec un ordinateur ? Et si c'est bien le cas, quelles sont les différences entre les expérimentations avec de la matière vivante et celles sur ordinateur ? Jouent-elles le même rôle épistémologique dans nos pratiques scientifiques ? Une différence évidente est que les simulations numériques ne sont pas ancrées dans le réel au sens matériel. Elles sont bien sûr faites à partir d'hypothèses et d'ingrédients de départ empiriquement pertinents, mais ce qu'elles nous disent ne correspond pas nécessairement à un phénomène réel au sens d'actualisé. Que sont-elles donc ? Des outils nous permettant de construire des scénarios ou des histoires possibles et empiriquement plausibles ? Et quel est leur rôle ? Celui de nous permettre d'explorer l'espace des possibles ? J'aimerais entendre davantage Nicolas Glade à ce sujet.

5-La relation entre mécanisme et holisme et entre mécanisme et hasard

Au tout début de son mémoire d'HDR et à plusieurs reprises par la suite, Nicolas Glade exprime son penchant en faveur d'un certain holisme dans l'étude du vivant et, en même temps, développe une critique raisonnée de l'approche réductionniste et mécaniste dominante dans la biologie contemporaine. Si je comprends bien sa position, le holisme s'opposerait donc au mécanisme, ce qui me semble allez de soi, à première vue, car

L'approche mécaniste consiste à décomposer un système afin d'en identifier les parties, notamment les entités qui le composent et les activités qu'elles performant, étant ainsi souvent qualifiée de réductionniste. Cependant, ce travail de décomposition est en principe suivi par une recombinaison des différentes parties et des mécanismes les animant (cf. par exemple les travaux de William Bechtel ou Carl Craver). L'approche mécaniste ne s'oppose donc pas nécessairement à un certain holisme, du moins en principe (car en pratique, je vous l'accorde, la recherche du ou de la biologiste est souvent focalisée sur une sous-partie spécifique d'un système biologique, qu'il ou elle analyse finement, sans forcément prendre en compte le contexte dans lequel ce sous-système s'insère). Ma remarque vise ici à pousser Nicolas Glade à spécifier davantage son holisme, et pourquoi il s'oppose au mécanisme. Je suggère qu'il s'agit d'un holisme de type organiciste, selon lequel les parties ont besoin du tout car ce dernier les détermine (et inversement), ce qui s'oppose nécessairement à une vision mécaniste et réductionniste à la fois en termes épistémologiques et ontologiques (cf. notamment les travaux de mon collègue Matteo Mossio).

En lisant le mémoire de Nicolas Glade, le mécanisme se révèle aussi être incompatible avec le hasard, l'aléatoire, la stochasticité (plusieurs termes souvent utilisés de manière interchangeable dans la littérature biologique, et que l'on pourrait discuter ensemble). Est-ce bien le cas ? Un mécanisme est-il nécessairement déterministe ? Il est vrai que, dans les différentes définitions que plusieurs philosophes de la biologie ont donné du mécanisme, figure l'idée qu'un mécanisme, en vertu des entités et des activités qui le composent, produit de manière régulière un certain état final (e.g., Machamer, Darden, Craver 2000). Cependant, il a aussi été proposé de parler de « mécanisme stochastique » (e.g., Bogen 2005, Andres 2012) afin d'intégrer la stochasticité dans les explications de type mécaniste. A mon avis, il serait intéressant pour Nicolas Glade de se plonger dans cette littérature et de s'y confronter afin de développer davantage l'idée d'une tension entre mécanisme et hasard.

6-La notion de finalité intrinsèque

Nicolas Glade introduit la notion de « intrinsically purposive » dans la conclusion de son mémoire d'HDR en citant les travaux de Daniel Nicholson. En quelques mots, c'est l'idée que le vivant, contrairement à une machine, est caractérisé par une finalité propre et un contrôle interne (bien que l'environnement, comme le reconnaît Nicolas Glade, exerce des pressions et donc un contrôle externe sur le vivant). C'est ici que j'aurais aimé voir cité et discuté le travail de Jacques Monod, en particulier les premières pages de son livre "Le hasard et la nécessité" où il est question de « projet téléonomique ».

Indépendamment de cela, ce passage confirme à mes yeux que Nicolas Glade adhère à un holisme organiciste (cf. le point précédent). Une question cependant se pose : la mise en évidence de cette finalité intrinsèque, qui est une forme d'agentivité propre au vivant, est-elle une autre façon de mettre en évidence les limites qui caractérise, selon Nicolas Glade, l'approche mécaniste du vivant et plus généralement les insuffisances des modèles et explications élaborés dans la biologie contemporaine à son sujet ? Et surtout, d'où vient une telle finalité ? Est-elle aussi nécessaire, selon Nicolas Glade, pour rendre compte non seulement du fonctionnement des systèmes biologiques mais aussi de leur évolution ?

Je m'arrête là même si j'ai encore plusieurs interrogations sur plein d'autres points soulevés par Nicolas Glade dans son mémoire d'HDR, par exemple le concept d'analogie et le rôle de l'analogie en science, le type de transdisciplinarité qu'il pratique, la question du rôle du hasard dans les systèmes vivants (fonctionnel ?) et dans nos modèles (explicatif ?), le caractère processuel du monde biologique et l'intérêt épistémologique (ou pas) d'adopter une ontologie des processus en biologie (cf. John Dupré et Daniel Nicholson), etc. Je serai très heureuse de les discuter avec Nicolas Glade, et de vive voix cette fois-ci, très prochainement.

Francesca Merlin
Chargée de recherche - HDR
Institut d'histoire et de philosophie des sciences et des techniques
Centre national de la recherche scientifique (CNRS)
& Université Paris 1 Panthéon-Sorbonne
<https://sites.google.com/site/francescamerlin/home>
francesca.merlin@univ-paris1.fr
Tel: +33 (0)6 38865027



11.2. Réponses aux rapporteurs et examinateurs – Analyse a posteriori

Je tiens à remercier les 3 rapporteurs, Mme Francesca Merlin, M. Patrick Amar et M. Hugues Berry pour les remarques très pertinentes qu'ils ont formulées sur ce manuscrit et durant la soutenance d'Habilitation à Diriger des Recherches. Je remercie également les examinateurs et membres invités, Franck Delaplace, Victor Norris, Fabien Crauste, Pascal Ballet, Maël Montevil et Angélique Stéphanou pour les discussions que nous avons eues avant, pendant et après la soutenance.

J'ai inclus quelques modifications dans le texte d'origine (sans en changer la nature) pour tenir compte de ces remarques, en particulier dans l'introduction et la conclusion. La plupart des réponses que j'apporte n'ont cependant pas été incluses dans le texte. Dans les deux cas, j'y réponds dans cette section dont je fais un chapitre à part entière. Cela permettra au lecteur de savoir ce qui a été ajusté.

Je vais commencer par répondre aux remarques (faites dans le rapport d'Hugues Berry, mais aussi durant la soutenance) sur la taille du document et sa concision. S'il est vrai que je n'ai pas fait un cadeau aux rapporteurs, le document étant effectivement très long, j'ai aussi et surtout initialement écrit ce document comme un document de travail, un bilan destiné à moi-même, à mes futurs étudiants, à mes collaborateurs, comme je l'indique dans l'introduction. La partie des recherches actuelles sur les réseaux Booléens s'étend effectivement sur une cinquantaine de pages, même plus si l'on considère la section 9. A mon sens, une HDR peut être vue de 2 façons différentes au moins : soit on la considère comme un bilan validant la capacité à diriger des recherches (ce que je ne me suis pas limité à faire étant donné mon objectif initial), soit on la considère comme une thèse incluant des travaux en cours et des perspectives (ce qui ressemble davantage à ce que j'ai produit). Dans ce contexte et parce que les perspectives de travail que j'ai présentées sont assez larges, il m'a semblé nécessaire d'être assez exhaustif et peu avare en détails sur la description des réseaux et de ce que nous en faisons, même si ces travaux n'ont été encore que peu publiés. De plus, pour que les propositions que j'y fait, les réflexions que je développe, soient compréhensibles, il était nécessaire d'expliquer ce sur quoi elles reposent. Il me semble que les réflexions que je développe vers la fin du manuscrit (dans les chapitres 9, 10 et le présent chapitre 11), réflexions qui selon moi donnent son intérêt à ce manuscrit, sont complexes, très complexes. Ne pas en expliquer les fondations et la façon dont moi je les appréhende, eut été à mon sens une erreur : l'ensemble du manuscrit aurait dans ce cas été incompréhensible, sauf par 3 ou 4 initiés.

Concernant les expériences de paillasse proposées pour mettre en évidence des interactions faibles entre acides aminés et triplets de nucléotides, elles sont effectivement simples. C'est une remarque qui a été faite dans le rapport d'Hugues Berry et qui a été soulignée durant la soutenance. Nous en avons débattu, en particulier avec Victor Norris et Patrick Amar. Mon point de vue en particulier est que d'abord ces expériences peuvent suffire et sont très sensibles, en particulier les expériences en ITC. Ensuite, elles ne font pas intervenir de molécules ou surfaces adjacentes qui pourraient perturber

une interaction faible. Elles sont de plus simples à mettre en œuvre, ce qui, pour s'attaquer à une série de mesures, est un avantage. Enfin, elles sont peu coûteuses et ne requièrent que peu de matériel, matériel plutôt accessible. Cela est, surtout dans le contexte de la recherche actuel, un point à ne pas ignorer je pense, en particulier dans le cas de recherches qui n'ont pas plus que cela le vent en poupe. Par ailleurs, commencer avec de telles expériences dont le principe est simple ne signifie pas que ces recherches ne puissent pas se poursuivre, si les premiers résultats s'avéraient prometteurs, avec des expériences dont le principe serait plus complexe. Maintenant, fallait-il décrire ces expériences, leur fonctionnement ? Je pense que oui. J'ai bien pensé en les décrivant que tout cela pourrait s'avérer source d'ennui pour le biologiste ou le physicien averti. Le manuscrit est cependant clairement transdisciplinaire mais cette culture (interdisciplinaire) n'est absolument pas le fait de tous. Parmi ceux qui seront amenés à lire ce document, il y aura des biologistes, des physiciens, mais aussi des personnes au profil plus théorique, venant des mathématiques appliquées ou de l'informatique. Je pense que pour ces personnes là, étudiants compris, cette partie n'est pas de trop.

Les rapporteurs ont tous retenu, à juste titre d'ailleurs, qu'une des originalités de notre travail (à mes étudiants, mes collaborateurs et moi) est « *de ne pas se focaliser sur un réseau unique pour modéliser un phénomène biologique étudié, mais une population de réseaux exhibant les propriétés structurales et dynamiques du phénomène modélisé* ». Oui, cela fait clairement partie des propositions intéressantes que nous faisons ; nous sommes d'ailleurs très peu à aller dans ce sens et je suis heureux que nous ayons très tôt proposé cette vision des choses et que celle-ci commence à être considérée. Il reste donc beaucoup à faire à ce sujet.

Bien que les rapporteurs soulignent les autres développements de mon travail, et je les en remercie, je voudrais quand même apporter deux précisions. D'abord l'essentiel de mon activité n'est à mon sens pas de la modélisation mais de l'expérimentation numérique. Comme je développerai ce point plus loin, je ne vais pas m'étendre là dessus, mais ce qu'il faut retenir c'est que l'objectif est davantage le développement de réflexions et la mise à jour de concepts sur le vivant que la représentation même des objets biologiques. Ensuite et surtout, je pense que mon travail, mon intérêt, mes objectifs, ce que nous souhaitons faire avec des réseaux va plus loin, bien plus loin que la simple représentation des systèmes biologiques « *en termes de populations ou d'ensembles de réseaux* ». Pour ma part, mon regard se tourne principalement sur ce qui trait à l'évolution, aux notions associées (complexité, évolutivité, robustesse, viabilité), aux implications de ce travail sur l'émergence du vivant et sur sa nature que je vois comme un long héritage, que je vois aussi comme reposant probablement sur de très nombreux processus que j'ai qualifiés de rares, faibles et amorphes. Il est aussi tourné vers les questions connexes qui touchent autant la biologie théorique que l'informatique théorique comme ce que je commence à développer sur les réseaux pluripotents et le contrôle (NB : partie qui a été davantage développée par rapport au document d'origine).

Je vais maintenant entrer dans les nombreuses réponses à faire sur la partie philosophique de ce

travail, réponses que je fais autant aux rapports (principalement celui de Francesca Merlin) qu'aux questions et discussions que j'ai eues avant, pendant et après la soutenance (avec en particulier Franck Delaplace, Maël Montevil, Victor Norris, Patrick Amar, Angélique Stéphanou et Pascal Ballet).

Parmi les points les plus discutés, il y en a en premier lieu la question de ce que j'ai qualifié de subjectivité en sciences, des biais auxquels sont soumises les sciences. Ensuite, il y a celle du rôle des modèles (des représentations abstraites en général) ou plus précisément de leur nature ontologique, et de l'importance que je leur confère. Enfin, il y a la définition du mécanisme et de la machine qui demande à être précisée, en particulier les relations entre mécanisme et holisme/organiscime, entre mécanisme et hasard, mais aussi la notion de finalité. Je vais me servir de ces 3 axes pour développer mes réponses à des questions plus précises qui m'ont été posées.

Commençons par la « subjectivité ». Le terme que j'ai employé pour désigner une des sources importantes de biais en sciences, la *subjectivité*, est effectivement un terme fort (opposé à l'objectivité et se référant principalement à la préférence personnelle d'un individu) et ce n'est pas forcément toujours le plus adapté (NB : on remarquera cependant que J. Monod [Monod, 1970] lui-même emploie ce terme). F. Merlin propose plutôt de se référer à la notion de *perspective*. Je remercie infiniment Francesca Merlin pour le long développement à ce sujet, développement qui m'a amené à y réfléchir davantage et à préciser ma pensée. Voici mes réponses, celle d'un chercheur et non d'un philosophe : En tant qu'individu qui pratique la science dans un domaine frontière à la fois théorique et expérimental, à la fois biologique et mathématique, à la fois mécaniste et holiste, formé à l'intersection de perspectives, de points de vue, de courants de pensées scientifiques, mais aussi philosophiques, sociologiques, puis je conserve une approche objective ? Dans quelle mesure la part subjective (préférences personnelles dont j'ai du mal à m'abstraire) que j'injecte, que chacun injecte dans les sciences, joue-t-elle un rôle, influence-t-elle leur développement ? Dans un domaine frontière, peu consensuel encore, je pense que la part subjective (personnelle) est influente et c'est l'effort déontologique que consent à faire chaque individu qui permet de limiter cette influence. J'ai cependant rencontré au cours de ma carrière plusieurs chercheurs qui, devant la singularité de comportements biologiques les laissant perplexes, ont trop laissé exprimer leur sentiment, ont fait des choix scientifiques très personnels pour les expliquer et, contre tout (et tous), les ont maintenus, à tort souvent. C'est en partie ce sentiment de méfiance, y compris de mes propres idées, que j'ai mis en avant. Je suis cependant rassuré par le fait que la recherche n'est fort heureusement pas un processus individuel, que les recherches isolées, si elles ne rejoignent pas rapidement le consensus, l'acceptation des idées par les pairs, sont vite éliminées et oubliées et en ce sens ne constituent qu'un danger d'origine individuelle limité.

Mais F. Merlin a raison bien sûr, il est aussi question derrière cela de la question de la neutralité en sciences donc de la question du point de vue, celui d'un individu situé (scientifiquement, philosophiquement, technologiquement, sociologiquement situé) et celui, plus général, d'un ensemble de scientifiques, point de vue qui prendra la forme d'une perspective. Le point de vue de

nulle part n'existe pas ou est inatteignable : la réalité même de la nature est un objet au mieux inaccessible. Réaliser ceci, à mon sens, permet de relativiser ne serait-ce que l'importance de toute réflexion, représentation, interprétation que l'on puisse avoir sur cette même nature, et même sur l'importance relative accordée aux modèles face à l'observation (je reviendrai sur ce point plus loin). Mais sans aller jusque là, en restant scientifiquement pragmatique, chercher à atteindre ce que peut être le point de vue de nulle part (celui qui est neutre, qui ne dépend d'aucun facteur extrinsèque, humain en particulier) est un des objectifs vers lequel tendre. En est-on capable quand on pratique la biologie théorique et qu'on infère des concepts à partir de nos expériences personnelles ou d'un ensemble d'expériences personnelles d'un groupe de chercheurs qui sont 1) humains, 2) formés de manière finalement assez semblable, 3) s'intéressant à ces domaines particuliers des sciences touchant à la nature de la matière vivante ? Je ne sais pas. J'en doute. Et c'est bien ce que j'ai aussi voulu signifier plusieurs fois dans ce document derrière le terme « subjectivité ».

Je pense donc en effet que pour chacun et même pour des groupes entiers de chercheurs, la connaissance, surtout dans les domaines frontière, est située dans des référentiels particuliers, des chapelles, des disciplines. Cependant ... cependant je pense que justement, embrasser la transdisciplinarité, c'est aussi épouser une forme de refus de rester situé ! Refuser de devenir trop spécialiste, c'est refuser d'adopter trop fortement un point de vue particulier, c'est accepter de considérer les autres avec tout autant d'importance. C'est aussi une lutte interne permanente pour se maintenir sur ce point d'équilibre instable. On en prend l'habitude mais même après de nombreuses années et au-delà de la simple question de l'identité dans un système de recherche (ce qui est en soit un vrai problème, pratique celui-là), je trouve cela mentalement difficile.

Maintenant, je pense aussi qu'adopter un point de vue non neutre, en tant que chercheur, est quand même important. Que nous cherchions à interpréter ce que nous voyons ou à imaginer le monde tel qu'il est réellement, nous adoptons une démarche créative : nous proposons des représentations du monde, nous appuyons dessus pour asseoir nos réflexions, et bien entendu cherchons à les tester (expérimentalement et numériquement, via des modèles). Permettre le point de vue (situé), c'est autoriser une créativité dont nous avons besoin pour faire ces propositions, pour suggérer le monde. Les connaissances produites contiennent le point de vue humain et le temps, le consensus qui s'établit peu à peu, vise à éliminer cette influence. Je pense donc que la circulation des idées et connaissances dans le monde scientifique a besoin de fluidité pour que le consensus s'établisse, mais pas trop pour que le point de vue (individuel ou de groupes) puisse s'exprimer.

Les grands courants de pensée, les cadres conceptuels dans lesquels nous travaillons, selon lesquels nous réfléchissons, faisons et projetons nos observations du monde vivant, constituent des perspectives que nous décidons de suivre. L'adoption de telle ou telle perspective est pragmatique ; elle doit l'être, sinon il s'agit d'un choix non objectif. Je ne mets pas en opposition scientifique les perspectives « vivant machine » et « vivant non-machine » ou encore « vision mécaniste » et « le vivant rare, faible et amorphe », je serais bien en peine de le faire, n'ayant aucune idée de la nature réelle de la matière vivante ; je les mets en opposition pragmatique. Autrement dit, je ne propose pas un meilleur modèle de ce que peut être la vie. Je propose « juste » un cadre alternatif selon lequel

réfléchir et développer de nouveaux outils.

Concernant la nature des modèles, leur importance et le rôle que je leur attribue, effectivement, je pense que mon positionnement doit être précisé.

Avant tout, je distingue (et je ne suis pas seul, certains de mes collègues font aussi cette distinction plutôt subtile) la *modélisation* de l'*expérimentation numérique*. La distinction ne relève pas particulièrement de la technique, bien que je pense que la modélisation s'appuie davantage sur des fondements mathématiques (équations différentielles, ...) tandis que l'expérimentation numérique repose davantage sur l'utilisation d'objets provenant du monde de l'informatique (réseaux formels, automates, P-Systems et autres systèmes de réécriture, etc). Je pense que la distinction que nous introduisons ainsi tient davantage à la finalité que nous leur associons. La finalité d'un modèle est de tester des hypothèses formulées sur la base d'observations, de vérifier que les mécanismes que l'on propose pour expliquer ce qui est observé sont bien compris, de valider des connaissances. De là, ils peuvent servir éventuellement de prédicteurs de comportements non encore observés. Par exemple, si un modèle de métabolisme prévoit l'existence d'attracteurs (comportements asymptotiques) ou de transitions particulières entre attracteurs qui n'ont pas encore été observés ou dont l'observation est incertaine, un tel modèle permettra de suggérer des expériences biologiques à faire, les conditions à réunir pour obtenir les comportement souhaité étant connues grâce au modèle, et de trancher quant au(x) mécanisme(s) sous-jacent(s) [Jacquet, 2023]. L'expérimentation numérique, elle, est moins fondée sur la donnée mais davantage sur l'inférence de mondes, de comportements, de mécanismes, inférence fondée, elle-même, sur une large expérience des processus connus, y compris dans des domaines non biologiques. L'étude des mondes inférés et simulés génère en nous des idées, nous fait imaginer des concepts que l'on cherche ensuite à tester sur la matière vivante elle-même, sur les systèmes vivant, ou sur leurs modèles, ce que défend bien D. G. Feitelson dans son article *Experimental Computer Science: The Need for a Cultural Change* [Feitelson, 2006]. C'est une approche bien connue de ceux qui œuvrent en vie artificielle. J. -P. Rennard [Rennard, 2002] va même jusqu'à parler de *binologistes*, des biologistes du monde binaire, dont l'activité consiste à étudier les mondes simulés, découvrir des formes de vie artificielle, des processus, au milieu de tout ce qui émerge dans ces environnements virtuels.

En l'état actuel, je ne crois pas que les deux (simulations fondées sur des modèles et expériences numériques) jouent « *le même rôle épistémologique dans nos pratiques scientifiques* », car on fait davantage confiance à la démarche scientifique classique qui de l'observation et de la mesure propose des hypothèses au regard des représentations établies et teste celles-ci (que ce soit par modélisation comme par expérimentation). Je suis cependant convaincu que la démarche scientifique est une construction et que cette construction ne peut reposer intégralement sur ce que l'on déduit directement de l'observation immédiate. Elle doit aussi laisser une place importante à ce que l'on peut inférer, suggérer, à partir d'un paysage bien plus large de savoirs, donc à une certaine créativité d'origine humaine.

Quant à la question qui m'est posée de me positionner sur le rôle épistémologique de la simulation numérique (en général) vis-à-vis de l'observation directe, là encore, j'estime que son rôle est sous

évalué. Je pense qu'il est illusoire de croire que parce que l'expérimentation manipule la matière plus directement, ce que l'on pourra en tirer sera beaucoup plus ancré dans le réel (et personnellement, je serais même assez facilement tenté de retirer le « beaucoup » de ma phrase). Seuls les faits ont de la valeur ... mais, qu'est ce qu'un fait ? Est ce que ce que je peux mesurer et interpréter, que je décide de mesurer et interpréter avec ce que la science me permet de mesurer et interpréter dans une perspective donnée a plus de valeur que ce que je peux construire de façon complètement artificielle (en ne se fondant que sur des objets abstraits, mathématiques, informatiques, sur mon vécu scientifique notamment, *etc*) ? Les deux alimentent en réalité efficacement mon esprit. J'ai beaucoup réfléchi à cette question (et lu dessus aussi) dont la réponse semble évidemment donner l'avantage à l'expérimentation pour comprendre le monde, pour y accéder. Je crois qu'en réalité ce n'est pas le cas. Tout ce que nos sens saisissent et que notre mémoire enregistre est forcément interprété dans le cadre de modèles (de représentations mentales) qui ont été peu à peu construits et éventuellement bien éprouvés par le consensus scientifique. Ce que nous observons n'a du sens que dans ce cadre. F. Merlin « *rappelle que tout modèle est faux stricto sensu car il ne correspond jamais à la réalité qu'il vise à décrire, représenter* » ; certes mais à mon sens ce n'est guère mieux en ce qui concerne l'observation (expérimentale) et surtout l'interprétation que l'on en fait, car l'ensemble repose sur notre perception qui elle-même repose sur notre expérience, sur les représentations mentales que nous avons des choses et que nous catégorisons comme connaissances admises. En ce sens, l'approche théorique n'est de plus pas moins entachée des biais liés à l'interprétation, à la dépendance vis-à-vis des modèles sous-jacents, que l'observation. Tout enseignement qui sortira d'un modèle, d'une expérimentation numérique se heurtera tôt ou tard au cadre conceptuel, aux représentations établies, à la critique des pairs. La réalité perçue, comprise, n'est pas un fait, elle est le résultat d'une interprétation.

Dès lors, pour moi, qu'elle soit issue de l'expérience de paillasse ou suggérée par l'expérimentation numérique ou la démarche de modélisation-simulation, toute connaissance a en réalité exactement le même statut scientifique ; elle n'acquiert sa validité, sa vraie valeur explicative, qu'après passage au crible du temps, de la publication et l'évaluation par les pairs, qu'après l'épreuve du consensus.

Finalement, la partie difficile : la définition du mécanisme et de la machine, les relations entre mécanisme et holisme/organicisme, entre mécanisme et hasard, et la question du projet, de la finalité dans le vivant. Difficile parce qu'elle demande, je crois, un avis plus personnel que les autres.

Comme je l'ai écrit dans le manuscrit, quand j'écris « machine », je désigne un concept général, mécaniste. Je ne distingue pas la machine faite de métal, de bois, de plastique ou de silicium du programme informatique, ni même des objets théoriques comme les réseaux, des systèmes de réécriture, des automates cellulaires, *etc*, qui se pensent de la même manière que ces machines faites de matière. En revanche, j'apporte une distinction particulière à la question de la « machine stochastique » pour les besoins de la réflexion. Dans ces machines, la notion de stochasticité est un ajout à ce qu'est la machine déterministe et nécessite un contre-processus assurant sa robustesse, celle-ci jouant le rôle de test *a posteriori* garantissant son fonctionnement. Dans ce cas, la

stochasticité est une fonctionnalité ajoutée à la machine (NB : je me rends bien compte qu'il y a une sorte de paradoxe à dire que la stochasticité est fonctionnelle) qui lui donne accès à un paysage de fonctionnalités plus large, en particulier une capacité exploratoire plus grande. Mais je crois qu'il faut bien prendre soin de la décrire comme cela : une base de machine + une source de stochasticité + une source de robustesse.

Selon Monod « *La notion de téléonomie⁶² implique l'idée d'une activité orientée, cohérente et constructive* » et il ajoute « *Les protéines doivent être considérées comme des agents moléculaires essentiels des performances téléonomiques des êtres vivants* » [Monod, 1970] . Je reviendrai plus loin à la discussion sur la notion de téléonomie pour rester concentré ici sur les agents du vivant et la vision mécaniste de Monod. Oui, bien sûr, les protéines sont indéniablement (avec d'autres « agents » comme les ARN-enzymes par exemple) les effecteurs qui provoquent cette activité si particulière du vivant. Monod présente néanmoins tout cela, la spécificité, le contrôle, comme mécanistes, programmatiques, cybernétiques, et responsables du fonctionnement téléonomique. Il associe explicitement le vivant à la machine : « *telle une machine* », « *le vivant est une machine* », il parle de « *puissance cybernétique* », il parle d'un système qu'on n'a pas complètement « *démonté* », de « *machine chimique* », *etc.* Et si ces agents étaient moins spécifiques que la vision de l'effecteur mécaniste qu'on leur colle ? Ils n'en restent pas moins effecteurs téléonomiques. Les IDP (*intrinsically disordered proteins* [voir sections 8.3.6.1 c et d, 9.4, 9.6.2]) ... en sont une manifestation. Il n'y a pas besoin de la vision très mécaniste, cybernétique de Monod pour valider le principe téléonomique. En réalité, on peut s'affranchir de ce point de vue et même aller plus loin, ce que je fais, en disant que les interactions faibles, que le fonctionnement amorphe, peu spécifique, dynamique, comptent et peut-être même beaucoup plus qu'on ne l'imagine dans le fonctionnement (téléonomique) de la cellule. Je dirais même que cela procure probablement une meilleure performance téléonomique au vivant, meilleure que sa version spécifique, mécaniste, programmatique. La vision mécaniste confère davantage une caractéristique de projet à ces objets (on voit où ces systèmes se dirigent, on les comprend mieux individuellement en les décortiquant, on visualise leur « objectif ») mais n'augmente pas nécessairement la performance de ce pour quoi l'évolution les a retenus.

Quand je présente le mécanisme comme incompatible avec le hasard, quand j'oppose le vivant à la machine et que je le considère selon une nature faite de rare, de faible et d'amorphe, c'est-à-dire naturellement stochastique, désorganisé, déstructuré, *etc.*, je dis là que sa nature fondamentale diffère de celle de la machine, même stochastique, dont les composants sont déterminés pour un nombre limité d'interactions. Ce que je propose comme cadre alternatif (et non de remplacement) de réflexion, entre autres, c'est d'abord de trouver un moyen de nous libérer de notre dépendance à des outils théoriques (mathématiques et informatiques) qui par leur nature même de machine constituent certainement une limite aux représentations que nous pouvons faire du vivant. C'est aussi d'essayer d'adopter une vision dans laquelle la base du vivant n'est pas une machine, mais justement un système dans lequel chaque constituant et chacune des interactions qu'ils peuvent avoir entre eux

62 En bref : la téléonomie se réfère à un « projet » sans cause finale (autrement dit, non pensé par une entité extérieure préalablement). Il s'agit bien entendu d'une illusion de projet. Cela sera plus largement développé plus loin dans le texte.

sont de façon inhérente porteurs de hasard, sont changeants, un système dans lequel le hasard et le désordre sont loi et dans lequel, l'apport de l'évolution a été celui de processus de contrôle, d'ordre, guidant un fonctionnement biochimique fondamental. Derrière la machine, il y a la notion d'ingénierie et la notion de contrôle. Pour la vie, c'est l'évolution elle-même qui contient une partie du contrôle puisque c'est elle qui, avec le temps dont elle dispose, autorise la plus ou moins grande malléabilité des éléments cellulaires, leur rareté (par l'intermédiaire de la mise en place de systèmes de contrôle génétique et métaboliques notamment, et de leur ajustement fin incluant les variations fines de la structure des éléments participant à la régulation).

Et finalement c'est poser la question de ce que nous percevons comme effectif, de ce que le système fini par intégrer comme interactions utiles (que nous percevons et jugeons utiles), produisant effectivement des chaînes de causalité qui font que la cellule vit, une réflexion proche de celle proposée par M. Conrad dans le cadre du calcul naturel lorsqu'il présente chaque élément de la vie comme processeur d'information, certains éléments étant des processeurs « *efficient* » et d'autres « *non efficient* », tous co-existant ensemble dans un système vivant. Conrad introduit donc lui aussi la distinction entre ce qui est effectivement processeur (machine) et ce qui ne l'est pas (ce qui n'est pas effecteur). C'est pratique et fonctionnel. Ce que j'essaie de considérer (mentalement au moins) ressemble davantage à un continuum où toute interaction, tout mouvement, tout mécanisme, du plus bas niveau (physico-chimique) au plus haut (protéique, biosystémique ...) fait partie d'un ensemble contigu dans lequel tous ces niveaux jouent un rôle.

Pour répondre à la question sur « mon holisme »⁶³ et à savoir si je « m'oppose au mécanisme », non, je pense pas m'opposer au mécanisme ni même à l'approche bottom-up car je les inclus dans les approches scientifiques nécessaires. Je le répète, je suis quelqu'un qui a appris à ne pas se positionner (trop fortement) sur la méthode scientifique ou sur la nature des choses. Cela ne m'empêche pas de le faire pour amorcer une réflexion dans un cadre conceptuel donné. Je reproche suffisamment un déséquilibre trop important en biologie (au moins) dans ce sens pour adopter moi-même une telle attitude. Je sais cependant qu'il est nécessaire que les chercheurs en biologie (notamment) adoptent plus fortement un certain positionnement qu'un théoricien en biologie.

Non, je vois cette usage de l'approche bottom-up et du point de vue mécaniciste, comme un moyen pratique de travailler en sciences, pour représenter les éléments et processus du vivant, un moyen qui nous parle, qui se rapproche de l'outil contrôlable. C'est utile, efficace, mais je n'en fait pas une vérité. Pas plus que son alternative. Cette proposition de point de vue holiste, non-mécaniciste, en tant que chercheur, je la choisis comme une voie de recherche et rien d'autre. En tant que personne cependant je trouve à la fois fascinant et inquiétant que le vivant puisse aussi bien et fréquemment s'expliquer par une approche mécaniste (un peu comme Max Tegmark⁶⁴ trouve fascinant le fait que les mathématiques parviennent à aussi bien expliquer la nature). En tant qu'informaticien, je trouve évidemment très pratique et formidable la possibilité d'un vivant aussi programmable qu'un ordinateur, dès lors que l'on dispose de son code. Mais comme von Neumann, et bien que mon

63 Anecdote amusante : après ma soutenance d'HDR, mes collègues de TIMC m'ont offert un superbe t-shirt sur lequel est écrit face arrière « Le Rare, Le Faible et l'Amorphe » et face avant « I'm Physicist, Holist, Organicist ! I'm p53 »

64 Max Tegmark, Notre univers mathématique – En quête de la nature ultime du réel, Dunod, 2014

esprit soit biaisé par cette habitude de pensée mécaniste, cette « facilité » qui nous est offerte de penser le vivant, j'entrevois (avec difficulté) l'alternative non-mécaniste. Comme l'objet qu'elle vise à décrire, ma pensée même à ce sujet reste protéiforme. C'est donc notamment pour cette raison qu'adopter une attitude franche à ce propos aide à réfléchir : je décide de penser selon le point de vue holiste, selon le point de vue non-mécaniste ... et je décide d'écrire un manuscrit dans cette optique ; je ne me donne ainsi pas d'autre choix que de tout penser ainsi, que de donner du sens aux choses (réseaux, métabolisme, fonctionnement cellulaire, cancer, évolution, *etc*) selon ce point de vue et ... les choses prennent sens et des interrogations naissent (Comment faire pour mettre cela en évidence dans le vivant ? Comment modéliser selon ce principe ? Comment transposer des notions existantes (évolution, robustesse, complexité, *etc*) selon ce point de vue ? *etc*). Si on prend un peu de recul par rapport à ces points de vue particulier, de façon plus générale, je pense qu'en sciences, adopter des points de vue, travailler selon des prismes particuliers, est aussi une bonne manière de faire de la recherche, d'étudier les systèmes naturels ou artificiels. La pratique de la transdisciplinarité habitue à cette démarche.

Et maintenant la question que je redoute car je n'y ai pas assez réfléchi encore, celle à laquelle il va falloir dans les années à venir que j'accorde plus d'importance : celle de la finalité, celle d'un « projet » du vivant, celle du curseur entre projet téléonomique ou téléologique.

J'ai bien lu il y a quelques années Monod mais (curieusement oui) je ne l'ai pas cité dans ce document d'abord parce que ces quelques lignes que j'ai écrites à ce propos (sur ce qui est relatif au hasard et à la nécessité, sur la finalité) ne sont arrivées que vers la fin de la rédaction et puis, j'ai un peu honte de l'avouer, je ne sais pas ce que j'ai fait de mon exemplaire de ce livre ; j'ai dû le prêter comme d'autres bouquins que je n'ai jamais revus. Avant d'écrire cette réponse, j'ai pris le temps de le racheter et de le relire. Je réalise à quel point je n'en n'avais perçu que la surface tant cela m'a donné à réfléchir. Je réalise surtout à quel point il est à mon sens impossible de le saisir pleinement sans disposer d'un long, solide et complexe bagage en sciences, en pratique de la biologie, de la chimie, de la physique, de l'étude du monde naturel. Je pense cependant qu'il n'est finalement pas plus mal de ne pas l'avoir relu avant d'avoir rédigé mon HDR. Quoi qu'il en soit, voilà ce que je peux en dire...

Monod introduit la notion de téléonomie (de τέλος [télos], la finalité et νόμος [nomos], la loi, ce qui désigne une finalité se déroulant selon une loi ... de la nature) pour s'opposer fermement à la notion de téléologie (de τέλος [télos], la finalité et de λόγος [logos], le raisonnement, ce qui indique une finalité issue d'un raisonnement). Un système téléologique dépend d'une cause finale, d'un projet se déroulant selon une intentionnalité, sous-entendu extérieure au système. Le fonctionnement téléonomique présente l'apparence d'un projet mais, ne découlant d'aucune intentionnalité (extérieure), il n'en a que l'apparence ; un système au fonctionnement téléonomique donne l'illusion d'un projet mais tout dans ses éléments, dans la complexité de son émergence, dans la complexité de sa mise en œuvre, dans la perfection de ses ajustements, des mécanismes à l'œuvre, laisse penser qu'il y a projet.

Monod cherche à montrer que tout objet biologique, tout fonctionnement biologique, si complexe

soit-il, peut être compris dans le cadre d'un fonctionnement téléonomique (et ne requiert aucunement une cause finale). Il écrit par exemple « *Toutes les structures, toutes les performances, toutes les activités qui contribuent au succès du projet essentiel seront donc dites téléonomiques* ». Il définit le « *projet essentiel* » comme suit : « *Toutes les adaptations fonctionnelles des êtres vivants comme aussi tous les artefacts façonnés par eux, accomplissent des projets particuliers qu'il est possible de considérer comme des aspects ou des fragments d'un projet primitif unique, qui est la conservation et la multiplication de l'espèce* ».

Pour commencer, je n'aime pas vraiment le terme « téléonomique », en tout cas l'usage trop rapide qu'on pourrait en faire. S'il présente l'avantage indéniable de distinguer un fonctionnement naturel non lié à une cause finale⁶⁵ différant de celui de la téléologie aux consonances théologiques, religieuses, voyant la réalisation de la vie comme dérivant d'une cause finale, d'un projet non illusoire, bien « réel », extérieur à la vie, à la Nature même, est quelque chose que l'on peut envisager il me semble. Ce terme (téléonomie, fonctionnement téléonomique) aplanit un peu trop à mon goût toute la nuance que l'on peut apporter à la notion de « projet », de « finalité », d'« anticipation », *etc*, notions dont je discuterai mon point de vue par la suite, de même que j'explicitai ce en quoi, pour moi, le terme « téléonomique » aplanit ces notions. J'acte donc la téléonomie et même le principe téléonomique universel car je ne souhaite pas placer mon discours sur le plan d'une cause finale, d'un dessein prévu de longue date pour la vie, l'univers et le reste. Il reste aussi qu'il me paraît un peu rapide d'inclure toutes nos réalisations émergeant de la combinaison de nos intellects dans un vaste projet téléonomique comme le fait Monod, sans aller plus loin dans la réflexion. De plus, l'admettre comme un fonctionnement téléonomique destiné à faire survivre notre espèce (ou une quelconque espèce) est un positionnement également trop rapide à mes yeux (ce que défend un peu facilement Monod). Argument tout aussi rapide : on le saurait si nos réalisations technologiques visaient à contribuer à notre survie !

Pour ces raisons, je vais m'éloigner de ces termes manquant à mon goût de nuance et me servir davantage de termes comme l'intentionnalité intrinsèque ou extrinsèque, affiner la notion de projet et glisser vers celle d'anticipation.

Pour développer cela et pour rester dans ce que j'ai développé dans ce document, je vais m'essayer à une réflexion : l'évolution, la génération des êtres vivants, de leurs lignées, peuvent-elles contenir une part d'intentionnalité (donc quelque chose vu comme une forme de téléologie si on ne prend pas garde comme je vais le faire de bien définir ce qu'est l'intentionnalité, le projet, *etc*) ? Cette forme là fait appel à l'auto-référence et à une intentionnalité intrinsèque (donc le *intrinsically purposive*), ce qui, si on le traduit sous la forme téléonomique vs téléologique serait une sorte de fonctionnement auto-téléologique : la cause finale, l'intentionnalité, le projet, existent mais viennent de l'intérieur du système qui met en œuvre le fonctionnement, même si ce système est capable (et a besoin) de percevoir l'extérieur pour tester son adaptation / s'adapter (je maintiens ici l'ambiguïté avec à gauche la non-intentionnalité et à droite l'intentionnalité). En comparaison, la machine n'a pas de

⁶⁵ d'ailleurs est-ce que la « cause finale » a du sens en dehors du *dessein théologique* ? Je ne crois pas. Par contre, la notion de « projet » à courte ou moyenne échéance (en tout cas pas selon des échéances cosmiques / géologiques) est possible ; j'y reviendrai...

« projet ». Elle, ses manifestations en fonction des conditions qu'elle rencontre, sont l'extension de notre « projet ». Elle est *externally purposive*.

Le système en question, c'est la vie dans son ensemble et le processus d'évolution associé inclus. La question revient finalement à demander si ce système *vie+processus évolutif* est capable d'infléchir de lui-même son évolution pour la diriger vers une trajectoire plutôt qu'une autre ... en connaissance de cause ! Est-ce qu'il est capable d'anticiper le choix de la trajectoire « la plus adaptée à sa survie » sans percevoir pour autant ce que sera son avenir à long terme ?

Je pense effectivement, comme le dit F. Merlin, que de penser au développement et à l'évolution autonome des formes de vies, à leur agentivité propre, c'est-à-dire à l'idée que les formes de vie, bien qu'elles soient liées à la pression de l'environnement (une forme de contrôle extérieur) et aux interactions qu'elles établissent avec celui-ci, disposent d'une finalité propre (intrinsèque), permet d'introduire une distinction forte entre la machine et le vivant et surtout permet de réfléchir aux limites de notre approche expérimentale et explicative de la vie principalement fondée sur des modèles mécanistes. Quand je pense l'évolution comme des ensembles de trajectoires dans un espace de réseaux, quand je réfléchis aux processus observés comme la combinaison, la concomitance de multiples processus, je vise à m'affranchir de l'apposition stricte d'un mécanisme particulier et défini sur l'observable. Je pense qu'en réalité je discute bien de cela dans les chapitres 8 et surtout 9 quand j'aborde les questions d'atteignabilité et d'évolutivité. Existe-t-il un processus maintenant les systèmes évolutifs, contrôlant leur évolutivité, voire dirigeant leur évolution selon une trajectoire auto-supervisée ? Dans tous les cas, on pourrait parler de projet intrinsèque, de « programme » à visée évolutive. Je reste bien entendu là avec le curseur du côté téléonomique (finalité sans intentionnalité extrinsèque). La matière vivante suit une trajectoire dans le paysage morphogénétique qu'elle modifie elle-même. Son action sur son devenir peut rester comprise dans le cadre téléonomique (comme le pense Monod), le Darwinisme (le test extérieur) validant ou non les solutions successives donc, vu globalement, les trajectoires dans l'espace des solutions, ce que je développe dans le triptyque évolutivité, robustesse, viabilité. Cependant, lorsqu'on se met à parler d'atteignabilité, on émet l'idée d'un but « à atteindre ». Ça, ça peut tout à fait rester un concept utile pour notre compréhension *a posteriori* (la vie a effectivement suivi une trajectoire d'un point A à un point B, donc elle a atteint B) ou *a priori* (la modélisation prévoit les trajectoires viables à partir d'un point A connu, donc l'atteignabilité potentielle d'un point B). La question est de savoir si cela ou même une partie de cette notion d'atteignabilité (qui contient donc un projet, un objectif à atteindre) est envisageable d'un point de vue intrinsèque dans le vivant.

Je pense que cette question est intimement liée à celle du libre arbitre. Un des articles les plus intéressants qui m'ait été donné de lire à ce sujet est celui qui s'intitule « *A Turing test for free will* » écrit par Seth Lloyd [Lloyd, 2013]. Celui-ci ne se contente pas de passer en revue les arguments préalables qui ont été avancés en faveur du libre arbitre, de la question du déterminisme de l'univers (« *Is the universe deterministic or probabilistic ?* »), en particulier ceux fondés sur l'indétermination quantique ; il en explique les limites et finit par apporter une définition assez fine du libre arbitre et surtout il questionne son existence de manière assez intéressante puisqu'on revient sur le problème de l'illusion, ici l'illusion du libre arbitre (« *Why we feel free* »). Le libre arbitre est lié à la capacité

libre de décider, libre au sens où il n'est pas possible de prévoir à l'avance ce qu'un décideur prendra comme décision. Il propose que pour savoir si il y a libre arbitre ou au moins illusion de libre arbitre, on peut chercher à savoir s'il est plus difficile / plus long de déterminer une décision qu'un décideur prendra que de simplement suivre le processus qui amènera à la décision elle-même : « *That is, familiar experience of the decider – that she does not know the final decision until she has thought it all through – is a necessary feature of the decision-making process. The inability of the decider to predict her decision beforehand holds whether the decision-making process is deterministic or not.* ». Autrement dit, est-ce qu'il est possible de calculer la décision à l'avance dans la majorité des cas ? En se fondant sur la théorie de la calculabilité et la décidabilité, il prouve que ce n'est pas possible. Il prouve même que certaines décisions ne sont pas décidables ... en un temps plus court que le processus de décision : « *The uncomputability of the decision making process doesn't mean that all decisions are unpredictable, but some must always be. Moreover there is no way to determine beforehand just what decisions are predictable and which are not* ». Cela montre possiblement l'existence du libre arbitre ou à défaut, son illusion. Dans ce contexte, il était aussi évident que cela mènerait Lloyd à établir un lien avec l'auto-référence et le processus de raisonnement / calcul : « *[...] the capacity for self-reference and recursive thought prevents deciders from knowing what their decision will be beforehand [...]* ». De ce libre arbitre, il en a même une vision un peu désespérante : « *It seems that we are no more than biological machines and that free will is just an illusion* ». Ce qu'il désigne par illusion, c'est l'idée que le temps et l'énergie nécessaires pour prédire la décision sont trop importants pour que l'on puisse envisager sa prédiction ... mais qu'il reste envisageable que cela puisse être prédit quand même. Cependant, même l'illusion du libre arbitre est suffisante : Si un système qui décide n'a pas accès à la décision au préalable, il n'est pas possible de savoir si ses décisions sont déterministes ou pas et conserve ainsi l'illusion du libre arbitre. On conserve alors l'illusion de projet pour tout système doté de libre arbitre. Une vision mécaniste, s'oppose au libre arbitre. Ce qui est fixé par la machine, ce qui est prévu, n'a pas de libre arbitre. Lloyd cite S. Johnson : « *All theory is against the freedom of will ; all experience for it.* ». Au delà de la discussion sur ce que nous ressentons comme vrai, je retiens effectivement que nos théories, par leur nature analytique, mécaniste, déduisent le déterminisme du monde ! ... et rejettent ainsi le libre arbitre. C'est à mon sens ce fonctionnement mécaniste du vivant, d'apparence fortement déterministe que Monod qualifie de téléonomique, ce fonctionnement mécaniste qui commence avec les lois physico-chimiques et qui se fixe de plus en plus à mesure que le vivant se perfectionne, un fonctionnement mécaniste qui ne laisse pas sa place au libre arbitre. Et pourtant, Lloyd montre comment même des machines (des ordinateurs, des i-phones) peuvent aussi afficher une forme de libre arbitre, d'imprédictabilité de leurs opérations, par leur capacité d'auto-référence et de raisonnement récursif. De telles machines possèdent la capacité d'auto-référence dans le sens où elles sont capables d'allouer de leurs propres ressources (mémoire, threads) de façon automatique pour leurs propres opérations de contrôle et de cohérence, ainsi que pour des programmes externes. Ces machines sont ainsi dotées de décision et sont capables d'une forme de libre arbitre. La suite des opérations (allocations, threads sélectionnés, priorités successives des programmes exécutés, etc) que les ordinateurs opèrent ne sont pas prévisibles mais sont au moins partiellement anticipées. On

notera, pour la suite du raisonnement que je vais mener, que leur capacité d'anticipation résulte d'un apprentissage (soit automatique dans certaines situations, soit intégré par un ingénieur).

Sous réserve que le libre arbitre existe [Lloyd, 2023 ; Conway, 2006 & 2009⁶⁶], nous, humains (mais également la vie animale), sommes la preuve que le projet avec intentionnalité peut naître du vivant : nous pouvons influencer notre propre évolution dans un but précis, eugénique, thérapeutique, plus tard peut-être comme une nécessité pour aller vivre par delà l'atmosphère terrestre, qui sait, nous pouvons décider de politiques natalistes, de politiques de santé publique, nous pouvons influencer à tort ou au besoin notre climat dont nous connaissons au moins partiellement l'impact en retour sur notre santé, sur le devenir dans notre espèce, tout cela par notre « intelligence ». Nous sommes capables de superviser notre trajectoire pour aller d'un point A à un point B ; nous pouvons moduler le paysage morphogénétique et environnemental selon et dans lequel nous évoluons, pour rendre atteignables certaines solutions. Je n'ai pas dit que cela marchait bien ni que c'était une bonne chose, je dis juste que nous en sommes capables. C'est donc possible dans le vivant. Trop facile comme explication ? Non, je ne crois pas : nous ne sommes que de la matière vivante, des cellules agencées d'une certaine manière qui ont permis au fil du temps de créer culture et sociétés et nous amener à décider au moins partiellement de notre devenir. La possibilité de projet avec intentionnalité a émergé de l'évolution finalement. Et à plus petite échelle alors ? L'organisme non conscient ? L'organe ? Le tissu ? La cellule ? Nous avons trop vite tendance à rejeter l'idée pour les systèmes simples, cellulaires, subcellulaires, et au contraire à l'accepter pour les systèmes les plus susceptibles d'être intentionnels (d'après notre expérience, notre conception du monde) : l'homme, les grands singes, les cétacés ... à cause de l'intellect. Réfléchissons un peu à ce qu'est un projet et sa version moins anthropo-centrée, l'anticipation supervisée.

Un projet, c'est la projection dans un espace et un temps éloignés, d'une « intention » d'atteindre un but. Étant donné des entrées (un ensemble de perceptions du monde), un système ayant appris à intégrer ces entrées et à prendre des décisions successives concernant ses futurs comportements, sa trajectoire, sera vu comme réalisant un projet. Pour qu'un projet soit supervisé, il faut que le système puisse anticiper ce qu'une action provoque et qu'il ait une vision des possibles. Il faut qu'il ait appris ce que provoquent ses actions.

Le terme intention utilisé précédemment est associé à la notion de projet au sens humain du terme. Considérons un système très basique adoptant une succession de comportements atypiques. Plus ces comportements seront atypiques, non prévisibles, plus ils paraîtront planifiés (la construction d'un objet complexe comme un ordinateur, l'établissement d'une stratégie de chasse tenant compte de nombreux facteurs, l'établissement de conditions nécessaires à une meilleure survie future), plus on aura tendance à parler d'intentionnalité ou de projet. C'est justement ce que Monod désigne comme téléonomique. Il est possible de considérer une version moins humanisée : un système pourrait, étant donné une perception, adopter un ensemble de décisions complexes faisant se diriger le système vers un point éloigné de son état initial. Encore une fois, on aura tendance à y voir là un projet.

Avoir un projet c'est donc avoir la notion du temps et de l'espace et apprendre à les dompter. L'apprentissage est l'illusion qui nous laisse penser que nous sommes capables de projeter nos

66 Voir aussi à ce sujet : https://en.wikipedia.org/wiki/Free_will_theorem

décisions. En réalité, comme le montre Lloyd, un décideur ne peut pas prévoir préalablement ce que seront ses décisions sauf à procéder au processus de décision lui-même. Un apprentissage passé permet à un système de procéder à un raisonnement supervisé / un ensemble de décisions qui l'amèneront à une trajectoire particulière.

L'apprentissage est un processus qui permet de lier de façon causale un ensemble d'états perçus à des états futurs. Pour qu'un système soit capable d'agencer des éléments internes, il faut qu'il puisse mettre en relation des entrées et des sorties connues à un moment préalable à l'usage de ce qui a été appris. L'apprentissage du système est soumis (ou le système le soumet volontairement) à un test externe vérifiant son efficacité (ou bien est soumis à l'approbation par d'autres systèmes semblables ayant déjà appris). Sur cette base et sous réserve qu'ils perçoivent les variations spatiales ou temporelles de leur environnement, les systèmes peuvent prendre des décisions pour infléchir leur trajectoire, des décisions qui n'étaient pas toutes prévisibles facilement. L'apprentissage est une forme de mémoire. C'est une mémoire de ce qui a été traversé par le système en termes de conditions. L'historicité est contenue dans cet apprentissage ; plus elle est développée (plus le système est capable de mémoriser ce qu'il a traversé), plus on peut s'attendre pour un système donné, à des décisions élaborées en termes de complexité de décision ou de profondeur d'anticipation (ce qui sépare la perception du changement et l'effectivité des actions engagées). Plus l'action est complexe et anticipée, plus on s'attend à ce qu'il s'agisse d'un projet intrinsèque, d'une intentionnalité intrinsèque. Parler d'intentionnalité, même sous la dénomination « *intrinsically purposive* » comme le fait Nicholson [Nicholson, 2013 & 2019] est une notion humaine. Mais je viens de montrer que même notre intentionnalité est une forme d'illusion et qu'elle résulte en tout cas d'un apprentissage et du libre arbitre (ou de son illusion), c'est à dire de son imprédictibilité. Je trouve que ce n'est pas incompatible avec les systèmes naturels, même aux petites échelles.

Donc, est ce que de façon générale les systèmes naturels sont capables d'anticipation, de choix supervisés ? Je pense que l'on peut sans hésiter dire oui.

Est-ce que la vie et le processus évolutif en sont capables ? Est ce que l'évolution peut-être anticipée et non complètement due au hasard ? Est-ce que des systèmes vivants qui perçoivent leur environnement, ses variations spatiales et temporelles, sont capables dans une certaine mesure d'appliquer des processus qu'ils ont appris au cours de leurs trajectoires passées, pour changer leur évolution, la contrôler partiellement ? Je pense que oui aussi.

Comment mettre en évidence un tel processus d'évolution auto-référente, contrôlée partiellement par le système lui-même ? La performance évolutive, la vitesse évolutive comparée à une évolution purement darwiniste, est peut-être un élément de réponse à chercher dans l'évolution observée des êtres vivants. Mais cette approche me paraît malheureusement assez difficile.

Une autre piste est celle que suggère la simulation. Et c'est justement cette voie (de facilité) que j'ai choisi. C'est aussi celle qu'ont exploré A. Crombach et P. Hogeweg dans leur article « Evolution of evolvability in gene regulatory networks » [Crombach, 2008]. Dans leurs travaux, ces auteurs font évoluer des réseaux jusqu'à ce que leur comportement soit compatible avec celui qu'exige un environnement imposé (fitness) ...ou un autre. Ils provoquent des permutations entre conditions et montrent que les réseaux Booléens (modélisant des gènes et leurs interactions) apprennent ces deux

conditions et peuvent anticiper les évolutions futures. V. Norris également lorsqu'il propose un système doué d'apprentissage pour expliquer la cohérence compétitive à laquelle sont soumises les cellules vivantes [Norris, 1998, 2012 & 2021]. La cohérence compétitive est présentée comme un mécanisme visant à assurer les transitions entre changements d'états cellulaire. Les cellules lors de ces changements doivent *a priori* fonctionner dans des conditions paradoxales dans lesquelles un ensemble de changements est requis (changements de composition intracellulaire, de concentrations, *etc*) pour faire fonctionner un état futur, tout en permettant l'état présent de continuer à fonctionner (donc avec le moins de changements possibles). On notera au passage que son modèle sous-jacent est un modèle de réseau de neurones liant les états présents (*now*) et les états futurs (*next*). Dans les deux cas, cela s'apparente aux réseaux pluripotents dont je parle chapitre 9 et à l'apprentissage d'un contrôle cohérent (proche de la cohérence compétitive de V. Norris). L'historicité des trajectoires évolutives finit par être stockée dans des réseaux de contrôle, de telle manière que l'évolution future des systèmes vivants ne relève plus du hasard pur. Je pense donc que les travaux de recherche que je dirige peuvent amener des éléments de réponse au mieux, des suggestions en tout cas, sur la façon dont le vivant évolue, s'adapte et anticipe.

Pour conclure, je dirais donc que le vivant est doué de projection, d'anticipation auto-supervisée, bien que son fonctionnement soit sujet à des prédéterminations à cause de conditions initiales récurrentes, forcées par les lois de la physique ou de la chimie, forcées par les contraintes qu'imposent les structures biologiques et organisations supra-moléculaires et au-delà.

12. BIBLIOGRAPHIE

Index des types bibliographiques.

• Périodiques et actes de conférence	314 – 332
• Livres	332 – 335
• Thèses et Masters	335 – 336
• Posters	336
• Articles de presse	336 – 337
• Communications	337
• Vidéos	337
• Références Web	337 – 338
• Logiciels	338

Périodiques et actes de conférence.

Abbott R., *If a tree casts a shadow, is it telling the time ?* LNCS 4135 (2006)

Adam J.K., Odhav B., Bhoola K.D., *Immune responses in cancer*. Pharmacol. Therapeutics 99 (2003) 113-132

Adamatzky A., Arena P., Basile A., Carmona-Galán R., De Lacy Costello B., Fortuna L., Frasca M., Rodríguez-Vázquez A., *Reaction-Diffusion Navigation Robot Control: From Chemical to VLSI Analogic Processors*, IEEE Trans. Circ. Syst. I : Reg. Pap., 51 (2004) 926-938

Adamatzky A., Costello B., Shirakawa T., *Universal Computation with Limited Resources: Belousov-zhabotinsky and Physarum Computers*. Int. J. Bifurc. Chaos (2008) DOI:10.1142/S0218127408021750

Adami C., Ofira C., Collier T.C., *Evolution of biological complexity*, Proc. Nat. Acad. Sci. USA 97 (2000) 4463-4468

Alberts J.B., Odell G.M., *In Silico Reconstitution of Listeria Propulsion Exhibits Nano-Saltation*, PLOS Biol. 2 (2004) e412

Alushin G.M., Lander G.C., Kellogg E.H., Zhang R., Baker D., Nogales E., *High-Resolution Microtubule Structures Reveal the Structural Transitions in α -Tubulin upon GTP Hydrolysis*, Cell 157 (2014) 1117-1129

Amar P., Bernot G., Norris V., *HSIM : a simulation programme to study large assemblies of proteins*. J. Biol. Phys. Chem. 4 (2004) 79-84

Anantharaman V., Koonin E.V., Aravind L., *Comparative genomics and evolution of protein involved in RNA metabolism*. Nucleic Acids Res. 30 (2002) 1427–1464

Aravind L., Anantharaman V., Koonin E.V., *Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP,*

photolyase, and PP-ATPase nucleotide-binding domains : implication for protein evolution in the RNA. Proteins 48 (2002 (a)) 1–14

Aravind L., Mazumder R., Vasudevan S. Koonin E.V., *Trends in protein evolution inferred from sequence and structure analysis.* Curr. Opin. Struct. Biol. 12 (2002 (b)) 392–399

Artymiuk P.J., Rice D.W., Poirrette A.R., Willet P., *A tale of two synthetases.* Nat. Struct. Biol. 1 (1994) 758–760

Asai R., Taguchi E., Kume Y., Saito M., Kondo S., *Zebrafish leopard gene as a component of the putative reaction-diffusion system,* Mech. Dev. 1 (1999) 87-92

Asai A., De Lacy Costello B.P., Adamatzky A., *Silicon Implementation of a Chemical Reaction-diffusion Processor for Computation of Voronoi Diagram.* Int. J. Bifurcation and Chaos 15 (2005) 3307-3320

Aubin, J.-P., *Tychastic Viability: A Mathematical Approach to Time and Uncertainty,* Acta Biotheor. 61 (2013)

Avery O.T., MacLeod C.M., McCarty M., *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types – Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III,* J. Exp. Med., 79 (1944) 137-158

Bagorda A., Mihaylov V.A., Parent C.A., *Chemotaxis: moving forward and holding on to the past.* Thromb. Haemost. 95 (2006) 12-21

Baig H., Fontanarrosa P., Kulkarni V., McLaughlin J.A., Vaidyanathan P., Bartley B., Beal J., Crowther M., Gorochowski T.E., Grünberg R., Misirli G., Scott-Brown J., Oberortner E., Wipat A., Myers C.J., *Synthetic biology open language (SBOL) version 3.0.0,* J. Integr. Bioinfo. -1 (2020) DOI: 10.1515/jib-2020-0017

Ballet P., SimCells, *An advanced software for multicellular modeling Application to tumoral and blood vessel co-development,* preprint (2018) <https://hal.archives-ouvertes.fr/hal-01853293>

Barbier B., Brack A., *Search for catalytic properties of simple polypeptides,* Orig. Life 17 (1987) 381-390

Barbier B., Brack A., *Conformation-controlled hydrolysis of polyribonucleotides by sequential basic polypeptides,* J. Am. Chem. Soc. 114 (1992) 3511-3515

Barbier B, Caille A, Brack A., *Synthesis and beta-conformation of sequential polypeptides containing arginine, histidine, and leucine.* Biopolymers. 23 (1984) 2299-2310.

Barbier P., Zejneli O., Martinho M., Lasorsa A., Belle V., Smet-Nocca C., Tsvetkov P.O., Devred F., Landrieu I., *Role of Tau as a Microtubule-Associated Protein: Structural and Functional Aspects,* Frontiers Aging Neurosc. 11 (2019) DOI=10.3389/fnagi.2019.00204

Bard Jonathan B. L., *A model for generating aspects of Zebra and other mammalian coat patterns,* J. Theor. Biol. 93 (1981) 363-385

Bartel D.P., Chen C.Z., *Micromanagers of gene expression: The potentially widespread influence of metazoan*

microRNAs, Nat. Rev. Genet. 5 (2004) 396

Beckers R., Deneubourg J-L., Goss S., *Trails and U-turns in the selection of the path by the ant Lasius niger*, J. Theor. Biol. 159 (1992) 397-415

Belousov B. P., *A periodic chemical reaction and its mechanism*, Sb. Ref. Radiats. Med. Medgiz, Moscow (1958) 145-147.

Ben-Amor H., Cadau S., Elena A., Dhouailly D., Demongeot J., *Regulatory networks analysis: Robustness in biological regulatory networks*, IEEE Proceedings of International Conference on Advanced Information Networking and Applications Workshops, 2009 (AINA'09) 924--929

Ben Amor H., Corblin F., Fanchon E., Elena A., Trilling L., Demongeot J., Glade N., *Formal Methods for Hopfield-like networks*, Acta Biotheor. 61 (2013) 21-39

Ben-Amor H., Glade N., Demongeot J., *Mnesic Evocation: An isochron-based analysis*, IEEE AINA' 10 & BLSMC' 10, Perth, Australia, IEEE Proceedings, Piscataway, 745-750 (2010b)

Ben Amor H., Glade N., Lobos C., Demongeot J., *The isochronal fibration: characterization and implication in biology*, Acta Biotheor. 58 (2010a) 121-142

Berezikov E., Plasterk R.H.A., *Camels and zebrafish, viruses and cancer: a microRNA update*. Human Mol. Gen. 14 (2005) R183–R190

Bianchi-Smiraglia A., Rana M.S., Foley C.E., Paul L.M., Lipchick B.C., Moparthy S., Moparthy K., Fink E.E., Bagati A., Hurley E., Affronti H.C., Bakin A.V., Kandel E.S., Smiraglia D.J., Feltri M.L., Sousa R., Nikiforov M.A., *Internally ratiometric fluorescent sensors for evaluation of intracellular GTP levels and distribution*. Nat. Methods 14 (2017) DOI: 10.1038/nmeth.4404

Boldt J., *Machine metaphors and ethics in synthetic biology*, Life Sci., Soc. Pol. 14:12 (2018)

Bonnot A.M., Deldem M., Beaugnon E., Fournier T., Schouler M.C., Mermoux M., *Carbon nanostructures and diamond growth by HFCVD: role of the substrate preparation and synthesis conditions*. Diamond Rel. Mater. 8 (1999) 631–635

Bornhold S., *Boolean network models of cellular regulation: prospects and limitations*, J. R. Soc. Interface 5 (2008) S85-S94

Boukella H., Campas O., Joanny J.F., Prost J., Sykes C., *Soft Listeria: Actin-Based Propulsion of Liquid Drops*, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 69 (2004) e061906

Bracha H.S., *Human brain evolution and the “Neuroevolutionary Time-depth Principle:” Implications for the Reclassification of fear-circuitry-related traits in DSM-V and for studying resilience to warzone-related posttraumatic stress disorder*. Progr. Neuro-Psychopharmacology Biol. Psych. 30 (2006) 827-853

Brack A., Barbier B., *Early peptidic enzymes*, Adv. Space Res., 9 (1989) 683-687

Brack A., Barbier B., *Chemical activity of simple basic peptides*, Orig. Life Evol. Biosph. 20 (1990) 139-144

Bray W. C., *A periodic reaction in homogenous solution and its relation to catalysis*, J. Am. Chem. Soc. 43

(1921) 1262-1267

Brito D.A., Strauss J., Magidson V., Tikhonenko I., Khodjakov A., Koonce M.P., *Pushing forces drive the comet-like motility of microtubule arrays in Dictyostelium*, Mol. Biol. Cell 16 (2005) 3334-3340

Bu H., Yuan P., Liu H., Liu D., Qin Z., Zhong X., Song H., Li Y., *Formation of macromolecules with peptide bonds via the thermal evolution of amino acids in the presence of montmorillonite: Insight into prebiotic geochemistry on the early Earth*. Chem. Geol. 510 (2019) 72-83

Budd A., Stone C., Masere J., Adamatzky A., De Lacy Costello B., Bull L., *Towards machine learning control of chemical computers*, Proceedings of the Unconventional Computing 2006, in From Utopian to Genuine Unconventional Computers, Luniver Press, 2006, 17-36

Burton A.S., Stern J.C., Elsila J.E., Glavin D.P., Dworkin J.P., *Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites*. Chem. Soc. Rev. 41 (2012) 5459-5472

Bussey K.J., Cisneros L.H., Lineweaver C.H., Davies P.C.W., *Ancestral gene regulatory networks drive cancer*. Proc. Natl. Acad. Sci. USA 114 (2017) 6160-6162

Byl J., *Self-Reproduction in small cellular automata*, Physica D 34 (1989) 295–299

Carlier M.F., Melki R., Pantaloni D., Hill T.L., Chen Y., *Synchronous oscillations in microtubule polymerisation*, Proc. Natl. Acad. Sci. USA 84 (1987) 5257-5261

Castets V., Dulos E., Boissonade J., de Kepper P., *Experimental evidence of a sustained standing Turing-type non-equilibrium chemical pattern*, Phys. Rev. Lett. (1990) 64, 2953-2956

Chambers R.A., Bremner J.D., Moghaddam B., Southwick S.M., Charney D.S., Krystal J.H., *Glutamate and post-traumatic stress disorder: toward a psychobiology of dissociation*. Semin. Clin. Neuropsychiatry 4 (1999) 274-281

Chevalier S., Noël V., Calzone L., Zinovyev A., Paulevé L., *Synthesis and simulation of ensembles of Boolean networks for cell fate decision*, Comp. Meth. Syst. Biol., 18th International Conference, CMSB 2020, Konstanz, Germany, September 23–25, 2020, Proceedings, DOI: 10.1007/978-3-030-60327-4_11

Christen U., Ivanov S., Segretain R., Trilling L., Glade N., *On computing structural and behavioural complexities of threshold Boolean automata networks - Application to biological networks*. Acta Biotheor. 68 (2019) 119-138

Cleaves HJ, *The origin of the biologically coded amino acids*. J. Theor. Biol. 263 (2010) 490-498

Cohen M. H., Robertson A., *Wave propagation in the early stages of aggregation of cellular slime molds*, J. Theor. Biol. 31 (1971) 101-118

Conrad M., *On design principles for a molecular computer*, Comm. ACM 28 (1985) 464-480

Conrad M., *Scaling of efficiency in programmable and non-programmable systems*, Biosystems 35 (1995) 161-166

- Conway J., Kochen S., *The free will theorem*, Foundations of Physics. 36 (2006) 1441
- Conway J. H. , Kochen S., *The strong free will theorem*, Notices of the AMS. 56 (2009) 226–232
- Corblin F., Tripodi S., Fanchon E., Ropers D., Trilling L., *A declarative constraint-based method for analysing discrete genetic regulatory networks*, Biosystems 98 (2009) 91-104
- Corblin F., Fanchon E., Trilling L., *Applications of a formal approach to decipher discrete genetic networks*, BMC Bioinformatics 11 (2010) 385
- Corblin F., Fanchon E., Trilling L., Chaouiya C., Thieffry D., *Automatic inference of regulatory and dynamical properties from incomplete gene interaction and expression data*. Information Processing in Cells and Tissues. LNCS 7223 (2012) 25-30
- Corblin F., Bordeaux F., Fanchon E., Hamadi Y., Trilling L., *Connections and integration with sat solvers: A survey and a case study in computational biology*, Hybrid Optimization: Optimization and Its Applications 45 (2011) 425-461
- Cornish-Bowden A., Cardenas M.L., *Self-organization at the origin of Life*, J. Theor. Biol. 252 (2007) 411-418
- Cornish-Bowden A., Cardenas M.L., Letelier J.C., Soto-Andrade J., *Beyond reductionism : metabolic circularity as a guiding vision for a real biology of systems*, Proteomics 7 (2007) 839-845
- Cortès S., Glade N., Chartier I., Tabony J., *Microtubule self-organisation by reaction-diffusion processes in miniature cell-sized containers and phospholipid vesicles*, Biophysical Chemistry 120 (2006) 168-177
- Costa, da F. L., Rodrigues F.A., Travieso G., Ribeiro Villas Boas P., *Characterization of Complex Networks: A Survey of measurements*, Adv. Phys. 56 (2007) 167-242
- Costa D., Tougeri A., Tielens F., Gervais C., Steivano L., Lambert J.-F., *Exploring the Adsorption of Microsolvated Glycine on the Surface of Amorphous Silica with Periodic DFT*. Phys. Chem. 10 (2008) 6360-6368
- Courbet A., Molina F., Amar P., *Computing with Synthetic Protocells*. Acta Biotheor. 63 (2015) 309–323
- Courbet A., Amar P., Fages F., Renard E., Molina F.. *Computer-aided biochemical programming of synthetic microreactors as diagnostic devices*. Molec. Syst. Biol. 14 (2018) DOI : 10.15252/msb.20177845
- Crick F.H.C., *On Degenerate Templates and the Adaptor Hypothesis*, 1955, URL : <https://pdfs.semanticscholar.org/0b24/3b25d9a7cf48de876d0f631dda0e2aaf0415.pdf>
- Crick F.H., Barnett L., Brenner S., Watts-Tobin R.J., *General nature of the genetic code for proteins*. Nature 192 (1961) 1227–1232
- Crombach A., Hogeweg P., *Evolution of evolvability in gene regulatory networks*, PLOS Comp. Biol. 4 (2008) e1000112

- Damsa C., Lazignac C., Pirrotta R., Andreoli A., *Troubles dissociatifs: aspects cliniques, neurobiologiques et thérapeutiques*, Rev. Med. Suisse (2006) 52.
- Davies M.S., Blackwell J., *Energy Saving through Trail Following in a Marine Snail*, Proc. Roy. Soc. B: Biol. Sci. 274 (2007) 1233-1236
- Davies P.C.W., *Does quantum mechanics play a non-trivial rôle in life ?*, Biosystems 78 (2004) 69-79
- Demarse T.B., Wagenaar D.A., Blau A.W., Potter S.M., *The neurally controlled animat : Biological brains acting with simulated bodies*, Autonomous Robots 11 (2001) 305-310
- Demongeot J., Glade N., Moreira A., Vial L., *RNA Relics and the Origin of Life*, Int. J. Mol. Sci. 10 (2009) 3420-3441
- Dewdney A.K., *In the game called Core War hostile programs engage in a battle of bits*. Scientific American. (May 1984). URL: <https://corewar.co.uk/dewdney/1984-05.htm>
- Dewdney A.K., *A Core War bestiary of viruses, worms and other threats to computer memories*. Scientific American (March 1985). URL : <https://corewar.co.uk/dewdney/1985-03.htm>
- Diederichsen U., *Pairing properties of alanyl peptide nucleic acids containing an amino acid backbone with alternating configuration*. Angew Chemie 35 (1996) 445–448
- Diederichsen U., *Alanyl PNA: Evidence for linear band structures based on guanine– cytosine base pairs*. Angew Chemie 36 (1997) 1886–1889
- Di Giulio M., *The coevolution theory of the origin of the genetic code*, Phys. Life Rev. 1 (2004) 128–137
- Dolan G.F., Müller U.F., *Trans-splicing with the group I intron ribozyme from Azoarcus*, RNA 20 (2014) 202–213
- Doolittle W.F., *Uprooting the tree of life*. Scientific American, 282 (2000) 90–95
- Douillet A., Ballet P., *A GPU Algorithm for Agent-Based Models to Simulate the Integration of Cell Membrane Signals*. Acta Biotheor. 68 (2020) 61–71
- Draper W.E., Hayden E.J., Lehman N., *Mechanisms of covalent self-assembly of the Azoarcus ribozyme from four fragment oligonucleotides*, Nucleic Acids Res. 36 (2008) 520–531.
- Dunnill P., *Triplet nucleotide-amino-acid pairing ; a stereochemical basis for the division between protein and non-protein amino-acids*. Nature 210 (1966) 1267–1268
- Dyson F., *Dynamic Universe*, Nature 435 (2005) 1033
- Egholm M., Buchardt O., Christensen L., Behrens C., Freier S.M., Driver D.A., Berg R.H., Kim S.K., Norden B., Nielson P.E., *PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules*. Nature 365 (1993) 566– 568.
- Egholm M., Buchardt O., Nielson P.E., Berg R.H., *Peptide nucleic acids (PNA). Oligonucleotide analogues with an achiral peptide backbone*. J. Am. Chem. Soc. 114 (1992) 1895–1897

- Elena A., Ben-Amor H., Glade N., Demongeot J.: Motifs in Regulatory Networks and their Structural Robustness. IEEE Transactions on Information Technology in Biomedicine, 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. BIBE 2008., Athens : Grèce (2008)
- Ellington A.D., Khrapov M., Shaw C.A., The scene of a frozen accident. RNA 6 (2000) 485–498
- Fanchon E., Corblin F., Trilling L., Hermant B., Gulino D., *Modeling the Molecular Network Controlling Adhesion Between Human Endothelial Cells: Inference and Simulation Using Constraint Logic Programming*, Comp. Meth. Syst. Biol. 3082 (2005) 104-118
- Feitelson D. G., *Experimental Computer Science: The Need for a Cultural Change*, (2006) Online paper <https://www.cs.huji.ac.il/w~feit/papers/exp05.pdf>
- Ferreira C.M.H., Pinto I.S.S., Soares E.V., Soares H.M.V.M., *(Un)suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions – a review*. RCS Advances 5 (2015) 30989
- Forterre P., *The origin of viruses and their possible roles in major evolutionary transitions*. Virus Research 117 (2006) 5–16
- Francés-Monerris A., Carmona-García J., Trabelsi, T., Saiz-Lopez A., Lyons J.R., Francisco J.S., Roca-Sanjuán D., *Photochemical and thermochemical pathways to S₂ and polysulfur formation in the atmosphere of Venus*. Nat Commun 13 (2022) 4425
- Franklin R., Gosling R. G., *Crystallographic photo of Sodium Thymonucleate, Type B. "Photo 51." May 1952*. [archive], sur Université d'État de l'Oregon, mai 1952
- Freeland S.J., *The genetic code is one in a million*. J. Mol. Evol. 47 (1998) 238-248
- Freeland S.J., Knight R.D., Landweber L.F., Hurst L.D., *Early fixation of an optimal genetic code*. Mol. Biol. Evol. 17 (2000) 511-518
- Freeland S.J., Wu T., Keulmann N., *The case for an error minimizing in the genetic code*. Orig. Life. Evol. Biosph. 33 (2007) 457-477
- Garcia-Ruiz J.M., Nakouzi E., kotopoulou E., Tamborrino L., Steinbock O., *Biomimetic mineral self-organization from silica-rich spring waters*, Sci. Adv. 3 (2017) e1602285
- Gardner M. *Mathematical games. the fantastic combinations of john conway's new solitaire game "life"*. Scientific American 223 (1970) 120–123
- Gast K, Damaschun H., Eckert K., Schulze-Forster K., Maurer H.R., Müller-Frohne M., Zirwer D., Czarnecki J., Damaschun G., *Prothymosin alpha: a biologically active protein with random coil conformation*, Biochemistry 34 (1995) 13211-13218
- Gebser M., Kaminski R., Kaufmann B., Ostrowski M., Schaub T., Schneider M. *Potassco: The Potsdam Answer Set Solving Collection*. AI Comm. 24 (2011) 107–124
- Georgiev D., Papaioanou S., Glazebrook J. F., *Neuronic system inside neurons: molecular biology and biophysics of neuronal microtubules*, Biomed. Rev. 15 (2004), 67–75

- Georgiev D.D., Papaioanou S.N., Glazebrook J.F., *Solitonic Effects of the Local Electromagnetic Field on Neuronal Microtubules*, *NeuroQuant.* 5 (2007) 276-291
- Gilis D. Masar S. Cerf N.J., Rooman M., *Optimality of the genetic code with the respect to protein stability and amino-acid frequencies*, *Gen. Biol.* 2 (2001) 1-49
- Gillespie D.T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*. *J. Comput. Phys.* 22 (1976) 403-434
- Gillespie D.T., *Exact stochastic simulation of coupled chemical reactions*. *J. Phys. Chem.* 81 (1977) 2340-2361
- Glade N., *On the nature and shape of tubulin trails : Implications on microtubule self-organization*, *Acta Biotheor.* 60 (2012) 55-82
- Glade N., Bastien O., Ballet P., *Diversity and survival of artificial lifeforms under sedimentation and random motion*. *Th. Biosci.*, 136 (2017) 153-167
- Glade N., Beaugnon E., Tabony J., *Ground-based methods reproduce space-flight experiments and show that weak vibrations trigger microtubule self-organisation*, *Biophys. Chem.* 121 (2006) 1-6
- Glade N., Ben Amor H., Bastien O., *Trail Systems as fault tolerant wires and their use in bio-processors*, In 'Modelling Complex Biological Systems in the Context of Genomics', Proceedings of the La Colle sur Loup Spring School (2009) P. Amar, F. Kepes, V. Norris (Eds) 85-119
- Glade N., Demongeot J., Tabony J., *Numerical simulations of microtubule self-organisation by reaction-diffusion*, *Acta Biotheor.* 50 (2002) 239-268
- Glade N., Elena A., Corblin F., Fanchon E., Demongeot J., Ben-Amor H., *Determination, optimization and taxonomy of regulatory networks - The example of Arabidopsis thaliana flower morphogenesis*, IEEE AINA' 11 & BLSMC' 11, Singapore, IEEE Proceedings, Piscataway (2011)
- Glade N., Tabony J., *The effect on microtubule self-organisation of reduced gravity levels produced in ground based experiments; comparison with space flight experiments*, *J. Phys. IV, France* 6 (2001) 255-260
- Glade N., Tabony J., *Brief exposure to magnetic fields determine microtubule self-organisation by reaction-diffusion processes*, *Biophys. Chem.* 115 (2005) 29-35
- Glansdorff N., Xu Y., Labedan B., *The Last Universal Common Ancestor: Emergence, constitution and genetic legacy of an elusive forerunner*. *Biol. Direct* 3 (2008) 29
- Glass L., Kauffman S., *The logical analysis of continuous, nonlinear biochemical control networks*, *J. Theor. Biol.* 39 (1973) 103-129
- Goldstone R.L., Roberts M.E., *Self-organized Trail Systems in Groups of Humans*. *Complexity* 11 (2006) 43-50
- Gorecki J., Gorecka J.H., Adamatzky A., *Information coding with frequency of oscillations in Belousov-*

Zhabotinsky encapsulated disks. Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys. 89 (2014) 042910.
doi: 10.1103/PhysRevE.89.042910.

Goss S., Aron S., Deneubour J-L., Pasteels J.M., *Self-organized shortcuts in the Argentine ant*,
Naturwissenschaften 76 (1989) 579-581

Grassé P. P., *La reconstruction du nid et les coordinations inter-individuelles chez Bellicositermes natalensis et Cubitermes sp. La théorie de la stigmergie : Essai d'interprétation du comportement des termites constructeurs*, Ins. Soc. 6 (1959) 41-80

Greaves J.S., Richards A.M.S., Bains W., Rimmer P.B., Sagawa H., Clements D.L., Seager S., Petkowski J.J., Sousa-Silva C., Ranjan S., Drabek-Maunder E., Fraser H.J., Cartwright A., Mueller-Wodarg I., Zhan Z., Friberg P., Coulson I., Lee E., Hoge J., *Phosphine gas in the cloud decks of Venus*, Nat. Astron. (2020)
<https://doi.org/10.1038/s41550-020-1174-4>

Grzybowski B.A., Bishop K.J.M., Campbell C.J., Fialkowski M., Smoukov S.K., *Micro- and nanotechnology via reaction–diffusion*, Soft Matter 1 (2005) 114-128

Hagan S., Hameroff S., Tuszynski J., *Quantum Computation in Brain Microtubules? Decoherence and Biological Feasibility*, Phys. Rev. E 65 (2002) epub061901.

Haig D. Hurst L.D., *A quantitative measure of error minimization in the genetic code*. J. Mol. Evol. 33 (1991) 412-417

Hameroff S., *Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness*, Philos. Trans. R. Soc. London Ser. A 356 (1998) 1869-1896

Hameroff S.R., Penrose R., *Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness*. In: *Toward a Science of Consciousness - The First Tucson Discussions and Debates*. (Hameroff, S.R., Kaszniak, and Scott, A.C., eds.), 507-540, MIT Press (1996)

Hameroff S.R., Watt R.C. *Information processing in microtubules*, J. Theor. Biol. 98 (1982) 549-561

Harrison L., *What is the status of Reaction-Diffusion theory thirty-four years after Turing*. J. Theor. Biol. 125 (1987) 369-384

Hattne J., Fange D., Elf J., *Stochastic reaction-diffusion with MesoRD*, Bioinformatics 21 (2005) 2923-2924

Helbing D., Keltsch J., Molnár P., *Modelling the evolution of human trail systems*, Nature 388 (1997) 47-50

Hendry L.B., Bransome E.D., Hutson M.S., Campbell L.K., *First approximation of the stereochemical*

Hernandez-Vega A., Braun M., Scharrel L., Jahnel M., Wegmann S., Hyman B.T., Alberti S., Diez S., Hyman A.A., *Local Nucleation of Microtubule Bundles through Tubulin Concentration into a Condensed Tau Phase*, Cell Reports 20 (2017) 2304–2312

Hershey A.D., Chase M., *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. J. Gen. Physiol., 36 (1952) 39-56

Hickman R., Van Verk M.C., Van Dijken A.J.H., Pereira Mendes M., Vroegop-Vos I.A., Caarls L.,

Steenbergen, Ivo Van der Nagel, Gert Jan Wesselink, Aleksey Jironkin, Adam Talbot, Johanna Rhodes, Michel De Vries, Robert C. Schuurink, Katherine Denby, Corné M.J. Pieterse, Saskia C.M. Van Wees, *Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network*,

Higgs P.G., Pudritz R.E., *A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code*. *Astrobiology* 9 (2009) 483–490

Hirokawa N., *Kinesin and Dynein Superfamily Proteins and the Mechanism of Organelle Transport*. *Science* 279 (1998) 519-526

Ho M.-W., Ulanowicz R., *Sustainable systems as organisms?* *BioSystems* 82 (2005) 39–51

Hoagland M.B., Stephenson M.L., Scott J.F., Hecht H.I., Zamecnik P.C., *A soluble ribonucleic acid intermediate in protein synthesis*. *J. Biol. Chem.* 231 (1958) 241-257

Hoover R.B., *Fossils of Cyanobacteria in CII Carbonaceous Meteorites: Implications to Life on Comets, Europa, and Enceladus*, *J. Cosmology*. 15 : 6249 (2011)

Hopfield J.J., *Origin of the genetic code : a testable hypothesis based on tRNA structure, sequence and kinetic proofreading*. *Proc. Natl. acad. Sci. USA* 75 (1978) 4334–4338

Hopfield J.J., *Neural networks and physical systems with emergent collective computational abilities*, *Proc. Natl. Acad. Sci. USA*. 79 (1982) 2554--2558

Hsu D.C., Ford E.B., Ragozzine D., Ashby K., *Occurrence Rates of Planets Orbiting FGK Stars: Combining Kepler DR25, Gaia DR2, and Bayesian Inference*. *the Astronomical J.* 158 (2019) 109

Hunding A., Képès F., Lancet D., Minsky A., Norris V., Raine D., Sriram K., Root-Bernstein R., *Compositional complementarity and prebiotic ecology in the origin of life*, *Bioessays*. 28 (2006) 399-412

Imai E.-I, Honda H., Hatori K., Brack A., Matsuno K., *Elongation of oligopeptides in a simulated submarine hydrothermal system*. *Science* 283 (1999) 831-833

Israel L., *Tumour progression : random mutations or an integrated survival response to cellular stress conserved from unicellular organisms*. *J. Theor. Biol.* 178 (1996) 375-380

Jeancolas C., *Modéliser la transition de l'inerte au vivant. Ethnographie d'un laboratoire de recherche microfluidique*. *Techniques & Culture* 73 (2020) URL : <https://journals.openedition.org/tc/13902>

Jacquet P., Stéphanou A., *Metabolic Reprogramming, Questioning, and Implications for Cancer*. *Biology (Basel)*. 10(2):129 (2021) doi: 10.3390/biology10020129

Jacquet P., Stéphanou A., *A reduced model of cell metabolism to revisit the glycolysis-OXPHOS relationship in the deregulated tumor microenvironment*, *J. Theor. Biol.* 562 (2023) 111434

Johnson K., Rossi L.F., *A mathematical and experimental study of ant foraging trail dynamics*. *J. Theor. Biol.* 241 (2006) 360-369

Kauffman S., *Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets*, *J. Theor. Biol.* 22 (1969) 437--467

- Kauffman S., *A proposal for using the ensemble approach to understand genetic regulatory networks*, J. Theor. Biol. 230 (2004) 581-590
- Khuong A., Gautrai J., Perna A., Sbaï C., Combe M., Kuntz P., Jost C., Theraulaz G., *Stigmergic construction and topochemical information shape and nest architecture*, Proc. Natl. Acad. Sci. USA, 113 (2016) 1303-1308
- Knight R.D., Landweber L.F., *Rhyme or reason: RNA-arginine interactions and the genetic code*. Chem. & Biol. 5 (1998) R215
- Knight R.D., Landweber L.F., *The early evolution of the genetic code*. Cell, 101 (2000) 569
- Kolmogorov A.N., *Three approaches to the quantitative definition of information*, Int. J. Comp. Math., 2 (1968) 157–168
- Kolmogorov A. N., Petrovsky I. G., Piskunov N. S., *Une investigation de l'équation de la diffusion combinée avec un accroissement de masse et son application à un problème biologique* (Traduit du Russe) 1937, dans 'Selected works of A. N. Kolmogorov' V. M. Tikhomirov (Ed.), Kluwer Academic Publishers, 1991
- Kondo S., Asai R. *A Reaction-Diffusion wave on the skin of the marine angelfish Pomacanthus*, Nature 376 (1995) 765-768
- Koonin E.V., *The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life*, Biology Direct, (2007) 2:15
- Koonin E.V., *The logic of chance : The nature and the origin of biological evolution*. Upper Saddle River NJ : FT Press (2011)
- Koonin E.V., *Frozen accident pushing 50 : stereochemistry, expansion, and chance in the evolution of the genetic code*. Life 7 (2017 (b)) E22
- Koonin E.V., Dolja V.V., *Evolution of complexity in the viral world: The dawn of a new vision*. Virus Res. 117 (2006) 1–4
- Koonin E.V., Novozhilov A.S., *Origin and evolution of the genetic code: the universal enigma*. IUBMB Life. 61 (2009) 99–111
- Koonin E.V., Novozhilov A.S., *Origin and Evolution of the Universal Genetic Code*. Ann. Rev. Gen. 51 (2017 (a)) 45–62
- Krawitz P., Shmulevich I., *Basin entropy in Boolean network ensembles*. Phys. Rev. Lett. 98 (2007) 158701
- Kulkarni P., *Intrinsically Disordered Proteins : Insights from Poincaré, Waddington and Lamark*. Biomolecules 10 (2020) 1490
- Langton C.G., *Self-reproduction in cellular automata*, Physica D 10 (1984) 135-144
- Leboucher G., *Quand l'éthologie s'intéresse au lien social. Processus biologiques et plaisir*, Études rurales

189 (2012) 47-56, URL : <https://journals.openedition.org/etudesrurales/9559>

Leder Z.P., Nirenberg M.W., *RNA Codewords and Protein Synthesis, III. On the Nucleotide Sequence of a Cysteine and a Leucine RNA Codeword*, Proc. Natl. Acad. Sci. USA 52 (1964) 1521–1529

Lespinats S., De Clerck O., Colange B., Gorelova V., Grando D., Maréchal E., Van Der Straeten D., Rébeillé F., Bastien O. *Phylogeny and Sequence Space: A Combined Approach to Analyze the Evolutionary Trajectories of Homologous Proteins. The Case Study of Aminodeoxychorismate Synthase*, Acta Biotheor.68 (2020) 139–156

Limaye S.S., Mogul R., Smith D.J., Ansari A.H., Slowik G.P., Vaishampayan P., *Venus' Spectral Signatures and the Potential for Life in the Clouds*, Astrobiology 18 (2018) <https://doi.org/10.1089/ast.2017.1783>

Lineweaver C.H., Davies P.C.W., Vincent M.D., *Targeting cancer's weaknesses (not its strengths): Therapeutic strategies suggested by the atavistic model*. BioEssays 36 (2014) 827-835

Lopez A., Fiore M., *Investigating prebiotic protocells for a comprehensive understanding of the origins of life : A prebiotic systems chemistry perspective*, Life 9 (2019) doi:10.3390/life9020049

Lynn C., Kahn A.E., Bassett D.S., *Structure from noise : Mental errors yield abstract representations of events*, 2018 Conference on Cognitive Computational Neuroscience

Maini P., *How the mouse got its stripes ?* Proc. Natl. Acad. Sci. USA 100 (2003) 9656-9657

Mandelkow E., Mandelkow E-M., Hotani H., Hess B., Müller S., *Spatial patterns from oscillating microtubules*, Science 246 (1989) 1201-1203

Margolin G., Gregoretti I.V., Cickovski T.M., Li C., Shi W., Alber M.S., Goodson H.V., *The mechanisms of microtubule catastrophe and rescue: implications from analysis of a dimer-scale computational model*, Mol. Biol. Cell 23 (2012) 642-656

Matthaei H.J., Jones O.W., Martin R.G., Nirenberg M.W., *Characteristics and Composition of RNA Coding Units*, Proc. Natl. Acad. Sci. USA. 48 (1962a) 666–677

Matthaei H., Nirenberg M.W., *Characteristics and Stabilization of DNAase-Sensitive Protein Synthesis in E. coli Extracts*, Proc. Natl. Acad. Sci. USA. 47 (1962b) 1580–1588.

McCulloch W.S., Pitts W., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, Bull. Math. Biophys. 5 (1943) 115-133

McKay D.S., Thomas-Keprta K.L., Clemett S.J., Gibson E.K. Jr., Spencer L., Wentworth S.J., *Life on Mars: new evidence from martian meteorites*, Proc. SPIE, 774102 (2009)

McMahon S., Bosak T., Gotzinger J.P., Milliken R.E., Summons R.E., Daye M., Newman S.A., Fraeman A., Williford K.H., Briggs D.E.G., *A field guide to finding fossils on Mars*, J. Geophys. Res. Planets 123 (2018) 1012-1040

Melo A.M., Coraor J., Alpha-Cobb G., Elbaum-Garfinkle S., Nath A., Rhoades E., *A functional role for intrinsic disorder in the tau-tubulin complex*, Proc. Natl. Acad. Sci. USA. 113 (2016) 14336-14341

- Mendoza L., Alvarez-Buylla E.R., *Dynamics of the Genetic Regulatory Network: Arabidopsis Thaliana Flower Morphogenesis*, J. Theor. Biol. 193 (1998) 307--319
- Meshkini A., Yazdanparast R., Nouri K., *Intracellular GTP level determines cell's fate toward differentiation and apoptosis*, Toxicol. Appl. Pharmacol. 253 (2011) 188-96
- Michelson A.A., Morley E.W., *On the Relative Motion of the Earth and the Luminiferous Ether*, Philos. Mag. S.5, 24 (1887) 449-463
- Miller S.L., *A Production of Amino Acids Under Possible Primitive Earth Conditions*. Science 117 (1953) 528-529
- Miller S.L., Urey H.C., *Organic Compound Synthesis on the Primitive Earth*. Science 130 (1959) 245-251
- Mitchison T., Kirschner M., *Dynamic instability of microtubule growth*, Nature. 312 (1984) 237-242
- Mitić S., Strampraad M.J.F., Hagen W.R., de Viries S., *Microsecond time-scale kinetics of transient biochemical reactions*, Plos One (2017) <https://doi.org/10.1371/journal.pone.0185888>
- Morowitz H., Smith E., *Energy flow and the organization of Life*. Complexity 13 (2007) 51-59
- Muchowska K.B., Varma S.J., Chevallot-Beroux E., Lethuiller-Karl L., Li G., Moran J., *Metals promote sequences of the reverse Krebs cycle*. Nat. Ecol. Evol. 1 (2017) 1716-1721
- Muchowska K.B., Varma S.J., Moran J., *Synthesis and breakdown of universal metabolic precursors promoted by iron*. Nature 569 (2019) 104-107
- Muller A.W.J., *Cancer is an adaptation that selects in animals against dissipation*, Med. Hyp. 104 (2017) 104-115
- Muz B., de la Puente P., Azab F., Azab A.K., *The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy*, Hypoxia (Auckl) 3 (2015) 83-92
- Newsome T.P., Scaplehorn N., Way M., *SRC mediates a switch from microtubule- to actin-based motility of vaccinia virus*, Science 306 (2004) 124-129
- Nicholson D.J., *Organisms ≠ Machines*, Studies Hist. Philo. Biol. Biomed. Sci. 44 (2013) 669-678
- Nicholson D.J., *Is the cell really a machine ?*, J. Theor. Biol. 477 (2019) 108-126
- Nirenberg M.W., Matthaei H.J. *The Dependence Of Cell- Free Protein Synthesis In E. coli Upon Naturally Occurring Or Synthetic Polyribonucleotides*, Proc. Natl. Acad. Sci. USA. 47 (1961) 1588-1602
- Nishinaria K., Sugawarab K., Kazamac T., Schadschneiderd A., Chowdhury D., *Modelling of self-driven particles: Foraging ants and pedestrians*. Physica A 372 (2006) 132-141
- Norris V., *Modelling Escherichia coli. The concept of competitive coherence*, Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie, 321 (1998) 777-787

- Norris V., Hunding A., Képès F., Lancet D., Minsky A., Raine D., Root-Bernstein R., Sriram K., Question 7: the first units of life were not simple cells, *Orig. Life Evol. Biosph.* 37 (2007) 429-443
- Norris V., Engel M., Demarty M., *Modelling Biological Systems with Competitive Coherence*, *Advances in Artificial Neural Systems.* (2012) 10.1155/2012/703878.
- Norris V., *Competitive Coherence Generates Qualia in Bacteria and Other Living Systems*, *Biology.* 10 (2021) 1034
- Novozhilov A.S., Koonin E.V., *Exceptional error minimization in putative primordial genetic codes.* *Biol. Direct* 4:44 (2009)
- Novozhilov A.S., Wolf Y.I., Koonin E.V., Evolution of the genetic code : partial optimization of a random code for robustness to translation error in a rugged fitness landscape, *Biol. direct.* 2:24 (2007)
- Norris V., den Blaauwen T., Cabin-Flaman A., Doi R.H., Harshey R., Janniere L., Jimenez-Sanchez A., Jin D.J., Levin P.A., Mileykovskaya E., Minsky A., Saier M. Jr., Skarstad K., *Functional Taxonomy of Bacterial Hyperstructures.* *Microbiol, Mol. Biol. Rev.* 71 (2007) 230-253
- Norris V., Demarty M., Raine D., Cabin-Flaman A., Le Sceller L., *Hypothesis: hyperstructures regulate initiation in Escherichia coli and other bacteria.* *Biochimie* 84 (2002) 341–347
- Nushi E., Popescu V.-B., Sanchez Martin J.A., Ivanov S., Czeizler E., Petre I., *Network modeling methods for precision medicine*, arXiv (2021) arXiv:2104.09206.
- Odde D., *Diffusion inside microtubules*, *Eur. Biophys. J.* 27 (1998) 514–520
- Pagel M., O'Donovan C., Meade A., *General statistical model shows that macroevolutionary patterns and processes are consistent with Darwinian gradualism.* *Nature Comm.* 13:1113 (2022) doi: 10.1038/s41467-022-28595-z.
- Pelc S.R., *Correlation between codon-triplet and amino acids.* *Nature* 207 -1965) 597–599
- Pelc S.R., Welton M.G.E., *Stereochemical relationship between coding triplets and amino-amids.* *Nature* 209 (1966) 868–870
- Perello M., Barbier B., Brack A., *Hydrolysis of oligoribonucleorides by α -helical basic peptides*, *Int. J. Peptide Protein Res.* 38 (1991) 154-160
- Pervushin K., Vamvaca K., Vögeli B., Hilvert D., *Structure and dynamics of a molten globular enzyme.* *Nat. Struct. Mol. Biol.* 14 (2007) 1202–1206
- Pesic P.D., *The smallest clock*, *Eur. J. Phys.* 14 (1993) 90-92
- Pirollet F., Job D., Margolis R.L., Garel J.-R., *An oscillatory mode for microtubule assembly*, *The EMBO J.* 6 (1987) 3247-3252
- Pfaffmann J.O., Conrad M., *Adaptative information processing in microtubule networks*, *BioSystems* 55 (2000) 47-58

- Pizzarello S., *The chemistry of Life's origin : a carbonaceous meteorite perspective*. Acc. Chem. Res. 39 (2006) 231-237
- Qian L, Winfree E., *Parallel and scalable computation and spatial dynamics with DNA-based chemical reaction networks on a surface*. LNCS 8727 (2014) 114-131
- Rashevsky N., *An approach to the mathematical biophysics of biological self-regulation and of cell polarity*, Bull. Math. Biophys. 2 (1940) 15-25
- Rasmussen S., Feldberg R., Hinsholm M, Knudsen C., *Core evolution: Emergence of cooperative structures in a computational chemistry*, Proceedings of the Los Alamos sponsored annual 9th conference *Emergent computation* (1989), URL : https://digital.library.unt.edu/ark:/67531/metadc1103422/m2/1/high_res_d/5858089.pdf
- Rasmussen S., Knudsen C., Feldberg R., Hinsholm M., *The coreworld: Emergence and evolution of cooperative structures in a computational chemistry*, Physica D: Nonlinear Phenomena 42 (1990) 111-134
- Ray S.T., *Evolution, ecology, and optimization of digital organisms*, SFI Working Paper 1992-08-042 (1992)
- Ray T., *Evolution, complexity, entropy and artificial reality*, Physica D: Nonlinear Phenomena, 75 (1994) 239-263
- Reggia J.A., Armentrout S.L., Chou H.-H., Peng Y., *Simple systems that exhibit self-directed replication*. Science 259 (1993) 1282–1287
- Robertson M.P., Joyce G.F., *The Origins of the RNA World*. Cold Spring Harb. Perspect. Biol. 4 (2012) a003608
- Rodionov V.I., Borisov G.G., *Microtubule Treadmilling in Vivo*, Science. 275 (1997) 215–218
- Roll-Mecak A., *Intrinsically disordered tubulin tails: complex tuners of microtubule functions?* Sem. Cell Dev. Biol. (2015) 17,11-19
- Ronald E.M.A., Sipper M., Capcarrère M.S., *Testing for emergence in Artificial Life*, ECAL '99: Proceedings of the 5th European Conference on Advances in Artificial Life September (1999) 13-20
- Rothmund P.K., Papadakis N., Winfree E.. *Algorithmic Self-Assembly of DNA Sierpinski Triangles*. PLoS Biol. 2 (2004) e424
- Rothmund P.W.K., *Folding DNA to create nanoscale shapes and patterns*. Nature 440 (2006) 297–302
- Rovinsky A. B., Menzinger M., *Self-organization induced by the differential flow of activator and inhibitor*, Phys. Rev. Lett. (1993) 70, 778-781
- Sakhno Y., Battistella A., Mezzetti A., Jaber M., Georgelin T., Michot L., Lambert J.-F., *One Step up the Ladder of Prebiotic Complexity: Formation of Nonrandom Linear Polypeptides from Binary Systems of Amino Acids on Silica*. Chem. Eur. J., 25 (2019) 1275-1285
- Santarella R.A., Skiniotis G., Goldie K.N., Tittmann P., Gross H., Mandelkow E.-M., Mandelkow E., Hoenger A., *Surface-decoration of microtubules by human tau*, J. Mol. Biol. 339 (2004) 539-553

- Sayer R.M.P., *Self-organizing proto-replicators and the origin of life*, Biosystems, 90 (2007) 121–138
- Saxinger C., Ponnamperuma C. Woese C. *Evidence for the interaction of nucleotides with immobilized amino-acids and its significance for the origin of the genetic code*. Nat. New Biol. 234 (1971) 172–174
- Sayama H., *Toward the Realization of an Evolving Ecosystem on Cellular Automata*, Proc. 4th Int. Symp. Art. Life Rob. (AROB 4th '99), M. Sugisaka and H. Tanaka (eds.), Beppu, Oita, Japan (1999) 254–257
- Schulte-Pelkum N., Wieskotten S., Hanke W., Dehnhardt G., Mauck B., *Tracking of Biogenic Hydrodynamic Trails in Harbour Seals (Phoca vitulina)*, The J. Exp. Biol. 210 (2007) 781-787
- Segretain R., Ivanov S., Trilling L., Glade N., *Implementation of a Computing Pipeline for Evaluating the Extensibility of Boolean Networks' Structure and Function*. bioRxiv (2020a) DOI: 10.1101/2020.10.02.323949
- Segretain R., Ivanov S., Trilling L., Glade N., *A Methodology for Evaluating the Extensibility of Boolean Networks' Structure and Function*. Proc. Complex Networks 2020 conference (2020b)
- Sel R., *Dissociation as complex adaptation*. Med. Hypotheses 48 (1997) 205-208
- Sept D., Limbach H. J., Bolterauer H., Tuszynski J.A., *A chemical kinetics model for microtubule oscillations*, J. Theor. Biol. 197 (1999a) 77-88
- Sept D., *Model for spatial microtubule oscillations*, Phys. Rev. E. 60 (1999b) 838-841
- Sims K., *Evolving 3D morphology and behavior by competition*, Artificial Life IV, MIT Press, Brooks R. & Maes P. (Eds), 28-39 (1994 a)
- Sims K. *Evolving virtual creatures*, SIGGRAPH'94, ACM Press, 15-22 (1994 b)
- Singh K., Muljadi B.P., Raeini A.Q., Jost C., Vandeginste V., Blunt M.J., Theraulaz G., Degond P., *The architectural design of smart ventilation and drainage systems in termite nests*, Sci. Adv. 5 (2019) eaat8520
- Suzuki N., Hirata M., Kondo S., *Traveling stripes on the skin of a mutant mouse*, Proc. Natl. Acad. Sci. USA 100 (2003) 9680-9685
- Schweiger R., Seibert H., *Graph representations of monotonic Boolean model pools*, Comp. Models Syst. Biol. - Springer Int. Pub. (2017) 233-248
- Soler-Toscano F., Zenil H., Delahaye J.-P., Gauvrit N., *Calculating Kolmogorov Complexity from the Output Frequency Distributions of Small Turing Machines*, PLoS ONE, 9 (2014) e96223
- Tabony J., Glade N., Papaseit C., Demongeot J., *The effect of gravity on microtubule self-organisation*, J. Phys. IV, France 11 (2001) 239-246
- Tabony J., Glade N., Papaseit C., Demongeot J., *Microtubule self-organisation and its gravity dependence*, Chapter in: 'Advances in Space Biology and Medicine: Cell biology and biotechnology in space' (2002a), Elsevier Science, A. Cogoli (Ed.) 8, pp 19-58. ISBN: 0444507353

- Tabony J., Glade N., Demongeot J., Papaseit C., *Biological self-organization by way of microtubule reaction-diffusion processes*, Langmuir 18 (2002b) 7196-7207
- Thellier M., Legent G., Amar P., Norris V., Ripoll C., *Steady-state kinetic behaviour of functioning-dependent structures*. FEBS Journal 273 (2006) 4287–4299
- Thellier M., Legent G., Norris V., Baron C., Ripoll C., *Introduction to the concept of functioning-dependent structures in living cells*, C. R. Biologies 327 (2004) 1017–1024
- Theraulaz G., Bonabeau E., Nicolis S.C., Solé R.V., Fourcassié V., Blanco S., Fournier R., Joly J.-L., Fernandez P., Grimal A., Dalle P., Deneubourg J.-L., *Spatial Patterns in Ant Colonies*, Proc. Natl. Acad. Sci. USA 99 (2002) 9645-9649
- Theriot J.A., Rosenblatt J., Portnoy D.A., Goldschmidt-Clermont P.J., Mitchison T.J., *Involvement of profilin in the actin-based motility of L. monocytogenes in cells and in cell-free extracts*. Cell 76 (1994) 505-517
- Thomas R., *On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations*, Springer Ser. Synerg. 9 (1980) 180-193
- Tran V., McCall M.N., McMurray H.R., Almudevar A., *On the underlying assumptions of threshold Boolean networks as a model for genetic regulatory network behavior*, Frontiers Gen. 4 (2013) 263
- Traut W.T., *Physiological concentrations of purines and pyrimidines*. Mol. Cell. Biochem. 140 (1994) 1–22
- Turing A., *On Computable Numbers, with an Application to the Entscheidungsproblem* : Proceedings of the London Mathematical Society, London Mathematical Society, 1937
- Turing A. M., *The chemical basis of morphogenesis*, Phil. Trans. R. Soc. Lond. (1952) B237, 37-72A
- Tuszyński J.A., Hameroff S., Sataric M.V., Trpisova B., Nip L.A., *Ferroelectric behavior in microtubule dipole lattices: implications for information processing, signaling and assembly/disassembly*, J. Theor. Biol. 174 (1995) 371-380
- Tuszyński J.A., Portet S., Dixon J.M., Luxford C., Cantiello H.F., *Ionic Wave Propagation along Actin Filaments*, Biophys. J., 86 (2004) 1890-1903
- Tyson J., Alexander K., Manoranjan V., Murray J., *Spiral waves of cyclic AMP in a model of slime mold aggregation*, Physica D 34 (1989) 193-207
- Uversky V.N., Oldfield C.J., Midic U., Xie H., Xue B., Vucetic S., Iakoucheva L.M., Obradovic Z., Dunker A.K., *Unfoldomics of human diseases: linking protein intrinsic disorder with diseases*, BMC Genomics 10, S7 (2009). <https://doi.org/10.1186/1471-2164-10-S1-S7>
- Vanag V.K., Epstein I.R., *Segmented spiral waves in a reaction-diffusion system*, Proc. Natl. Acad. Sci. USA 100 (2003) 14635–14638
- VanBuren V., Cassimeris L., Odde, D.J., *Mechanochemical Model of Microtubule Structure and Self-Assembly Kinetics*, Biophys. J., 89 (2005) 2911-2926

- Varela F.G., Maturana F.H., Uribe R., *Autopoiesis: The organization of living systems, its characterization and a model*, Biosystems, 5 (1974) 187
- Varetto J., *Typogenetics: an artificial genetic system*, J. Theor. Biol., 193 (1993) 257
- Varetto L., *Studying artificial Life with a molecular automaton*, J. Theor. Biol. (1998) 193
- Vassilev P.M., Dronzine R.T., Vassileva M.P., Georgiev G.A., *Parallel arrays of microtubules formed in electric and magnetic fields*, Biosci. Rep. 2 (1982) 1025-1029
- Vaya I., Gustavsson T., Miannay F.-A., Douki T., Makovitsi D., *Fluorescence of Natural DNA: From the Femtosecond to the Nanosecond Time Scales*. J. Am. Chem. Soc. 132 (2010) 11834-11835
- Vetsigian K, Woese C, Goldenfield N, *Collective evolution of the genetic code*. Proc. Natl. Acad. Sci. 103 (2006) 10696-10701
- Vincent M., *Cancer: A de-repression of a default survival program common to all cells ?* BioEssays 34 (2011) 72-82
- Wagner A., *Robustness and evolvability : a paradox resolved*. Proc. Roy. Soc. B 275 (2008) 91-100
- Watson J.D., Crick F.H.C., *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid*. Nature 171 (1953) 737-738
- Webster C.R., Mahaffy P.R., Atreya S.K., Flesch G.J., Mischna M.A., Meslin P.-Y., Farley K.A., Conrad P.G., Christensen L.E., *Mars methane detection and variability at Gale crater*, Science 347 (2015) 415–417
- Welch M.D., Way M., *Arp2/3-Mediated Actin-Based Motility: A Tail of Pathogen Abuse*, Cell Host & Microbe, 14 (2013) 242-255
- Weinreb P.H., Zhen W., Poon A.W., Conway K.A., Lansbury P.T., *NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded*, Biochemistry 35 (1996) 13709–13715
- Whitfield J., *Nascence man*, Nature 459 (2009) 316–319
- Wittung P., Nielsen P.E., Buchardt O., Egholm M., Norden B., *DNA-like double helix formed by peptide nucleic acid*. Nature 368 (1994) 561–563
- Witzany G., *Natural genome-editing competences of viruses*. Acta Biotheor. .54 (2006) 235-53.
- Woese C.R., *A new biology for a new century*, Microbiol. Molec. Biol. Rev. 68 (2004) 173-186
- Woese C.R., Dugre D.H., Saxinger W.C., Dugre S.A., *The molecular basis for the genetic code*, Proc. Natl. Acad. Sci. USA 55 (1966) 966–974
- Wong J.T.F., *Coevolution theory of the genetic code at age thirty*, Bioessays 27 (2005) 416–425
- Wong J.T.F., *Question 6 : Coevolution theory of the genetic code : a proven theory*. Orig. Life Evol. Biosph. 37 (2007) 403–408

- Yarus M., *Amino acids as RNA ligands : a direct-RNA-template theory for the code's origin*. J. Mol. Evol. 47 (1998) 109-117
- Yarus M., *RNA-ligand chemistry : a testable source for the genetic code*. RNA 6 (2000) 475-484
- Yarus M., Caporaso J.G., Knight R., *Origin of the genetic code : the escaped triplet theory*. Annu. Rev. biochem. 74 (2005) 179-198
- Yarus M., Wildmann JJ., Knight R., *RNA-amino acid binding : a stereochemical era for the genetic code*. J. Mol. Evol. 69 (2009) 406-429
- Yeh I.-C., Hummer G., *Peptide Loop-Closure Kinetics from Microsecond Molecular Dynamics Simulations in Explicit Solvent*, J. Am. Chem. Soc. 124 (2002) 6563–6568
- Zaia D.A., Zaia C.T., De Santana H., *Which amino acids should be used in prebiotic chemistry studies ?* Orig. Life Evol. Biosph. 38 (2008) 469-488
- Zenil H., Hernández-Orozco S., Kiani N.A., Soler-Toscano F., Rueda-Toicen A., Tegnér, J., *A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity*, Entropy, 20 (2018)
- Zhabotinski A. M., dans Frank, G. M., Ed., *Oscillating processes in biological and chemical systems*, Science Publ. Moscow, (1967), p149

Livres et chapitres de livres.

- Ahrweiler J., *Ecrits sur la vie*. Seghers, Paris (Eds), Collection dirigée par J. Ahrweiler, 1964
- Atlan H., *La fin du "tout génétique" ? Vers de nouveaux paradigmes en biologie*, Sciences en Question, INRA (Eds), 1999.
- Bachelard G., *La formation de l'esprit scientifique*, Bibliothèque des Textes Philosophiques, Librairie Philosophique J. Vrin, Paris, 1989
- Barrau A., *De la vérité dans les sciences*, 2016, Dunod
- Beal J., Phillips A., Densmore D.M., Cai Y., *High-Level Programming Languages for Biomolecular Systems*, 2011, Design and Analysis of Biomolecular Circuits, DOI: 10.1007/978-1-4419-6766-4_11
- Belousov B. P., *A periodic reaction and its mechanism*, 1951, reimprimé dans "*Oscillations and travelling waves in chemical systems*", Eds. R. Field and M. Burger. Wiley, New-York (1985) pp 605-613.
- Bonabeau E., Dorigo M., Theraulaz G., *Swarm intelligence. From natural to artificial systems*, Oxford University Press, New York, 1999
- Bray D., *Wetware. A computer in every living cell*, Yale University press, New Haven & London, 2009
- Buffon [George-Louis Leclerc, Comte de] *Histoire Naturelle Générale et Particulière : avec la description du*

Cabinet du Roi', (Histoire Naturelle : 1746-1788), Num. BNF de l'éd. de Leiden. Reprod. de l'éd. de A Paris : de l'Impr. royale ; Tome I, Premier Discours "De la manière d' étudier & de traiter l'Histoire Naturelle"

Cariou J.-Y., *Histoire des démarches scientifiques. De l'antiquité au monde contemporain*, Editions Matériologiques, 2019

Chalmer A.F., *Qu'est ce que la science ? Popper Kuhn, Lakatos, Feyerabend*, 1987, La découverte (Eds), réimpression de *What is this thing called science ? An assessment of the nature and status of science and its methods*, U. Queensland Press, St Lucia 1976.

Chapoutier G., Kaplan F., *L'homme, l'animal et la machine*, Collection *Le bouquet scientifique*, CNRS Editions (Eds), 2011

Codd E.F., *Cellular Automata*, Academic Press, New York (1968)

Dawkins R., *The selfish gene*, 1976, Trad. fr. *Le gène égoïste*, Odile Jacob, 2003

Deamer D., *Semisynthetic minimal cells: New advancements and perspectives*, *Protocells : Bridging nonliving and living matter*. Rasmussen et al. (Eds), The MIT Press, Cambridge, Massachusetts; London, England (2008)

Descartes R., *Oeuvres choisies, Tomes I et II*, Librairie Garnier Frères, paris, 1930

Deutsch D., *The fabric of reality*, The Penguin Press, NY, 1997, online : https://ia800107.us.archive.org/16/items/TheFabricOfReality/The_Fabric_of_Reality.pdf

Edelman G.M., Tononi G., *Comment la matière devient conscience*, Odile Jacob, 2000

Feyerabend P., *Contre la méthode. Esquisse d'une théorie anarchiste de la connaissance*, Points (Eds), Sciences, 2005

Glade N., Bastien O., HessamiH., *Phénomènes critiques en épidémiologie: résistance, spatialité, dynamiques humaines. Des modèles aux applications*, chapitre dans *Le Vivant Critique et Chaotique*, N. Glade & A. Stephanou (Eds), Collection "Modélisations, simulations, systèmes complexes", Editions Matériologiques, 2015

Hanczyc M.M., *The early history of protocells: The search for the recipe of Life*, *Protocells : Bridging nonliving and living matter*. Rasmussen et al. (Eds), The MIT Press, Cambridge, Massachusetts; London, England (2008)

Heisenberg W., *Le Manuscrit de 1942*, Allia, 2010

Hodges A., *Alan Turing: The Enigma*, Vintage Books, London (1992)

Hofstadter D.R., *Gödel, Escher, Bach: An eternal golden braid*, Basic Books, Inc. New York, 1979

Hofstadter D.R., *Je suis une boucle étrange*, Dunod, 2013

Isaacson W., *Les Innovateurs*. JC Lattès (Eds), 2015

Keller F., *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*. Harvard University Press, Cambridge, 2002

Kroo I., *Collectives and complex system design*. VKI lecture series on Optimization Methods & Tools for Multicriteria/Multidisciplinary Design, 2004

Kuhn T., *La structure des révolutions scientifiques*, Champs Flammarion (Eds), Sciences, 2008.

Kupiec J.-J., Gandrillon O., Morange M., Silberstein M., *Le hasard au coeur de la cellule*, Editions Matériologiques, Paris, 2016.

Leduc S., *Théorie physico-chimique de la vie et générations spontanées*, A. Poinat, 1910.

Lewis H.R., Papadimitriou C., *Elements of the Theory of Computation*. Prentice-Hall, 1982; second edition September 1997.

Liehr A.W., *Dissipative Solitons in Reaction Diffusion Systems. Mechanism, Dynamics, Interaction*. Springer Series in Synergetics, Springer, Berlin Heidelberg, vol. 70, 2013

Lodge O., *Life and Matter*, 1907, Trad. Française par J. Maxwell, *La Vie et la Matière*, Bibliothèque de Philosophie Contemporaine, Librairie Félix Alcan, Paris, 1930

Markovitsi D., Gustavsson T., Banyasz A., *DNA Fluorescence*. CRC Handbook of organic Photochemistry and Photobiology, Taylor & Francis Group, 1057-1079, 2012.

[Materiologique 2013] H. Berry , S. R. Thomas, J. Demongeot , S. Bernard, V. Volpert, N. Bessonov, N. Eymard, A. Tosenberger , A. Chauvière, H. Hatzikirou, M. Tektonidis, A. Deutsch, E. Promayon, J. Berro, S. Portet, P. Amar, P. Baconnier, E. Fanchon, S. Lespinats, P. Ballet, A. Jeannin-Girardon, A. Pothet, G. Misevic, A. Fronville, V. Rodin, A. Stéphanou, N. Glade, *Le Vivant Discret et Continu*, A. Stephanou & N. Glade (Eds), Collection "Sciences et Philosophie", Editions Matériologiques (2013)

[Materiologique 2015] F. Varenne, H. Berry , J. Demongeot , A. Chauvière, A. Stéphanou, N. Glade, M. Montévil, J.-P. Françoise, M. Lahutte-Auboin, F. Peruani, B. Caré, A. Lo Van, H. Soula, O. Bastien, H. Hessami, S. Jacquir, B. Xu, S. Binczak, J.-M. Bilbaut, V. Thomas-Vaslin, *Le Vivant Critique et Chaotique*, N. Glade & A. Stephanou (Eds), Collection "Modélisations, simulations, systèmes complexes", Editions Matériologiques (2015)

Meinhardt H., *Models of biological pattern formation*, Academic Press, London, 1982

Moatti A., *Einstein. Un siècle contre lui*, Odile Jacob, 2007

Monod J., *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Editions du Seuil, 1970

Moussaïd M., *Fouloscopie. Ce que la foule dit de nous*, Editions HumenSciences, 2019

Morin E., *Science avec Conscience*, Points (Eds) – Sciences. 2017

Murray J.D., *Mathematical biology*, Springer, 1993

Nicolis G. and Prigogine I. *Self-organisation in non-equilibrium systems*, Wiley, New York, 1977.

von Neumann J., Burks A.W., *Theory of Self-Reproducing Automata*, University of Illinois Press, 1966

Pignon D., *Ordinateur et cerveau machines - Les machines molles de von Neumann*, MAJ Masse (Eds) 1992

Popper K., *La connaissance objective*, Champs Flammarion (Eds), 1991.

Rashevsky N., *Mathematical biophysics : Physico-mathematical foundations of biology*, Dover Publications, Inc. New York, New York, 3ème édition révisée, 1948

Rennard J.-P., *Vie Artificielle - Où la biologie rencontre l'informatique*, Vuibert, 2002

Rosen R., *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of Life*, New York: Columbia Univeristy Press, 2001

Rousseau, J.-J., *Discours sur l'origine et les fondements de l'inégalité parmi les hommes. Discours sur les sciences et les arts*, Garnier-Flammarion, 1971.

Shannon C.E., Weaver W., *The mathematical Theory of communication*, University of Illinois, Urbana III, 1949

Thompson W. Sir D'Arcy (1917) *On growth and form*, Cambridge University Press. Réédition version française (Forme et croissance), 1994, le seuil.

Thorne K.S., *Trous noirs et distorsions du temps*, Champs Flammarion, Sciences, 1997.

Triclot M., *Le moment cybernétique – La constitution de la notion d'information*, Collection Milieux, Champ Vallon (Eds), 2008

Univers. *Voyage à travers l'univers – La physique du Cosmos*, Time-Life, 1989

Voltaire [Mr. de , François-Marie Arouet], (1757) Collection complète des Œuvres de Mr. de Voltaire – Première édition Tome 3, Mélanges de Philosophie avec des figures – Éléments de philosophie de Newton

Waddington C.H., *The Strategy of the Genes*, George Allen & Unwin, 1957

Wiener N., *Cybernetics: Or Control and Communication in the Animal and the Machine*. Paris, (Hermann & Cie) & Camb. Mass. (MIT Press) ISBN 978-0-262-73009-9; 1948, 2nd revised ed. 1961

Thèses et Masters.

Bastien O., Développements théoriques et méthodes numériques pour les analyses comparatives de génomes et protéomes biaisés, (2006) PhD Thesis 2003-2006

Ben Amor H.M., *Etude de la robustesse dans les réseaux de neurones et les réseaux de régulation biologique*, (2008) Stage de fin d'études d'ingénieur - ENSI Tunis – Directeurs de stage : Jacques Demongeot et Nicolas Glade

Ben Amor H.M., *Mémoire dynamique et évocation mnésique / Dynamical memory and mnesic evocation*,

(2011) PhD Thesis 2008-2011, TIMC-IMAG, Grenoble, Directeurs de thèse : Jacques Demongeot et Nicolas Glade

Christen U., *Evaluation of distance between Thresholded Boolean Automaton Networks*, (2018) M1 Internship report, Directeurs de stage : Laurent Trilling et Nicolas Glade. DOI:10.13140/RG.2.2.27568.69120 Grand prix de la recherche de l'École Polytechnique, 2018.

Elena A., *Robustesse des réseaux d'automates Booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique*, (2009) Ph.D. Thesis, TIMC-IMAG, Grenoble France, Directeur de thèse : Jacques Demongeot.

Glade N., *Effet des champs externes faibles sur l'auto-organisation des microtubules. Comparaison entre expériences et simulations numériques*. (2002) Ph.D. Thesis, CEA / TIMC-IMAG, Grenoble France, Directeurs de thèse : James Tabony et Jacques Demongeot.

Hegge C., *Étude par expérimentation numérique de la diversité et de la survie de formes de vie artificielles dans un environnement stratifié par sédimentation différentielle* (2017a) Master Thesis, TIMC-IMAG, Grenoble, France, Directeur de stage : Nicolas Glade et Sergiu Ivanov

Hessami M. H., *Modélisation multi-échelle et hybride des maladies contagieuses : vers le développement de nouveaux outils de simulation pour contrôler les épidémies*. (2016) Ph.D. Thesis, TIMC-IMAG, Grenoble France, Directeur de thèse : Nicolas Glade

Larchères A., Wang M., *Évolution artificielle des réseaux booléens à signe* (2020) TER Master 1, Directeur de stage : Sergiu Ivanov.

Segretain R., *Inference, simulation et analyse des réseaux d'automates Booléens à seuil*, (2018) Master Thesis, TIMC-IMAG, Grenoble, France, Directeur de stage : Nicolas Glade et Laurent Trilling

Tempesti G., *A Self-Repairing Multiplexer-Based FPGA Inspired by Biological Processes* (1998) Ph.D. Thesis, EPFL Lausanne, Suisse

Posters.

Bonnot A.M., Filippini C., Chaumont M., Marty L., Chenot S., Fournier T., Tabony J., Glade N., *Self organized growth mode of self-assembled carbon nanotubes synthesized by hot filament assisted CVD*, '1st Workshop of the COST Chemistry Action D27', Origin of Life And Early Evolution, Ravello, Italie.(2002)

Muller A.W., *The origin of Life explained by thermogenesis* (2019) Astrobiology Science Conference (AbSciCon) 2006, Washington DC, DOI: [10.13140/RG.2.2.24506.54721](https://doi.org/10.13140/RG.2.2.24506.54721) URL: <https://www.researchgate.net/publication/334593356> [The Origin of Life Explained by Thermosynthesis](https://www.researchgate.net/publication/334593356)

Articles de presse.

Chang K., *Life on Mars? Rover's Latest Discovery Puts It 'On the Table' - The identification of organic molecules in rocks on the red planet does not necessarily point to life there, past or present, but does indicate that some of the building blocks were present*. The New York Times. 7 June 2018.

<https://www.nytimes.com/2018/06/07/science/mars-nasa-life.html>

Communications.

[Comm. SFBT 2014] Discussion en table ronde sur la vie artificielle, XXXIV^{ème} séminaire de la SFBT, « La vie artificielle et le point de vue de la biologie théorique », St Flour, 2014

Videos.

[Adami, 2011] Adami C., *Finding alien life: Christoph Adami at TEDxUIUC*. (2011), <https://www.youtube.com/watch?v=BBao3D6sBto>

[BioVision, 2012] BioVision The inner Life in a Cell, Harvard University. (2012), <https://www.youtube.com/watch?v=WFq6aavkqG4>

[Theriot, 2016] Theriot J., *Protein Polymers, Crawling Cells and Comet Tails*. (2016) Stanford HHMI, <https://www.youtube.com/watch?v=FIT0fdt6c3Y>

Références Web.

[ATDBio] <https://www.atdbio.com/content/1/Ultraviolet-absorbance-of-oligonucleotides>

[BiochimieAnger] <http://biochimej.univ-angers.fr/>

[Edwards, 2018] Edwards C., *Sols 1913-1924: Curiosity's Working Holiday* (2018) URL: <https://mars.nasa.gov/msl/mission-updates/sols-1913-1924-curiositys-working-holiday/?mu=sols-1913-1924-curiositys-working-holiday>

[Ellis, 2018] Ellis M., *Une brève histoire du design graphique* (2018), URL: <https://99designs.fr/blog/histoire-design-genres-mouvements/histoire-design-graphique/>

[Hege; 2017b] Hege C., Glade N., *A Turing test for improving Artificial Life and Biological Models* (2017) URL: https://docs.google.com/forms/d/e/1FAIpQLSezaLyva4zCQIe32Ejp01ovxIUSc8MS1vS3W783ppUwcpFOA/viewform?usp=sf_link

[Jones, 1984] Jones D.G., Dewdney A.K., *Core War Guidelines* (1984), URL : <https://corewar.co.uk/standards/cwg.txt>

Matulewicz, C. (2017, February 8). *Is Dissociation a Coping Skill or a Dysfunction in DID?*, HealthyPlace. <https://www.healthyplace.com/blogs/dissociativeliving/2017/02/dissociation-is-not-always-dysfunction>

[Petri, 2008] Carl Adam Petri and Wolfgang Reisig (2008), Scholarpedia, 3(4):6477.

http://www.scholarpedia.org/article/Petri_net

[Phys.org, 2017] *Fossil or inorganic structure? Scientists dig into early life forms* (2017, March 15) 28 URL:

<https://phys.org/news/2017-03-fossil-inorganic-scientists-early-life.html>

[Ray, 1993] Ray S.T., *How I created Life in a virtual Universe* (1993) URL : <http://life.ou.edu/pubs/nathist/>

[SP3, 2013] SP3 convert, MIT Licence (2013), Sourceforge URL : <http://freshmeat.sourceforge.net/projects/csp-convert-sp3>

[Space.com, 2011] Moskowitz C., *Scientists Dubious Over Claim of Alien Life Evidence in Meteorite* (2011) URL: <https://www.space.com/11049-alien-life-meteorites-skepticism.html>

Logiciels.

[AEVOL] <http://www.aevol.fr/>

[BVS] <https://docs.unrealengine.com/en-US/ProgrammingAndScripting/Blueprints/index.html>

[CamiTK] <https://camitk.imag.fr/>

[HSIM] <https://www.lri.fr/~pa/Hsim/>

[JPP] Segretain R., JPP : Java-Parallel-Pipeline, DOI: 10.13140/RG.2.2.11262.10562

[NetBioDyn] <http://virtulab.univ-brest.fr/netbiodyn.html>

[PhysiBOSS] <https://github.com/gletort/PhysiBoSS>

[PhysiCell] <http://physicell.mathcancer.org/>

[SimCells] <http://virtulab.univ-brest.fr/simcells.html>

13. ANNEXES

1. Projets de recherches soumis pour financement

Ci dessous sont listés les demandes de financement faites récemment et en lien direct avec mes activités de recherche actuelles, regroupés par thèmes. Ils sont restitués tels qu'ils ont été soumis (et pour certains acceptés pour financement), en indiquant les partenaires et leurs affiliations, le contexte des projets, la méthodologie, les moyens à disposition, les résultats attendus. J'ai imaginé et rédigé ces projets. Je ne suis que co-PI pour certains pour de simples raisons de stratégie de soumission des projets.

1.1. Origines du vivant – Ecosystèmes et microcosmes.

DESCRIPTION DU PROJET IDEX CDP COOL 1 – Confirmation expérimentale de la possible origine stéréochimique du code génétique.

Scientifique responsable et Co-I. PI : Nicolas Glade (TIMC-IMAG), co-I : Eric Beaugnon (LNCMI)

Laboratoires et équipes impliqués. 1) Equipe DyCTiM2, laboratoire TIMC-IMAG, CNRS UMR5525 / Université Grenoble Alpes, 38706 La Tronche ; 2) Laboratoire LNCMI-G, CNRS UPR3228, 25 rue des martyrs, 38042, Grenoble ; 3) NanoBio Campus - 570 rue de la Chimie - 38610 Gières ; 4) Plateforme IBS, Institut de Biologie Structurale, 71 avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9

Description du projet.

Contexte. Ce projet se situe dans le contexte de la chimie prébiotique entre l'apparition des premières briques du vivant et l'émergence des premières formes (proto-métabolismes, proto-cellules ...) ordonnées et auto-entretenu, voire amplifiées, précurseurs de systèmes de complexité croissante amenant à l'apparition de la matière vivante. Il s'intéresse à une manière possible dont les métabolismes autopoïétiques [15], émergent de 'la soupe primitive', ont pu être encodés dans des génomes primitifs. Ce projet expérimental s'appuie sur l'hypothèse de l'origine stéréochimique du code génétique [3,8,11-12,14] qu'il cherche à valider.

La chimie prébiotique est souvent pensée dans le contexte de soupes chimiques homogènes contenant des composés fondamentaux comme les acides aminés, certains précurseurs des sucres, certaines bases azotées et quelques autres composés chimiquement actifs. Dans ces solutions, des molécules composées de ces éléments simples (des polypeptides et des acides polynucléotidiques (ADN, ARN, PNA)) peuvent se former spontanément ou sous l'effet de catalyseurs et posséder eux-mêmes une activité catalytique. Lorsque les conditions le permettent, en particulier lorsque la concentration en composés est suffisamment élevée, certains composés peuvent s'assembler transitoirement, établissant en un même complexe supra-moléculaire une activité composite. Des études théoriques ont montré, par simulation, qu'il devrait être possible dans ce contexte d'obtenir des cycles autopoïétiques au cours desquels sont entretenus périodiquement et sont amplifiés certains composés et complexes moléculaires [9,10,15,16]. L'obtention d'un métabolisme primordial, préalable à l'émergence de la vie organisée fait partie des hypothèses plausibles d'un point de vue théorique et expérimental [13,15]. La fixation dans une forme de génome a pu se faire ultérieurement, peut-être par l'intermédiaire d'assemblage supramoléculaires se renforçant au contact des composés les plus souvent formés dans ces métabolismes primordiaux.

Cependant, ce scénario auto-catalytique présente une faille à laquelle de nombreuses personnes ont tenté de répondre: comment s'est établie la dualité "code génétique – protéines" et ses processus associés, la traduction en particulier. Les modèles d'ensembles auto-catalytiques de Kauffman [10] et typogénétiques de Hofstadter et Varetto [9,16] admettent au préalable l'existence d'une relation fonctionnelle (la traduction implicite d'un code génétique en protéines dotées d'activité) entre les séquences polynucléotidiques et les

peptides actifs sans en préciser le mécanisme. Une hypothèse, ancienne maintenant, concernant le code génétique fut précisée en 1954 par G. Gamow [3], un an après la découverte de la structure de l'ADN par Watson et Crick. Il proposait que les acides aminés puissent s'apparier directement avec certaines portions de la séquence génétique (dans des cavités délimitées par les blocs de 4 bases que forment les échelons de la double hélice d'ADN). En 1965 était découverte la correspondance que nous connaissons aujourd'hui entre chacun des acides aminés et leur(s) anticodon(s) respectifs. Les recherches de biologie moléculaire se sont alors focalisées sur cette machinerie de traduction fondée sur l'existence de molécules adaptatrices (les ARN de transfert) pour passer d'un code (génétique, à triplets de 4 bases azotées possibles) à l'autre (protéique). Bien que les recherches concernant les origines du code génétique aient aussi rapidement suivi le même courant (recherche de molécules adaptatrices ancestrales nommées aptamères [1,2,11,12]), quelques équipes ont néanmoins continué à s'intéresser aux fondements moléculaires du code génétique et notamment à la possible relation stéréochimique directe, sans molécules adaptatrices, entre les acides aminés et leurs triplets nucléotidiques respectifs. Ces travaux n'ont été effectués qu'avec des modèles moléculaires ('Legos moléculaires' et plus récemment des modèles informatiques) et n'ont jamais été testés expérimentalement (seuls des aptamères ont été testés, souvent avec succès). Ils concluent pourtant à une possible relation stéréochimique spécifique entre les acides aminés et leurs anticodons respectifs [8,14].

Objectifs. Nous souhaitons réaliser une étude exploratoire dans le contexte des origines du vivant. Ce projet est expérimental. Il consiste en la mesure, par des techniques de calorimétrie, de RMN et de fluorimétrie, des forces d'interaction entre des acides-aminés et leurs anticodons respectifs dans le code génétique, et à mettre en évidence la spécificité de cette reconnaissance amino-acide – anticodon. Cette partie qui a des retombées potentielles importantes dans le domaine de la régulation cellulaire [1] visera à confirmer les résultats théoriques de Pelc et Welton [14] et de Hendry et al [8] favorables à une hypothèse stéréochimique du code génétique aux origines du vivant. Cette validation est essentielle pour accepter les modèles autopoïétiques de chimie prébiotique (en particulier les modèles typogénétiques) [9,10,15,16] car elle résoudrait en grande partie le problème de la traduction ancestrale de séquences polynucléotidiques en peptides. Ces mesures n'ont jamais été réalisées.

Méthodologie. L'objectif de notre expérience est de montrer que non seulement de simples triplets de nucléotides peuvent, avec une affinité faible mais non négligeable et surtout spécifique, complexer leurs acides-aminés correspondants (dans le code génétique), mais aussi que l'on peut confiner (rendre proches voisins) au moins 2 acides aminés donnés sur leurs triplets correspondants dans un hexanucléotide (séquence de 6 nucléotides, donc 2 codons adjacents). Nous utiliserons ainsi des di-codons mais également des polycodons de 21 à 30 nucléotides.

Un préalable à ces expériences a été de déterminer les séquences hexanucléotidiques (parmi les 4096 (4^6) existantes) que nous pourrions utiliser (il ne faut pas qu'elles puissent s'auto-hybrider et devenir compétitrices des acides aminés ; il faut aussi éviter qu'elles contiennent des anticodons parasites (un hexanucléotide contient 4 codons et nous souhaitons que seuls les 2 anticodons adjacents participent spécifiquement)) et celles qui donnent le plus de chance de réussite pour les premières mesures (celles qui contiennent les anticodons capables du plus grand nombre d'interactions avec leurs acides aminés). N. Glade a développé un programme pour 'scorer' ces séquences en se servant notamment des données de S. Pelc et M. Welton [14]. Étant donné les faibles niveaux d'affinités auxquels s'adresse cette expérience (les interactions entre un acide aminé et son triplet nucléotidique correspondant se limitent à seulement 4 à 6 liaisons hydrogène et électrostatiques, et quelques liaisons de van der Waals ; les acides aminés présentant un cycle peuvent en outre établir des liaisons hydrophobes avec les bases azotées [14]), nos méthodes de détection et de mesure des interactions doivent être très sensibles.

Le LNCMI avec N. Glade collabore depuis longtemps, dispose d'une plateforme RMN qui nous permettra de mettre en évidence et mesurer la force des interactions entre polycodons et acides aminés. Cela se traduira notamment par des effets d'élargissement des raies spectrales. Dans ce même laboratoire pourront aussi être menées des mesures de fluorimétrie : l'excitation dans l'UV des composés présents pourrait provoquer des transferts d'énergie (FRET) vers les molécules spatialement proches. Nous nous attendons à observer

notamment des transferts d'énergie des bases azotées (donneuses) vers les acides aminés (accepteurs), et ainsi une diminution de fluorescence des polynucléotides. Dans ce contexte, l'usage de bases analogues dotées de propriétés de fluorescence sera envisagé [17]. L'équipe 'Thermodynamique des petits systèmes' du département de la Matière Condensée et des Basses Températures du CNRS Grenoble dispose du matériel et du savoir faire pour détecter et mesurer des très faibles interactions dans de très faibles volumes avec des instruments de calorimétrie dédiés (micro-titration calorimétrique dans des gammes de température de 0,05K à 400 Kelvin par nanotitration NanoITC III)[4,5]. Nous allons procéder à des injections de quantités croissantes d'acides aminés dans des solutions de polynucléotides dans un tampon phosphate (peu enclin à interagir avec nos composés). Nous allons ainsi mesurer, par ITC (isotherm titration calorimetry) l'affinité de fixation (K_a), les variations d'enthalpie (ΔH) et la stoechiométrie de fixation de ces interactions. Nous procéderons ainsi pour des jeux d'acides aminés correspondants ou au contraire ne correspondant pas aux polycodons en solution.

Ce travail interdisciplinaire réalisable par un stagiaire de M2 permettra de vérifier un scénario plausible d'une étape cruciale des origines du vivant : il fournira des éléments de preuves expérimentales sur la possible origine stéréochimique (fortement pressentie et appuyée par plusieurs travaux théoriques, mais jamais montrée expérimentalement) du code génétique. Ce travail apportera ainsi une contribution à la compréhension des origines du vivant sur Terre et de la régulation génétique dans les cellules actuelles par les petits métabolites.

Références.

1. Demongeot J., Glade N., Moreira A., Vial L., *RNA Relics and the Origin of Life*. Int. J. Mol. Sci. 10 (2009) 3420
2. Demongeot J., Glade N., Moreira A., *Evolution and RNA relics. A system biology view*, Acta Biotheor. 56 (2008) 5
3. Gamow, G., *Possible relation between deoxyribonucleic acid and protein structures*. Nature 173 (1954) 318
4. Garden J.-L., Guillou H., Lopeandia A.F., Richard J., Heron J.-S., Souche G., Ong F., Vianay B., Bourgeois H., *Thermodynamics of small systems by nanocalorimetry: from physical to biological nano-objects*. ThermoChim. Acta, 492 (2009) 16
5. Garden J.-L., Richard J., Guillou H., *Temperature of systems out of thermodynamic equilibrium*. J. Chem. Phys. 129 (2008) epub044508
8. Hendry L.B., Bransome E.D., Hutson M.S., Campbell L.K., *First approximation of the stereochemical rationale for the genetic code based on the topography and physicochemical properties of "cavities" constructed from models of DNA*. Proc. Natl. Acad. Sci. USA 78 (1981) 7440
9. Hofstadter D.R., *Gödel, Escher, Bach: An eternal golden braid*, Basic Books, Inc. New York, 1979
10. Kauffman S., *Autocatalytic sets of proteins*, J. Theor. Biol. (1986) 119, 1
11. Knight R.D., Landweber L.F., *Rhyme or reason: RNA-arginine interactions and the genetic code*. Chem. & Biol. 5 (1998) R215
12. Knight R.D., Landweber L.F., *The early evolution of the genetic code*. Cell, 101 (2000) 569
13. Muchowska K.B., Varma S.J., Chevallot-Beroux E., Lethuillier-Karl L., Li G., Moran J., *Metals promote sequences of the reverse Krebs cycle*, Nat. Ecol. Evol., 1 (2017) 1716
14. Pelc S.R., Welton M.G.E., *Stereochemical relationship between coding triplets and amino-acids*, Nature, 209 (1966) 868
15. Varela F.G., Maturana F.H., Uribe R., *Autopoiesis: The organization of living systems, its characterization and a model*, Biosystems, 5 (1974) 187
16. Varetto J., *Typogenetics: an artificial genetic system*, J. Theor. Biol., 193 (1993) 257
17. Wilhelmsson L.M., *Fluorescent nucleic acid base analogues*, Q. Rev. Biophys., 2 (2010) 159

DESCRIPTION DU PROJET IDEX CDP COOL 2 – Modélisation des phénomènes spatiaux émergents aux origines du vivant.

Scientifique responsable et Co-I. PI : Nicolas Glade (TIMC-IMAG), co-I : Olivier Bastien (LPCV)

Laboratoires et équipes impliqués. 1) Equipe DyCTiM2, laboratoire TIMC-IMAG, CNRS UMR5525, Université Grenoble Alpes. ; 2) Laboratoire LPCV - CEA, UGA - CNRS UMR5168, INRA USC1359, rue des martyrs, 38054, Grenoble

Description du projet.

Contexte. Ce projet se situe dans le contexte de la chimie prébiotique entre l'apparition des premières briques du vivant et l'émergence des premières formes (proto-métabolismes, proto-cellules ...[15]) ordonnées et auto-entretenues et auto-amplifiées, précurseurs de systèmes de complexité croissante amenant à l'apparition de la matière vivante. Il est focalisé en particulier sur les processus autopoïétiques [18] permettant l'émergence de

la complexité fonctionnelle et spatiale à partir de 'la soupe primitive', c'est-à-dire l'émergence des premières boucles métaboliques et leur auto-organisation dans l'espace sous forme de micro-environnements réactionnels.

La chimie prébiotique est souvent pensée dans le contexte de soupes chimiques homogènes contenant des composés fondamentaux comme les acides aminés, certains précurseurs des sucres, certaines bases azotées et quelques autres composés chimiquement actifs. Dans ces solutions, des molécules composées de ces éléments simples (des polypeptides et des acides polynucléotidiques (ADN, ARN, PNA)) peuvent se former spontanément ou sous l'effet de catalyseurs et posséder eux-mêmes une activité catalytique [12]. Lorsque les conditions le permettent, en particulier lorsque la concentration en composés est suffisamment élevée, certains composés peuvent s'assembler transitoirement, établissant en un même complexe supra-moléculaire une activité composite. Des études théoriques ont montré, par simulation, qu'il devrait être possible dans ce contexte d'obtenir des cycles autopoïétiques dissipatifs au cours desquels sont entretenus périodiquement et sont amplifiés certains composés et complexes moléculaires [10,18,19]. L'obtention d'un métabolisme primordial, préalable à l'émergence de la vie organisée fait partie des hypothèses plausibles d'un point de vue théorique et expérimental [12,18]. La fixation dans une forme de génome a pu se faire plus tardivement par l'intermédiaire d'assemblage supramoléculaires se renforçant au contact des composés les plus souvent formés dans ces métabolismes primordiaux.

Les modèles d'autopoïèse auxquels ce projet s'intéresse s'apparentent aux ensembles auto-catalytiques de Kauffman [10] et typogénétiques de Hofstadter et Varetto [19]. Leur principe est qu'à partir de briques de base interagissant faiblement entre elles et de faible réactivité, puissent se former, rarement, des briques plus complexes possédant parfois une activité simple (faciliter la liaison entre 2 autres composés en les complexant plus longtemps, avoir une activité lytique, ...). De là, un large bestiaire hétéroclite de composés se forme. Dans de tels systèmes, lourds à simuler, il a été montré qu'avec le temps, finissent par naître des cycles ouverts auto-entretenus et auto-amplifiés, tel le cycle de Krebs actuellement. Ces modèles reposent néanmoins sur un préalable : la dualité métabolisme / encodage génétique, l'existence d'une relation fonctionnelle (la traduction implicite d'un code génétique en protéines dotées d'activité) entre les séquences polynucléotidiques et les peptides actifs sans en préciser le mécanisme. Un mécanisme primordial proposé pourrait être celui de l'hypothèse stéréochimique du code génétique [8,14]. C'est une hypothèse qu'un autre projet, expérimental, soumis au CDP Origins of Life, entend renforcer.

La limitation de ces simulations de processus autopoïétiques est qu'elles ont toujours été faites en faisant abstraction de l'espace ; ces modèles omettent donc les processus physiques de transport comme la diffusion qui s'opposent au maintien de la cohérence de systèmes fonctionnels émergents et auto-entretenus. Les boucles métaboliques primordiales variées ayant émergé du 'chaos' physico-chimique régnant sont fragiles et requièrent en effet pour être maintenue une densité suffisante. L'auto-organisation spatio-temporelle (par des processus statiques d'une part comme les processus d'agrégation et d'autre part dynamiques par réaction-diffusion comme dans les milieux excitables de type Belousov-Zhabotinskii) et le confinement sur des surfaces (cristaux, empilements d'argiles), au niveau d'interfaces (front d'ondes chimiques stationnaires obtenus dans les cas d'instabilités de Turing [17], ou les séparateurs entre 2 phases liquides) [4,7,9,13] ou dans des régions closes (vésicules, suspensions lamellaires, micelles) [9,15] sont des processus qui permettent au contraire le maintien de cette cohérence locale. Cela autorise également l'obtention et le maintien à travers l'espace d'une grande variété de systèmes émergents [7]. Avec P. Ballet (Univ. Brest), nous avons réalisé une étude de l'effet de l'apport d'énergie sous forme d'agitation thermique et d'énergie provenant d'un champ vectoriel (gravité) sur un automate cellulaire (le *Jeu de la vie*). Ces travaux [7] nous ont permis de montrer une augmentation des 'espèces intéressantes' lorsque l'agitation thermique et la gravité étaient couplées. A l'interface entre une 'boue active' résultant de la sédimentation différentielle des espèces (liée à leur masse) et un 'océan' principalement vide d'espèces, se développaient de nombreuses espèces intéressantes, souvent propulsées dans l'océan. Une exploration systématique de l'espace des paramètres nous a permis de montrer une augmentation de la durée de vie moyenne de ces espèces intéressantes dans ces conditions.

Description du projet. Nous souhaitons réaliser une étude exploratoire théorique dans le contexte des origines du vivant. Plus particulièrement il s'agit d'étudier les facteurs qui peuvent favoriser un accroissement

de la diversité d'espèces biochimiques actives (molécules et complexes moléculaires), leurs interactions dans une soupe prébiotique et l'émergence de cycles autopoïétiques, tels que le confinement dans des hétérogénéités auto-organisées ou au niveau d'interfaces. Le travail consistera d'une part en l'apport de la spatialité dans les modèles de prébiotique de type typogénétique [19], et d'autre part leur couplage avec des processus physiques d'agrégation-sédimentation ou d'adsorption sur des surfaces en suspension [7,9]. Nous souhaitons ainsi provoquer l'émergence d'auto-organisations spatiales soit stationnaires, soit sous la forme d'ondes chimiques mobiles comme dans les milieux excitable de type Belousov-Zhabotinskii (BZ). Les motifs (bandes ou spots) ainsi formés seraient constitués de 'molécules' formées au cours de ces processus prébiotiques. Une des conditions nécessaires pour l'existence de ces motifs est que la production de ces molécules soit maintenue et amplifiée, c'est-à-dire qu'elles fassent parties de boucles d'auto-amplification. Nous pensons que les interfaces entre 2 milieux de composition différente, créées uniquement par des processus chimiques (type Turing ou BZ) ou résultant de mécanismes exogènes (notamment par séparation de phase liée à la gravité), ont pu servir, avant l'apparition des premières membranes, de proto-cellules, de lieux de confinement protégeant et amplifiant les processus prébiotiques autopoïétiques comme cela a été proposé dans [4,7,9,13,15]. L'instabilité de Turing, ou la formation d'ondes chimiques étant liées à un rapport entre les processus réactifs (faisant intervenir une boucle de rétroaction négative) et les processus de transport, nous souhaitons déterminer les conditions physico-chimiques nécessaires, dans des systèmes typogénétiques, pour obtenir l'émergence auto-organisations spatiales dans ces systèmes. Par des expérimentations numériques de biochimie artificielle nous montrerons que les frontières réactives stationnaires ou mobiles peuvent jouer le rôle de proto-membranes [4,7,9,13,15] isolant ces processus prébiotiques fragiles d'autres se développant en d'autres endroits de l'espace. Du point de vue technique, nous utiliserons un moteur de simulation haute-performance pour la modélisation des systèmes hybrides (multi-échelles, multi-formalismes) développé par N. Glade au laboratoire TIMC-IMAG. Ce moteur de simulation récent bénéficie des derniers raffinements en termes de calcul (calcul GPU en particulier). La partie coûteuse en calcul concernera le stockage, en chaque point de l'espace de larges dictionnaires répertoriant les différents composés générés et leurs quantités, ainsi que le moteur de simulation stochastique (algorithme de Gillespie [6]) déterminant les probabilités d'interactions entre ces composés en fonction de leurs quantités respectives. Un des challenges de ce projet est l'implémentation de cette partie en utilisant des tables de hachage performantes [16]. C'est sur ce sujet de modélisation que sera proposé un stage de M2 en informatique à TIMC-IMAG. Nous souhaitons également caractériser les métabolismes obtenus, les étudier comme nous le faisons les métabolismes actuels [11], afin notamment de comparer leur typologie avec ces derniers. Parmi les caractères que nous pouvons notamment étudier et pour lesquels nous avons de l'expérience [2,3,5,11,20] figurent la complexité du réseau, sa connectivité, mais également la complexité des composés (le potentiel chimique) engagés. Cela requerra une capacité à calculer des distances entre réseaux et à les classifier que nous avons développé récemment [2-4,11,20]. Ce travail pouvant être engagé immédiatement en simulant des systèmes typogénétiques en espace homogène, engagera un autre stagiaire de M2 à LPCV.

A terme, nous souhaitons développer le concept autopoïétique proposé par Varela [18] en le contraignant au concept de viabilité développé par Aubin [1], comme nous sommes en train de le faire actuellement à TIMC-IMAG pour les réseaux Booléens utilisés dans la modélisation des systèmes de régulation génétique et métabolique [2,3,20].

Références.

1. Aubin J.-P., Bayen A.M. Saint-Pierre P., *Viability Theory*, Springer, 2011
2. Christen U., Ivanov S., Segretain R., Trilling L., Glade N., *On computing structural and behavioural complexities of threshold Boolean networks*, Submitted to *Acta Biotheor.* (2018)
3. Christen U., Ivanov S., Segretain R., Trilling L., Glade N., *On computing structural and behavioural distances between threshold Boolean networks*, Submitted to *Acta Biotheor.* (2018)
4. Demongeot J., Glade N., Moreira A., Vial L., *RNA Relics and the Origin of Life*. *Int. J. Mol. Sci.* 10 (2009) 3420
5. Demongeot J., Glade N., Moreira A., *Evolution and RNA relics. A system biology view*, *Acta Biotheor.* 56 (2008) 5
6. Gillespie D.T., *Exact Stochastic Simulation of Coupled Chemical Reactions*, *J. Phys. Chem.* 81 (1977) 2340
7. Glade N., Bastien Q., Ballet P., Diversity and survival of artificial lifeforms under sedimentation and random motion, *Theor. Biosci.* (2017) doi: 10.1007/s12064-017-0254-1
8. Hendry, L. B., Bransome, E. D., Hutson, M. S., Campbell, L. K.: First approximation of the stereochemical rationale for the genetic code based on the topography and physicochemical properties of "cavities" constructed from models of DNA. *Proc. Natl. Acad. Sci. USA* 78 (1981) 7440
9. Hunding A., Képès F., Lancet D., Minsky A., Norris V., Raine D., Sriram K., Root-Bernstein R., Compositional complementarity and

- prebiotic ecology in the origin of life, *Bioessays* (2006) 28, 399-412
10. Kauffman S., Autocatalytic sets of proteins, *J. Theor. Biol.* (1986) 119, 1
11. Marechal E., Bastien O., Modeling of regulatory loops controlling galactolipid biosynthesis in the inner envelope membrane of chloroplasts, *J. Theor. Biol.*, 361 (2014) 1
12. Muchowska K.B., Varma S.J., Chevallot-Beroux E., Lethuillier-Karl L., Li G., Moran J., *Metals promote sequences of the reverse Krebs cycle*, *Nat. Ecol. Evol.*, 1 (2017) 1716
13. Norris V., Hunding A., Képès F., Lancet D., Minsky A., Raine D., Root-Bernstein R., Sriram K., Question 7: the first units of life were not simple cells, *Orig. Life Evol. Biosph.* (2007) 37, 429-443
14. Pelc S.R., Welton M.G.E., *Stereochemical relationship between coding triplets and amino-acids*, *Nature*, 209 (1966) 868
15. Rasmussen S., *Protocells*, MIT Press, 2008
16. Skarupke M., *I Wrote The Fastest Hashtable*, <https://probablydance.com>
17. Turing A., *The Chemical Basis of Morphogenesis*, *Phil. Trans. Roy. Soc. London*, 237 (1952) 37
18. Varela F.G., Maturana F.H., Uribe R., *Autopoiesis: The organization of living systems, its characterization and a model*, *Biosyst.*, 5 (1974) 187
19. Varetto J., Typogenetics: an artificial genetic system, *J. Theor. Biol.*, 193 (1993) 257
20. Vuong Q.T., Chauvin R., Ivanov S., Glade N., Trilling L., *A logical constraint-based approach to infer and explore diversity and composition in thresholded Boolean automaton networks*, *Studies in Computational Intelligence Series, Proceedings of the Complex Networks 2017 conference* (2017)

1.2. Réseaux de régulation génétique et réseaux d'oscillateurs – Application à la théorie ataviste du cancer

DESCRIPTION DU PROJET IDEX IRS 2019 – Inférence de réseaux Booléens à seuil par programmation logique par contrainte pour modéliser l'homéostasie cellulaire et tester la théorie atavistique du cancer.

Scientifique responsable et Co-I. PI : Nicolas Glade (TIMC-IMAG)

Partenaires. Pascal Ballet (LATIM, UBO Brest), Arnaud Chauvière (TIMC-IMAG), Ibrahim Cheddadi (TIMC-IMAG), Eric Fanchon (TIMC-IMAG), Nicolas Glade (TIMC-IMAG), Sergiu Ivanov (Université Evry-val d'Essonne), Claudio Lobos (Universidad Técnica Federico Santa María, Chili), Angélique Stephanou (TIMC-IMAG), Laurent Trilling (TIMC-IMAG)

Laboratoires et équipes impliqués. 1) Equipe DyCTiM2, laboratoire TIMC-IMAG, CNRS UMR5525 / Université Grenoble Alpes, 38706 La Tronche ; 2) IBISC, Université Evry-val d'Essonne ; 3) Universidad Técnica Federico Santa María, Chili

Positionnement, contexte et objectifs.

Notre projet de modélisation en biologie se place dans le contexte des réseaux de régulation biologique et de la tumorigenèse. Il s'agit d'un projet interdisciplinaire à visée applicative en biomédecine. Il sera réalisé à TIMC-IMAG où il fera l'objet d'un dialogue permanent entre biologistes et modélisateurs. Le laboratoire TIMC-IMAG a en effet un long savoir faire dans les mathématiques et informatique appliquées à la biologie et à la médecine et, parmi les thématiques biomédicales qui y sont développées actuellement, en particulier au sein des équipes BCM et DyCTiM2, la différenciation cellulaire et tissulaire et la tumorigenèse figurent en bonne place [4,20,23-27]. La régulation ou au contraire le dérèglement génétique et métabolique font partie des mécanismes microscopiques sous-jacents dont dépendent ces processus ; ce sont thèmes de recherche également très développés à TIMC-IMAG [2,3,5-7,9,10,12,22,24,30]. Ce projet est à dominante MSTIC : ce travail impliquera plusieurs développements en informatique concernant les réseaux d'automates Booléens à seuil et il devra relever plusieurs défis numériques de modélisation hybride de tissus biologiques.

Durant ce projet, nous souhaitons tester, par modélisation, la théorie atavistique du cancer qui propose que la prolifération tumorale soit davantage liée à la perte de contrôle d'un ensemble de systèmes génétiques ancestraux, toujours présents mais inhibés dans les cellules actuelles, assurant la survie des organismes unicellulaires [8,15,18,29]. Cette théorie atavistique s'appuie notamment sur de solides observations de phylostratigraphie faisant le constat de la sur-expression de gènes qualifiés d'ancestraux, et sur l'inhibition progressive des systèmes génétiques récemment acquis par les organismes évolués. Ceci a pour conséquence une perte progressive, au fur et à mesure de l'évolution des grades tumoraux, des caractères de la vie pluricellulaire et une ré-acquisition des caractères adaptés à la vie unicellulaire.

La théorie mathématique de la viabilité, initialement développée dans le cadre des systèmes dynamiques [1] et que nous cherchons actuellement à appliquer aux réseaux Booléens à seuil, est bien adapté pour réfléchir

sur ce problème. La théorie de la viabilité permet d'exprimer les contraintes d'atteignabilité de fonctions cibles par un ensemble de systèmes évolutifs (systèmes d'équations différentielles, équations mutationnelles, réseaux d'automates Booléens ...). Elle permet notamment de déterminer l'ensemble des stratégies existantes (ensemble des trajectoires qu'une collection de systèmes évolutifs peut suivre pour réaliser une fonction) dans un système vivant (et que l'évolution des espèces a contraint) pour s'adapter aux différentes situations qu'il peut rencontrer, y compris un stress cellulaire. Le phénomène de déprogrammation progressive des fonctions essentielles et le changement de mode de fonctionnement correspondent dans ce cadre à un éloignement du système vivant d'un noyau de viabilité 'normal' pour rejoindre peu à peu un autre plus adapté à sa situation. Cela peut donc être compris comme une adaptation des cellules à un stress : elles deviennent tumorales et adaptent leur environnement avec. La théorie atavistique soutient que le passage à l'état tumoral et la progression de la tumeur ne relève pas ou peu de mutations ponctuelles dans l'ADN ; cela implique en revanche une "déprogrammation" pas-à-pas (par strates) des systèmes de régulation. Les fonctions cellulaires normales ainsi que leur dérive épigénétique par dérégulations successives vers des systèmes de régulation anormaux peuvent être modélisés dans le cadre formel des réseaux d'automates Booléens à seuil. A TIMC-IMAG [2,3,5,6,7,9,10,12,30], ces dernières années, nous avons considérablement développé des méthodes de programmation logique par contraintes [11] dans le cadre des réseaux Booléens à seuil [14,21,28,31]. Nous pouvons ainsi déterminer, par inférence, de larges ensembles de réseaux satisfiables pour une dynamique (fonction) donnée et pour un ensemble de connaissances structurelles (existence de gènes ; relations entre gènes) [5,6,30]. En biologie, on associe souvent à une fonction (de régulation notamment) une structure de réseau unique. Les systèmes biologiques sont pourtant imparfaits, certaines interactions restant faibles ou peu efficaces, changeantes dans le temps... La fonction vitale doit cependant être maintenue. Nous proposons que, considérer un ensemble entier de réseaux, tous différents, certains se ressemblant, d'autres moins [5,6], mais tous réalisant la même fonction, permet de prendre en compte, dans la modélisation des réseaux de régulation, une forme de robustesse au stress : les réseaux satisfiables pour une fonction et leurs trajectoires menant à cette même fonction constituent un noyau de viabilité.

En définitive, nous proposons un projet novateur car s'appuyant sur de multiples changements conceptuels (théorie atavistique du cancer, ensemble de réseaux satisfiables, théorie de la viabilité) pour modéliser le déclenchement et l'évolution d'un tissu cellulaire vers une forme cancéreuse. Ces concepts prendront forme au cours d'une thèse, via la modélisation d'un système tissulaire homéostatique.

Programme et résultats attendus.

Le projet consiste à réaliser un modèle numérique de tissu vivant doté de systèmes de régulation génétique munissant les cellules de moyens d'intégration de stimuli externes (biomécaniques, chimiques ...) et leur donnant un cycle de vie (divisions, quiescence) et un comportement de cellule sociale (capacité d'adhérence, synchronisations cellulaires ...). Il y a donc deux composantes distinctes dans cette réalisation :

- la recherche de réseaux de régulation adaptés à notre problématique par inférence
- la modélisation d'un tissu cellulaire ayant des propriétés biomécaniques

Ces deux parties dont nous allons détailler la mise en œuvre devront s'articuler entre elles.

Un effort initial important sera fourni sur la partie concernant les réseaux de régulation. Ceux-ci devront notamment contrôler les comportements individuels et sociaux des cellules et disposer de points d'entrée (nœuds d'entrée) pour rendre les cellules sensibles aux stimuli externes (biomécaniques) que nous aurons définis. Ce travail de traduction des connaissances biologiques en systèmes de régulation Booléens se fera en étroite collaboration avec les biologistes de l'équipe. L'intégration dans un modèle de biomécanique tissulaire se fera dans un second temps avec les modélisateurs de l'équipe.

En pratique, nous proposons de faire avancer le projet selon une méthode SCRUM, à savoir une méthode itérative incrémentale centrée sur la réalisation d'un produit fonctionnel. L'idée est de commencer par la détermination de réseaux ayant des fonctionnalités simples comme le contrôle de l'adhérence cellulaire, de réaliser un modèle de biomécanique tissulaire simple, puis de réunir les deux. Une fois ce modèle simple fonctionnel, le cycle de développement pourra être repris avec la détermination de réseaux plus complexes assurant à la fois la fonction précédente (ex : adhérence) et de nouvelles fonctions, notamment un cycle

cellulaire dépendant de l'état biomécanique cellulaire... et ainsi de suite jusqu'à obtention d'un modèle de plus en plus complexe, c'est-à-dire possédant de plus en plus de couches imbriquées de fonctionnalités biomécaniques et de cycle cellulaires.

A chaque étape du développement, le modèle fonctionnel pourra être utilisé pour étudier la théorie atavistique : le système cellulaire subira des stress biomécaniques importants ou répétés, visant à éloigner le système de l'homéostasie. La robustesse du noyau de viabilité associé à la fonction sera étudiée. La bascule vers des comportements primitifs (perte d'une fonctionnalité essentielle à l'homéostasie tissulaire, adoption de comportements antagonistes au maintien tissulaire) pourra aussi être étudiée à chaque étape. De même que la recherche de perturbations permettant la récupération des fonctionnalités importantes dans le maintien tissulaire qui permettra de tester des possibilités de thérapies fondées sur le principe du rétablissement de la 'normalité cellulaire' par forçage des conditions environnementales normales.

Donnons maintenant le détail des deux composantes de notre projet :

1. Inférence et modélisation des réseaux Booléens à seuil permettant la régulation de la biomécanique permettant la morphogenèse et l'homéostasie des tissus cellulaires.

De nombreuses fonctions sont affectées par la déprogrammation ayant lieu dans les cellules cancéreuses. Il n'est pas possible de toutes les prendre en compte. Pour tester la théorie atavistique du cancer, il nous paraît judicieux de nous appuyer sur un système emblématique affecté dans cette théorie : celui qui permet le passage d'un mode de vie autonome, unicellulaire, à un mode de vie social, pluricellulaire, c'est-à-dire celui qui permet aux cellules d'adhérer entre elles, de se synchroniser, de s'intercaler ou de modifier leurs formes pour réduire des tensions tissulaires, ... le système qui permet la régulation [29] de la morphogenèse et de l'homéostasie tissulaire d'un point de vue mécanique [13]. Le choix de ce système de régulation biomécanique présente d'autres intérêts : notre équipe est composée de plusieurs chercheurs spécialisés dans la biomécanique des cellules [4,19,20,23,25,26] ; il présente aussi l'avantage de pouvoir être facilement intégré dans un modèle de tissu cellulaire.

Quels que soient les systèmes régulés, les réseaux de régulation que nous devons chercher doivent être composés de sous réseaux successivement imbriqués les un dans les autres pour correspondre aux strates phylogénétiques mises en évidence par M. Vincent [28], chaque strate supérieure (mise en place plus récemment dans l'évolution) contrôlant les strates plus primitives. Dans le cas de la biomécanique tissulaire, les cellules cancéreuses en devenir peuvent ainsi perdre successivement différentes fonctionnalités impliquées dans la morphogenèse et l'homéostasie tissulaire.

Pour modéliser ces réseaux de régulation, nous allons travailler avec les réseaux d'automates Booléens à seuil (RABS) [2,3,5,6,9,10,12,14,21,28,30,31] qui présentent l'avantage de capturer les relations génétiques avec peu de paramètres. Les fonctions Booléennes à seuil les composant sont définies par leurs paramètres (un seuil et 1..N poids pour les arcs entrants), à la différence des fonctions Booléennes utilisées dans les réseaux de Thomas qui sont définies par des tables de vérité. De plus, les RABS peuvent être facilement assimilés à des hybrides entre réseaux Booléens (dont la dynamique est facilement interprétable) et systèmes différentiels (facilitant ainsi l'analyse mathématique). Enfin, nous savons bien travailler avec ces RABS ; en particulier, nous avons montré que leur structure, leur dynamique et leurs propriétés peuvent être déterminées par inférence [2,3,5,6,10,30] en utilisant un langage de programmation logique non-monotone par contrainte, le langage ASP (Answer Set Programming)[11]. Toute la connaissance biologique disponible, structurelle comme dynamique, est encodée sous la forme de combinaisons de contraintes, elles-mêmes traduites sous une forme logique (CNF) transmise à un solveur de satisfiabilité (type SAT). Le résultat est l'obtention d'ensembles de réseaux dits satisfiables, c'est-à-dire vérifiant tous les contraintes de structure et de dynamique imposées. Nous allons maintenant calculer la distance à la fois structurelle (proximité dans l'espace des paramètres) [5] et dynamique (bassins d'attraction et trajectoires) entre réseaux afin de déterminer un espace composé de l'ensemble des réseaux. Dans cet ensemble de systèmes évolutifs (ici des réseaux), certains possèdent des trajectoires viables (menant à la fonction requise) ; la conjonction constitue le noyau de viabilité dans lequel évolue un système à l'état physiologique. La sortie du noyau de viabilité est faite par éloignements successifs d'une part d'un point de vue fonctionnel (sortie des bassins d'attraction

menant à la fonction), ce qui s'apparente à la dérive fonctionnelle et épigénétique d'un système vivant subissant un stress important, mais aussi, progressivement, d'un point de vue structurel (changement radical de structure du réseau par variations successives de ses paramètres), ceci s'apparentant à la survenue de mutations non corrigées se produisant dans le génome de ces organismes. Nous allons chercher à appliquer la théorie de la viabilité développée par Aubin [1] pour les systèmes dynamiques, aux réseaux d'automates Booléens à seuil, afin de déterminer les contours de ces noyaux de viabilité, leurs maxima (zones de plus forte robustesse), et quantifier localement la robustesse associée à chaque couple (structure,dynamique) de cette espace. L'application de la théorie de la viabilité permettra d'en utiliser les théorèmes puissants. Ce travail déjà en développement et dont la publication débute [5,6], a été initié durant un stage de recherche co-encadré par L. Trilling et N. Glade, et a remporté le Grand Prix du Stage de Recherche de l'Ecole Polytechnique (Palaiseau) cette année.

2. Modélisation de l'homéostasie d'un tissu épithélial doté de comportements biomécaniques.

Cette partie s'appuiera sur les travaux existants de plusieurs participants de ce projet pour modéliser des tissus vivants [4,25,26], en particulier pour les aspects biomécaniques et computationnels [16,17,19,20,23]. La particularité qui devra être prise en compte est l'aspect modulaire, désactivable, de certains mécanismes biomécaniques indispensables à l'obtention d'un tissu épithélial homéostatique [13]. En effet, nous devons pouvoir, comme le prédit la théorie atavistique du cancer, progressivement désactiver, un à un, ces mécanismes, par le biais de l'éloignement progressif en fonctionnalité et en structure du noyau de viabilité définissant les systèmes de régulation (réseaux d'automates Booléens à seuil dans des états viables) de l'homéostasie biomécanique du tissu.

Les résultats attendus sont d'une part des livrables logiciels (logiciels développés pour l'inférence et l'analyse de réseaux d'automates Booléens à seuil ; algorithmes et implémentations numériques de la biomécanique tissulaire et d'un modèle hybride adjoignant un système de régulation génétique fondé sur des collections de réseaux d'automates Booléens à seuil), d'autre part une production scientifique écrite d'abord sur les développements qui continueront d'être faits en informatique théorique et en mathématiques appliquées concernant les réseaux d'automates Booléens à seuil (inférence, distances entre réseaux, viabilité, ...), ensuite sur le modèle biomécanique modulaire, sur son hybridation avec le système de régulation génétique aussi, enfin sur les enseignements que nous tirerons de cette étude théorique concernant la théorie atavistique du cancer.

Organisation.

Comme dit ci-dessus, les réalisations du projet suivront un cycle de développement SCRUM : la méthode incrémentale est en effet plus sûre pour faire progresser de façon viable le développement d'une étude impliquant une modélisation hybride. Les itérations seront ponctuées dans la mesure du possible de communications scientifiques orales ou écrites.

Un effort sera cependant fait au début du projet sur la recherche de réseaux de régulation permettant le contrôle, par couches interdépendantes, des processus biomécaniques et du cycle cellulaire assurant l'homéostasie d'un tissu de cellules.

Nous souhaitons très rapidement pouvoir disposer de premiers modèles hybrides, même simples, et ne pas chercher d'emblée à établir des modèles complexes. La complexification de fera au fur et à mesure des itérations du projet.

Comme le projet fait intervenir plusieurs participants, certains hors TIMC-IMAG, nous allons mettre en place, comme nous l'avons fait dans plusieurs projets que nous avons mené à bien ou encore en cours, des outils de travail collaboratifs. Parmi ces outils, nous utiliserons une plateforme GitLab pour la maintenance et la gestion des versions des programmes qui seront développés. Nous utiliserons aussi Gitter, un système particulièrement performant de communications instantanées conservant la trace de toutes les communications ayant été faites dans un canal de communication, permettant de poster code informatique, équations mathématiques et figures, et permettant d'être relié à la plateforme GitLab pour suivre les mises à jour des différents programmes qui seront réalisés. Nous avons pu vérifier plusieurs fois le très grand intérêt en termes de rapprochement, d'échanges, et de traçabilité de tels outils. Pour la rédaction d'articles, nous

utiliserons OverLeaf, un système de rédaction collaborative (en ligne) d'articles en Latex.

Les pourcentages de temps des collaborateurs de ce projet sont indiqués dans la fiche « personnels ». Le temps (50 %, c'est-à-dire 100 % du temps de recherche d'un MCF) du porteur de projet N. Glade, est du au fait que 100 % de son temps de recherche effectif correspond à ce sujet avec d'une part des recherches sur les réseaux d'automate Booléens à seuil, d'autre part des développement logiciels entrant dans le cadre de ce projet. L. Trilling est à 100 % car Pr. émérite. A. Stéphanou est à 30 % car associée à d'autres projets. De même pour E. Fanchon (35%). Les autres protagonistes sont à 20 %.

Outre le financement de la thèse et les coûts de fonctionnement des membres de ce projet, intra IDEX, nous demandons pour 8000 euros de matériel. Cela couvrira une évolution que nous souhaitons faire dans notre serveur de calcul : le munir d'une capacité en RAM très importante (256 Go de RAM EEC, pour un coût CNRS d'environ 3000 à 4000 euros) et lui adjoindre une carte dédiée au calcul GPU de type Nvidia Quadro P6000 (coût CNRS d'environ 3000 à 4000 euros). Une quantité très importante de RAM est en effet essentielle pour l'inférence de modèles utilisant ASP (coût mémoriel très important du grounder clingo et du solver SAT-like). L'achat d'un gros GPU est justifiée par le souhait de systématiquement utiliser OpenCL pour réaliser les calculs impliqués dans le modèle hybride que nous allons réaliser. C'est une technologie dont nous sommes coutumiers.

Sujet de thèse.

Cette thèse qui sera réalisée au laboratoire TIMC-IMAG, a pour objectif de tester, par l'intermédiaire de la modélisation, la théorie atavistique du cancer. L'émergence et le développement d'un cancer peut s'expliquer par la régression des phénotypes cellulaires vers des types primitifs, cela impliquant une désactivation et une altération progressive, au niveau génomique et épigénétique, des systèmes biologiques récemment acquis dans l'évolution, au profit de systèmes biologiques ancestraux donnant aux cellules un plus grande faculté à survivre aux environnements tissulaires altérés et instables. En cela, la théorie atavistique du cancer donne des clefs pour comprendre, mais également proposer des thérapies plus efficaces pour lutter contre la progression de tumeurs.

Le doctorant développera un modèle hybride de tissus cellulaire couplant des modèles de régulation génétique, fondés sur le formalisme des réseaux d'automates Booléens à seuil, avec un modèle de biomécanique tissulaire. En termes de réalisation, trois aboutissements devront être atteints :

- (i) la détermination, par des techniques d'inférence logique, de collections de réseaux d'automates Booléens à seuil satisfaisant les nécessités biologiques, comportementales et structurelles, organisées en couches interdépendantes, pour la régulation des processus intervenant dans l'homéostasie biomécanique des tissus. Les réseaux ainsi déterminés seront réfléchis dans le cadre conceptuel de l'évolution des systèmes de régulation, des processus de régulation épigénétiques et de la théorie de la viabilité.
- (ii) la modélisation par des techniques d'éléments finis notamment, d'un tissu cellulaire de type épithélial aux propriétés biomécanique faisant intervenir lesdits processus ;
- (iii) le couplage des processus biomécaniques avec les systèmes de régulation génétique déterminés préalablement.

Le doctorant devra montrer un intérêt fort pour l'interdisciplinarité pour être en mesure de mener à bien cette thèse. Les compétences requises sont principalement informatiques. Un niveau suffisant en mathématiques appliquées sera également nécessaire. Le contexte de sa thèse qui se déroulera en étroite collaboration entre modélisateurs et biologistes, lui permettra d'acquérir progressivement les compétences additionnelles.

Références bibliographiques du projet.

- [1] Aubin J.-P., Bayen A.M. Saint-Pierre P., Viability Theory, Springer, 2011
- [2] Ben Amor H., Méthodes numériques et formelles pour l'ingénierie des réseaux biologiques : Traitement de l'information par des populations d'oscillateurs, Approches par contraintes et Taxonomie des réseaux biologiques , Ph.D. Thesis, 2012, dir. J. Demongeot & N. Glade
- [3] Ben Amor H., Corblin F., Fanchon E., Elena A., Trilling L., Demongeot J., Glade N., Formal Methods for Hopfield-like networks, Acta Biotheor., 61 (2013) 21-39
- [4] Caraguel F., Lesart A.-C., Estève F., van der Sanden B., Stéphanou A., Towards the Design of a Patient-Specific Virtual Tumour, Comp. Math. Methods Med. 9 (2016) 1-12
- [5] Christen U., Ivanov S., Segretain R., Trilling L., Glade N., On computing structural and behavioural complexities of threshold Boolean networks,

- (2018) DOI: 10.13140/RG.2.2.15670.96327, Submitted to Acta Biotheor. (dec. 2018)
- [6] Christen U. Ivanov S., Segretain R. Glade N., Trilling L., Searching for Viability Graphs of Biological Functions by Computing Distances between Thresholded Boolean Automaton Networks, (2018) DOI: 10.13140/RG.2.2.18606.97603
- [7] Corblin F., Fanchon E., Trilling L., Chaouiya C., Thieffry D., Automatic Inference of Regulatory and Dynamical Properties from Incomplete Gene Interaction and Expression Data, In: Lones M.A., Smith S.L., Teichmann S., Naef F., Walker J.A., Trefzer M.A. (eds) Information Processing in Cells and Tissues. IPCAT 2012. Lecture Notes in Computer Science, vol 7223. Springer, Berlin, Heidelberg
- [8] Bussey K.J., Cisneros L.H., Lineweaver C.H., Davies P.C.W., Ancestral gene regulatory networks drive cancer, Proc. Natl. Acad. Sci. USA, 114 (2017) 6160–6162.
- [9] Elena A., Robustesse des réseaux d'automates booléens à seuil aux modes d'itération. Application à la modélisation des réseaux de régulation génétique, 2009, dir. J. Demongeot
- [10] Elena A., Ben-Amor H., Glade N., Demongeot J., Motifs in Regulatory Networks and their Structural Robustness, IEEE Transactions on Information Technology in Biomedicine, 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. BIBE 2008., Athens : Grèce (2008)
- [11] Gebser M., Kaminski R., Kaufmann B., Ostrowski M., Schaub T., Schneide, M., Potassco: The Potsdam Answer Set Solving Collection. A.I. Comm. 24 (2011) 107-124
- [12] Glade N., Elena A., Corblin F., Fanchon E., Demongeot J., Ben-Amor H., Determination, optimization and taxonomy of regulatory networks - The example of Arabidopsis thaliana flower morphogenesis, IEEE AINA' 11 & BLSMC' 11, Singapore, IEEE Proceedings, Piscataway (2011)
- [13] Guillot C., Lecuit T., Mechanics of epithelial tissue homeostasis and morphogenesis, Science 340 (2013) 1185-1189
- [14] Hopfield J.J., Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA 79 (1982) 2554-2558
- [15] Israel L., Tumour progression : random mutations or an integrated survival response to cellular stress conserved from unicellular organisms ?, J. Theor. Biol. 178 (1996) 375-380
- [16] Jeannin-Girardon A., Ballet P., Rodin V., A Software Architecture for Multi-Cellular System Simulations on Graphics Processing Units, Acta Biotheor. 61 (2013) 317-327
- [17] Jeannin-Girardon A., Ballet P., Rodin V., Large Scale Tissue Morphogenesis Simulation on Heterogenous Systems Based on a Flexible Biomechanical Cell Model, IEEE/ACM Trans. Comp. Biol. Bioinfo. 12 (2015) 1021-1033
- [18] Lineweaver C.H., Davies P.C.W., Vincent M.D., Targeting cancer's weaknesses (not its strengths): Therapeutic strategies suggested by the atavistic model, BioEssays 36 (2014) 827–835
- [19] Lobos C., A set of mixed-elements patterns for domain boundary approximation in hexahedral meshes, Studies Health Techno. Inf. 184 (2013) 268-272
- [20] Long Y., Cheddadi I., Mirabet V., Dumond M., Godin C., Boudaoud A., Cellular heterogeneity in pressure and growth emerges from tissue topology and geometry, bioRxiv, in review in Nature Physics, page 334664, 2018
- [21] McCulloch W.S., Pitts W., A Logical Calculus of the Ideas Immanent in Nervous Activity, Bull. Math. Biophys. 5(1943) 115-133
- [22] Mobilia N., Rocca A., Chorlton S., Fanchon E. Trilling L., Logical Modeling and Analysis of Regulatory Genetic Networks in a Non Monotonic Framework, In: Ortuño F., Rojas I. (eds) Bioinformatics and Biomedical Engineering. IWBBIO 2015. Lecture Notes in Computer Science, vol 9043. Springer, Cham
- [23] Ravasio A., Cheddadi I., Chen T., Pereira T., Ong H.T., Bertocchi C., Bruges A., Jacinto A., Kabla A.J., Toyama Y., Trepas X., Gov N., Neves de Almeida L., Ladoux B., Gap geometry dictates epithelial closure efficiency, Nat. Comm. 6, 7683 (2015)
- [24] Stephanou A., Fanchon E., Innominato P.F., Ballesta A., Systems Biology, Systems Medicine, Systems Pharmacology: The What and The Why, Acta Biotheor. (2018) doi: 10.1007/s10441-018-9330-2.
- [25] Stephanou A., Le Floc'h S., Chauviere A., A Hybrid Model to Test the Importance of Mechanical Cues Driving Cell Migration in Angiogenesis, Math. Mod. Nat. Phenom. 7 (2015) 32-55
- [26] Stephanou A., Lesart A.-C., Deverchère J., Juhem A., Popov A., Estève F., How tumour-induced vascular changes alter angiogenesis: Insights from a computational model, J. Theor. Biol, 419 (2017) 211-226
- [27] Stephanou A., Volpert V., Hybrid Modelling in Biology: A Classification Review, Math. Mod. Nat. Phenom. 11 (2016) 37-48
- [28] Tran V., McCall M.N., McMurray H.R., Almudevar A., On the underlying assumptions of threshold Boolean networks as a model for genetic regulatory network behavior, Frontiers in Genetics, 4 (2013) 1-10
- [29] Vincent M., Cancer : a de-repression of a default survival program common to all cells, Bioessay 34 (2011) 72-82
- [30] Vuong Q.T., Chauvin R., Ivanov S., Glade N., Trilling L., A logical constraint-based approach to infer and explore diversity and composition in thresholded Boolean automaton networks, Studies in Computational Intelligence Series, Proceedings of the Complex Networks 2017 conference (2017)
- [31] Zanudo J.G.T., Aldana M., Martinez-Mekler G. (2011) Boolean Threshold Networks: Virtues and Limitations for Biological Modeling. In: Niiranen S., Ribeiro A. (eds) Information Processing and Biological Systems. Intelligent Systems Reference Library, vol 11. Springer, Berlin, Heidelberg

1.3. Effet des champs magnétiques intenses sur le vivant – propriétés physico-chimiques de l'eau

DESCRIPTION DU PROJET PEPS 2016 – Intense magnetic fields effects on the physico-chemical properties of water. Implications in biology. (*Effet des champs magnétiques intenses sur les propriétés de l'eau. Conséquences en biologie*)

Scientifique responsable et Co-I. PI : Eric Beaugnon (LNCMI), co-I : Nicolas Glade (TIMC-IMAG)

Laboratoires et équipes impliqués. 1) Laboratoire LNCMI-G, CNRS UPR3228 , 25 rue des martyrs, 38042, Grenoble ; 2) Equipe DyCTiM2, laboratoire TIMC-IMAG, CNRS UMR5525 / Université Grenoble Alpes, 38706 La Tronche

Scientific context – scientific and technological objectives.

Water, more than being the liquid medium where biological particles move, is a polar solvent that is chemically implied in most biological processes notably by the way of the hydrogen bonds [Martin, 1999] water molecules establish between each others or with other molecules (amino acids, polar heads of lipids, ...), or during hydrolysis depending processes (*e.g.* in hydrolase enzymes). **The functioning of biological systems strongly depends on its physico-chemical properties.**

The different manners water plays a role in biological systems are however not all known nor well understood, and put aside most of the time, when trying to understand biological processes. Although water molecules mainly exist within clusters of water [Shelton, 2014], dynamical structures of various sizes that permanently form due to hydrogen bonding and dissolve due to thermal agitation, the importance of its structuration is rarely taken into account in the study of biological processes including biochemical reactions, first because the lifetime of water clusters is not comparable to the observation time of the biological processes, and second because it is hard (difficult to implement and huge computational cost) to take explicitly this solvent into account in the associated numerical models. Yet such clusters are observed even by neutron scattering or infrared absorbance [Pang, 2012] and may play an important role in biological dynamics : the disponibility of water molecules for solvation or biochemical reactions is limited by the strength of these structures.

Because of the partial covalence of hydrogen bonds, the electrons of water are largely delocalized within large coherent regions [Del Giudice, 2010 & 2013], the water clusters that are stable structures [Maheshwary, 2001] are shown to be able to interact with electric fields [Marín & Lohse, 2010]. **Intense magnetic fields** have also been shown to **change durably water properties**. Macroscopic properties of water are indeed modified after water sample were exposed to and removed from magnetic fields in the range of 0.1 T to up to 20 T [Otsuka & Ozeki, 2006; Pang, 2012; Chen & Beaugnon, 2015 (preliminary study)]. What is nothing less than surprising is the **long lifetime of this effect** : after exposure to a magnetic field of 6T during 150 minutes and removal of the samples from the magnetic field, the effect remains for durations in the **range of several dozen minutes to up to one hour** until the water recovers progressively its normal properties. Surface energy increases (*e.g.* the decrease of the contact angle of water droplets on metallic surfaces is in the range 8°-15° after 150 minutes of exposure to magnetic fields of 6 to 10 T), infrared absorption increases, and the oxydative power increases [Otsuka & Ozeki, 2006; Pang, 2012; Chen & Beaugnon, 2015 (preliminary study)]. At the molecular level, this change of physico-chemical properties of water seems to be due to an increase of the size of water clusters [Pang & Deng, 2008] because of an increase of the strength of hydrogen bonds [Cai, 2009]. In addition, our preliminary results [Chen & Beaugnon, 2015 (preliminary study)] showed that increasing the ionic charge in water samples also increases the effect of a magnetic field (8 T) on water interfacial energy ; Increasing amounts of dissolved oxygen is shown to increase this property [Otsuka & Ozeki, 2006]. In the absence of dissolved oxygen and in pure water, the duration of the effect is very limited. These observations might confirm the reinforcement of clathrate hydrates (water clusters organized around charged molecules) formation under the action of magnetic fields [Fesenko, 2002] and the subsequent amplification and stabilization of water clusters. Finally, we showed recently [Chen & Beaugnon, 2015 (preliminary study)] that the intensity and lifetime of the effect of magnetic fields on water properties is increased as the intensity of the magnetic field increases (we measured interfacial energy of water exposed to magnetic fields in the range 2 to 20 T). Other related macroscopic properties may also change in these conditions like the pH and electro-chemical properties [Pang, 2012]. The manner magnetic fields cause these changes is not known but could be due to the reorganisation of water clusters under their action.

Magnetic fields were showed to have strong effects on living matter. They exert notably a strong torque on magnetically anisotropic biological structures, *i.e.* elongated structures like the fibres of the cytoskeleton, and probably elongated membrane structures like mitochondria. By aligning such fibres, they change the global organisation of the cytoskeleton and its dynamics [Vassilev, 1982; Glade & Tabony, 2005; Glade 2006].

Magnetic fields also cause different effects on living cells like in prokaryotes where a strong reduction (70%) of their survival is observed in parallel to an increase of intracellular reactive oxygen species ROS, but how magnetic fields interact is still not well explained [Boda, 2015]. The increase of ROS could be associated to a malfunction in a enzymatic process due to a conformational change of enzymes in a solvent with different physico-chemical properties or a decrease of their activity. Intense magnetic fields are however frequently used in medicine and biology at similar values of intensity and exposure time. In medical MRI, subjects are exposed to fields of 1 to 4 T during at least 30 minutes. Biological experiments of metabolic NMR that use fields of 1 to 10 T can also go on for hours or days [Gout, 2014]. **One can then ask the question of the effect of intense magnetic fields mediated by water in living cells or organisms.**

Our **working hypothesis** is the following : **biological processes** from single biochemical interactions like protein adsorption on a substrate, or biochemical reactions such as the functioning of an hydrolase enzyme, to cellular processes such as cell adhesion or cytoskeleton assembly **may display different behaviors or kinetics when done in water exposed to magnetic fields compared to those occurring in normal water.**

The **long-term objective** of our project is to understand **how water exposed to intense magnetic fields causes changes in biological processes.** By changing temporarily the physico-chemical properties of saline buffers by exposure to intense magnetic fields, we aim to show evidences of changes in the functioning of different biological systems dissolved in such buffers.

The **main objective** of the **present project** is the following : the **understanding of the physico-chemical changes of water by physical experiments will provide clues to understand how normal water as well as water exposed to intense magnetic fields interact with biological systems.** The PhD thesis we will recruit will focus mainly on the physical part of the project.

Project organization and implemented means.

The project is composed of two main connected workpackages described below :

- A physical study of the properties of water and saline buffers exposed to intense magnetic fields.
- A biological study of the effect of saline buffers exposed to magnetic fields on 4 biological systems.

These two workpackages will communicate: the physical study will give clues for understanding the effects observed in the biological study, and in return the manner biological systems will respond to the exposed buffers will suggest new physical measurements or to adapt their experimental conditions to correspond to conditions encountered in biology (concentration of salts, temperature, oxygen level ...).

WPI) Study of the effect of intense magnetic fields on the physico-chemical properties of water and saline buffers. This workpackage implies the following partners: LNCMI and TIMC-IMAG. A Ph.D Thesis will be recruited (ANR funds or fellowship from the physical doctoral college) and shared between LNCMI and TIMC-IMAG laboratories. The Ph.D student, in addition to the physical studies will communicate his results to and discuss their implications with the biologists. He will also be in charge of managing mainly the physical experiments at the LNCMI, but also help the biologists to realize their experiments in the latter case.

The objective of this WP is both to investigate the atomic or molecular nature of the changes in water and saline (ionic) buffers exposed to magnetic fields, and to quantify the macroscopic changes of some of their physico-chemical properties, in particular those that are expected to play a role in the biological systems that will be studied. In particular, we raise the question of the influence in biology of both water clusters and hydrogen bonds whose strength seem to be reinforced after exposure to intense magnetic fields. We will then focus on quantifying them by the way of the techniques described below in different experimental conditions, particularly the ionic charge of the medium, the temperature, the intensity of the magnetic field (in the range 2-30 T) and the time exposure to the magnetic field. We will follow the evolution over time of the physico-chemical properties of water in the magnet and outside the magnet after exposure to an intense magnetic field. **Both the knowledge on the influence of these parameters and these kinetics are new.**

Measuring the **contact angle** of water or buffer droplets (sessile drop method) on different surfaces of known and different polar and dispersive **surface energies** (we will use glass, teflon, gold and copper) is the simplest

way to obtain information on the molecular forces implied and that are changed under the action of magnetic fields. Via the Owens & Wendt or the Fowkes formulae, one can indeed determine the detailed (*e.g.* dispersive and polar) parts of interfacial energy of the measured compounds. Their whole interfacial energy can be confirmed by direct measurements by pendant drop or drop volume tensiometry.

Electrochemistry will provide more details on the physico-chemical properties of the bulk water-based solutions: the conductivity informs on proton and ion mobility, so on the structure of water in the solution, while amperometry informs on the redox potential of the solution. Finally, we will realize **infrared spectroscopy** kinetic measurements on samples of exposed water-based solutions removed from the magnet, just as well as samples still in the magnet. The latter experiments will measure the kinetics of the strengthening and growth of water clusters as they are exposed to an intense magnetic field (as suggested by [Pang, 2012]) while the former will measure the relaxation of water structures and subsequent physico-chemical properties after they were exposed to the magnetic field.

WP2) Study of water exposed to intense magnetic fields on biological processes. This workpackage concerns mainly the following partners : TIMC-IMAG and LPCV. A demand has been made to the ANR to fund a 2-years post-doctoral contract.

In this WP we will **measure the kinetics and observe the behaviour of 4 different biological systems** dissolved in saline water-based buffers exposed to an intense magnetic field. These experiments will be realised outside of the magnet so that we will ensure that the observed effects come from the property changes of exposed water. The systems tested will represent several **biological systems at different scales and levels of complexity** : (i) hydration (solvation) and adsorption (on a solid surface) of single proteins followed by electro-chemistry [Zebda, 2011], (ii) kinetics of enzyme - substrate interactions (an enzyme that does not depend explicitly on the hydrolysis of water (a Glutathione-S-transferase), and a serine hydrolase (a alpha-mannosidase) where water is required to perform the enzymatic reaction, both used by partner 3 [Jacquinod, 2007]), (iii) microtubule assembly and self-organization that both depend on the properties of water [Chakrabarti, 1999] and for which we have a strong expertise [Glade 2005, 2006, 2012], and (iv) cell *primo* and long-term adhesion (observed by several techniques of microscopy for which TIMC-IMAG is equipped [Platform IBISA : Leray, 2012] and measured by radial flow detachment [Sordel, 2007]). As induced **secondary objectives**, this will allow us to **identify the biological processes sensitive to changes** in the physico-chemical properties of water, in particular when exposed to magnetic fields as it is the case in **Magnetic Resonance Imagery MRI** or in **Metabolic Magnetic Nuclear Resonance MMNR**. Moreover, it will allow us to envisage new ways to drive biological systems into different behaviors, *e.g.* to change temporarily the adsorption or the kinetics of biological compounds in biotechnological processes or in the context of drug delivery. Finally, we also aim to **take into account the physico-chemical changes of water** in the corresponding **models** of protein adsorption, enzyme kinetics, microtubule assembly and self-organisation and cell adhesion, **for which partners 1 and 2 have a large expertise**.

Analysis of the risks. This innovative project will yield numerous new results as well in physics as in biology. Our successful preliminary study on the physical part of this project [Beaugnon & Chen, 2015] secures the obtaining of such results in the WP1, *i.e.* it offers a student to realize his Ph.D thesis on a well-defined subject with a predictable research program and short-term results to count on. On the contrary, we will not engage the Ph.D student in the biological part of the project since it needs more experimental adjustments and contains more unknown aspects. Indeed, (i) it is still difficult to have an idea of how much the studied biological systems are sensitive to the changes of the physico-chemical properties of water (although there are elements of proofs in the literature to reinforce that point [Chakrabarti, 1999; Boda, 2015]), and (ii) the preparation and the realization of such biological experiments is difficult since the change of physico-chemical properties of water is limited in time (from about several dozen minutes to one hour). This work needs a strong expertise in biology and biophysical measurements, *i.e.* a very experienced postdoc. The Ph.D student will however work on the necessary adjustments of such experiments, notably by working on similar measurement devices with similar buffers.

Impact and Fallout from the Project. In this project, we ask the question of the indirect effect, mediated by the solvent, of magnetic fields on living matter. Our project will clearly provide new knowledge (articles and conferences), but should also have an impact in technological and biotechnological applications (new processes and patents).

Knowledge. The physical part (WP1) of the project will provide new knowledge on the changes of the properties of water due to magnetic fields, in particular the importance of the ionic charge in solution, the kinetics of both changes and relaxation of the physico-chemical properties of watery solutions exposed to magnetic fields, and the dependence of these changes and relaxation to the strength of the field. The physical measurements (IR spectroscopy, electrochemistry, surface energies) will also give clues to understand what happens at the molecular level. The biological part (WP2) will inform first on the possibility that the behaviour of biological systems (living organisms or biological samples) can be indirectly affected by the effect of magnetic fields on their water-based solvent. Second, we will be able to measure the differential sensitivity of such biological systems to the magnetically exposed water. Finally and more generally, this study will reinforce the knowledge on the role of water in biology, in particular the possible one of water clusters.

As a direct consequence to this biological study, this work may call into question some biological experimental studies in metabolic NMR (magnetically exposed water could induce a bias) and the innocuousness of medical MRI for which time exposures could be optimized.

Applications. Since the surface energy of magnetically treated water is higher and its hydrogen bonds stronger, one can expect that the hydration of surfaces and solutes including polar biological matter from proteins to cells will increase. This also means that one can expect less adsorption on solid surfaces and less cellular adhesion in such conditions. This allows us to have a glimmering of possible technological applications in microfluidics, e.g. for biochip applications [Vulto, 2006], in which the hydration power of magnetically treated water could be advantageous. The differential but temporary adsorption of biological compounds (proteins, cells...) in magnetically treated water compared to its normal state after relaxation could also be exploited in biotechnological processes or applications like biofuel cells [Zebda, 2011]. Finally, we think that, in a medical context, therapeutic exposure of patients to intense magnetic fields (i.e. in a MRI device) could be envisaged as a new way to enhance or control the effect of drugs delivery and activity. This could constitute a long term objective after this project will be completed.

Publication. All the partners in this project have a sustained publication activity in peer-review journals, often well impacted. In addition, this TIMC-IMAG team has a mentality directed to applications [Plissonnier, 2006; Zebda, 2013] ; we will file patents insofar as possible. The results will be published in journals of physics and in journals of biology.

Partners and Schedule.

Partners and Involvements. The main project implies 3 partners : 1. LNCMI, 2. TIMC-IMAG and 3. LPCV-CEA. The LNCMI will be in charge of the magnetic fields, and will manage the physical experiments (IR/UV spectroscopy, surface energy measurements, electrochemistry) in collaboration with the TIMC-IMAG. All biological experiments except microscopy observations of long term cell adhesion will be performed by the TIMC-IMAG and LPCV (2-years postdoc is required in another demand) at the LNCMI laboratory. The biological experiments (and the post-doctoral contract) will only start on the 6th month, after several developments and adjustments on the experimental benches and processes will have been made.

Partner 1. LNCMI (National Laboratory of Intense Magnetic Fields - CNRS) includes Eric BEAUGNON (Pr. CNU28 - Magnetic fields & Physics, 75%), Ivan BRESLAVETZ (IE CNRS - Optical measurements, 10%), Ghenadi NOVITCHI (IR CNRS - Chemistry, 10%).

The CNRS facility LNCMI holds physical or biological experiments in magnetic fields. This team has in

addition experienced biological studies and the physico-chemical characterization of liquids [Beaugnon & Tournier, 1991; Braithwaite, 1991; Beaugnon, 2001; Glade, 2006; Rivoirard, 2009, Novitchi, 2012].

Partner 2. TIMC-IMAG / IBISA (Techniques for biomedical engineering and complexity - informatics, mathematics and application - Grenoble. CNRS / University Joseph Fourier /) includes Nicolas GLADE (MC CNU27 - Physics, Biology and Modelling, 75%), Yves USSON (CR CNRS / IBISA facility - Microscopy, 10%), Abdelkader ZEBDA (CR CNRS - Electrochemistry & Biology, 10%).

TIMC-IMAG is an interdisciplinary laboratory. This team combines expertises [Glade & Tabony, 2005; Glade, 2006; Plissonnier, 2006; Sordel, 2007; Zebda, 2011 + patent; Glade, 2012] as well in biology (cell adhesion, cytoskeleton dynamics, microscopy), biophysics (Physico-chemical characterization of surfaces and solutions, protein adsorption kinetics and electrochemistry, MRI studies) and modelling (multiscale and biophysical modelling of complex systems). In addition, it includes a recognized microscopy facility called IBISA for which the team is very qualified [Leray, 2012].

Partner 3. LPCV-CEA (Laboratory of Vegetal Cell Physiology - French Alternative Energies and Atomic Energy Commission) includes Olivier BASTIEN (CR INRA - Enzyme kinetic models, 10%) and Florent VILLIERS (PostDoc - Biology, 30%).

This team of LPCV has a strong expertise on biological kinetic measurements and modelling [Villiers & Kwak, 2012; Marechal & Bastien, 2014].

Schedule. The present demand is scheduled for a 18 months project as required. It does not include the experiments on microtubules (that are planned for the last year of our 3-years main project), and only the preliminary experiments on cell adhesion (adjustments).

Months	Partner	1-6	6-12	12-18
WP1	LNCMI & TIMC	Combined physical measurements other 18 months : Surface energy measurements / IR spectra / electrochemical measures. Variable conditions : (months 1-6) composition of the solutions, (6-18) intensity of the field and exposure times (including also measures of relaxation dynamics)		
WP2	TIMC	Adjustments (buffers, preliminary biological experiments outside the magnet ...)	Measures of solvation and adsorption by electrochemistry.	Adjustments for cell adhesion experiments.
	LPCV		Enzymatic experiments.	

Projected expenses. We ask for a total of **60 k€** distributed as follows : **5k€** for 18 months of Helium for the supraconducting magnet + **35k€** of equipments for the LNCMI (for a tensiometer (25 k€) and for an infrared extension for our JASCO V700 spectrometer (10k€)) + **5k€** for two Master 2 internships + **15k€** of functioning (5k€ for each partner).

References.

- [Beaugnon & Tournier, 1991] E Beaugnon, R Tournier, Levitation of organic materials, *Nature* 349 (1991) 6309
 [Beaugnon, 2001] E Beaugnon, D Fabregue, D Billy, J Nappa, R Tournier, Dynamics of magnetically levitated droplets, *Physica B: Cond.Matter* 294 (2001) 715
 [Boda, 2015] SK Boda, K Ravikumar, DK Saini, B Basu, Differential viability response of prokaryotes and eukaryotes to high strength pulsed magnetic stimuli, *Bioelectrochem.* 106 (2015) 276
 [Braithwaite, 1991] D Braithwaite, E Beaugnon, R Tournier, Magnetically controlled convection in a paramagnetic fluid, *Nature* 354 (1991) 134
 [Cai, 2009] R Cai, H Yang, J He, W Zhu, The effects of magnetic fields on water molecular hydrogen bonds, *J. Mol. Struct.* 938 (2009) 15
 [Chakrabarti, 1999] G Chakrabarti, S Kim, MLJ Gupta, JS Barton, RH Himes, Stabilization of tubulin by deuterium oxide, *Bioelectrochem.* 38 (1999) 3067
 [Chen & Beaugnon, 2015] B Chen, E Beaugnon, "Effet des champs magnétiques sur les propriétés de l'eau". (2015) *preliminary study* (master internship)

- [DelGiudice, 2010] E Del Giudice, EC Fuchs, G Vitiello, Collective molecular dynamics of a floating water bridge, *WATER* 2 (2010) 69
- [DelGiudice, 2013] E Del Giudice, A Tedeschi, G Vitiello, V Voeikov, Coherent structures in liquid water close to hydrophilic surfaces, *J. Phys.: Conf. Ser.* 442 (2013) 012028; G. Preparata, QED Coherence in Matter (London: World Scientific, 1995).
- [Fesenko, 2002] EE Fesenko, VI Popov, VV Novikov, SS Khutsyan, Structure formation in water by the action of weak magnetic fields and xenon. Electron microscopic analysis, *Biophysics* 47 (2002) 365
- [Glade, 2012] N Glade, On the nature and shape of tubulin trails: implications on microtubule self-organization, *Acta Biotheor.* 60 (2012) 55
- [Glade, 2006] N Glade, E Beaugnon, J Tabony, Ground-based methods reproduce space-flight experiments and show that weak vibrations trigger microtubule self-organisation, *Biophys. Chem.* 121 (2006) 1
- [Glade & Tabony, 2005] N Glade, J Tabony, Brief exposure to magnetic fields determine microtubule self-organisation by reaction-diffusion processes, *Biophys. Chem.* (2005) 115, 29-35
- [Gout, 2014] E Gout, F Rébeillé, R Douce, R Bligny, Interplay of Mg^{2+} , ADP, and ATP in the cytosol and mitochondria: Unravelling the role of Mg^{2+} in cell respiration. *Proc. Natl. Acad. Sci. USA* 111 (2014) E4560
- [Jacquinod, 2007] M Jacquinod, F Villiers, S Kieffer-Jacquinod, V Hugouvieux, C Bruley, J Garin, J Bourguignon, A proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture, *Mol. Cell. Proteomics* 6 (2007) 394
- [Leray, 2012] A Leray, C Spriet, D Trinel, Y Usso, L Hélot, Generalization of the polar representation in time domain fluorescence lifetime imaging microscopy for biological applications: practical implementation. *J. Microsc.* 248 (2012) 66
- [Maheshwary, 2001] S Maheshwary, N Patel, N Sathyamurthy, AD Kulkarni, SR Gadre, Structure and Stability of Water Clusters $(H_2O)_n$, $n = 8-20$: An Ab Initio Investigation. *J. Phys. Chem.* 105 (2001) 10525
- [Marechal & Bastien, 2014] E Maréchal, O Bastien, Modeling of regulatory loops controlling galactolipid biosynthesis in the inner envelope membrane of chloroplasts. *J. Theor. Biol.* 36 (2014) 1
- [Marín & Lohse, 2010] ÁG Marín, D Lohse, Building water bridges in air: Electrohydrodynamics of the floating water bridge, *Physics of fluids* 22 (2010) 122104.
- [Martin, 1999] TW Martin, ZS Derewenda, The name is bond - H bond, *Nature Structural Biology*, 6 (1999) 403
- [Novitchi, 2012] G Novitchi, G Pilet, L Ungur, VV Moshchalkov, W Wernsdorfer, LF Chibotaru, D Luneau, AK Powel, Heterometallic Cu^{II}/Dy^{III} 1D chiral polymers: chirogenesis and exchange coupling of toroidal moments in trinuclear Dy_3 single molecule magnets, *Chem. Sci.* 3 (2012) 1169
- [Otsuka & Ozeki, 2006] I Otsuka, S Ozeki, Does magnetic treatment of water change its properties ?, *J. Phys. Chem. B* 110 (2006) 1509
- [Pang & Deng, 2008] XF Pang, B Deng, Investigation of changes in properties of water under the action of a magnetic field, *Science in China Series G-Physics Mechanics Astron.* 51(2008) 1621
- [Pang, 2012] XF Pang, B Deng, B Tang, Influences of magnetic field on macroscopic properties of water, *Mod. Phys. Lett. B* 26 (2012) 1250069 1
- [Plissonnier, 2006] M Plissonnier, I Chartier, M Borella, N Glade, Method for localizing a chemical or biological species on a substrate, analyzing microsystem and biochip. Patent W0 2006/005887 A1
- [Rivoirard, 2009] S Rivoirard, T Garcin, E Beaugnon, F Gaucherand, High temperature dilatation measurements by in situ laser interferometry under high magnetic field, *Rev. Sci. Instr.* 80 (2009) 103901
- [Shelton, 2014] DP Shelton, Long-range orientation correlation in water, *J. Chem. Phys.* 141 (2014) 224506
- [Sordel, 2007] T Sordel, F Kermarec-Marcel, S Garnier-Raveaud, N Glade, F Sauter-Starace, C Pudda, M Borella, F Chatelain, F Bruckert, N Piccollet-D'hahan, Influence of glass and polymer coatings on CHO cell morphology and adhesion, *Biomaterials* 27 (2007) 1572
- [Villiers & Kwak, 2012] F Villiers, JM Kwak, Rapid apoplastic pH measurement in Arabidopsis leaves using a fluorescent dye, *Plant Signalling & Behavior* 8 (2012) e22587
- [Vulto, 2006] P Vulto, N Glade, L Altomare, J Babet, G Medoro, I Chartier, N Manaresi, M Tartagni, R Guerrieri, Microfluidic channel fabrication in dry film resist for production and prototyping of hybrid chips, *Lab on a Chip* 5 (2005) 158
- [Zebda, 2011] A Zebda, C Gondran, A Le Goff, M Holzinger, P Cinquin, S Cosnier, Mediatorless high-power glucose biofuel cells based on compressed carbon nanotube-enzyme electrodes, *Nature Comm.* 370 (2011)
- [Zebda, 2011] A. Zebda, A Le Goff, M Holzinger, S Cosnier, Direct-transfer biopile, patent PCT/FR2011/051931

2. Développement logiciel

2.1. Contexte, besoin et état de l'art

Dans un laboratoire comme TIMC, ainsi que certainement dans de nombreux autres laboratoires attachés à la modélisation en biologie ou physique, ou au traitement des données, de nombreux programmes, algorithmes, bibliothèques logicielles sont développés.

Souvent, ceux qui contribuent, voire initient ces programmes sont les étudiants avec leurs divers niveaux d'exigences, de temps disponible pour la finalisation, les tests, la documentation. Ces programmes sont rarement aboutis, d'autant que les stages de recherche comme les thèses n'ont pas pour finalité le développement de logiciel destinés à être valorisés autrement que pour faire ce qu'on leur demande, à savoir des tâches précises: des simulations ou du traitement des données dans le cadre plus ou moins strict du sujet de recherche donné. Leur réutilisation en tant que telle en est d'autant plus compliquée. Ce qui est demandé, c'est que ces programmes remplissent leurs objectifs

qui ont pour finalité des résultats de recherche. Cependant, le temps de développement passé, qui relève de l'ingénierie, pour beaucoup d'un savoir faire en programmation, pourrait être mieux valorisé à défaut de se faire quasiment à perte.

La question de la perte de l'investissement homme-temps en génie logiciel est plus large que cela : de nombreuses parties des programmes développés sont ré-écrites à chaque nouveau projet, là où il pourrait y avoir capitalisation et factorisation, d'autant que ces briques logicielles sont souvent longues et pénibles à développer ; elles sont souvent indispensables bien que non centrales. Par exemple, les systèmes de visualisation graphique et les interfaces utilisateur pour manipuler de façon interactive certaines simulations sont souvent longues et complexes à développer. L'utilisateur du programme (le chercheur ou l'étudiant) peut en avoir besoin pour visualiser et contrôler l'évolution de la simulation en direct, en particulier pendant les phases de développement d'un modèle ; c'est donc une fonctionnalité nécessaire, qui ne constitue pourtant pas le cœur du projet qui est la simulation numérique elle-même. C'est un exemple classique mais récurrent. Un autre exemple, plus 'bas niveau' est celui des structures de données et des algorithmes fréquemment utilisés : maillages, conteneurs 2 ou 3D comme les grilles, solveurs, moteurs de diffusion ... sont des éléments presque toujours retrouvés dans les simulations en biologie.

Il existe certes des bibliothèques logicielles assez efficaces proposant ces parties. Mais c'est alors leur articulation qui pose problème. C'est d'ailleurs presque le point le plus critique : ces briques logicielles, souvent complexes et rarement pensées pour fonctionner directement avec d'autres, font souvent l'objet, par les modélisateurs, de mauvaises intégrations dans les programmes, les rendant peu performants et peu extensibles. Sans compter que s'atteler à une telle intégration prend du temps et, point important, impose au modélisateur dont l'objectif, rappelons-le, est de simuler un système biologique et non d'écrire un programme élégant, optimal et ergonomique, de se détacher de son sujet d'intérêt pour se concentrer sur des questions d'ingénierie logicielle. Ces questions peuvent s'avérer très complexes dans le cas de modèles hybrides, c'est-à-dire mixant plusieurs formalismes (donc plusieurs types de structures de données et algorithmes) de telle façon que, souvent, le modélisateur en restant focalisé sur l'implémentation du modèle en oublie toute réflexion sur la structure globale de son programme et sur son optimisation, au détriment de la modularité, de son extensibilité (ex : peut-on réutiliser un programme simulant une tumeur et intégrer facilement des réseaux de régulation génétique dans les cellules simulées ? Peut-on y ajouter une biomécanique tissulaire *a posteriori* ? etc. C'est la plupart du temps impossible car cela n'a pas été pensé comme une potentialité dès le départ) et de sa performance.

Ajoutons pour finir que dans les laboratoires pratiquant la transdisciplinarité comme TIMC, les compétences premières des modélisateurs ne font pas d'eux des experts du développement et du génie logiciel. Ils n'ont la plupart du temps pas le savoir-faire nécessaire en terme d'ingénierie logicielle pour mener de front, seuls ou même accompagnés d'un étudiant, l'implémentation numérique d'un modèle formel et l'ingénierie logicielle dans laquelle elle sera intégrée.

Les besoins en modélisation augmentant dans le domaine de la biologie, il y a depuis quelques années un challenge qui se fait pressant, de proposer aux modélisateurs des outils logiciels pour

faciliter l'implémentation rapide et performante de leurs modèles numériques. Parmi eux, des bibliothèques spécialisées, *bas niveau*, comme des solveurs d'équations différentielles, des outils mathématiques d'algèbre linéaire, des générateurs de maillages, fort utiles, mais pas *clefs en main*, c'est-à-dire des solutions intégrées permettant de modéliser rapidement des objets de *haut niveau* comme des cellules, des tissus, intégrant biomécanique, réseaux de régulation, diffusion, *etc* ; mais aussi des solutions intégrées (dont je qualifie les composantes de *haut niveau* au sens où elles leur niveau d'abstraction est faible tant elles s'approchent des objets biologiques d'intérêt) comme *SimCells*, *HSIM*, *Aevol* (dédié à la simulation de dynamiques darwiniennes sur des systèmes génétiques artificielles ; je ne le décrirai pas) [Logiciels :AEVOL] ou *PhysiCell* et *PhysiBOSS*. Il existe d'autres solutions de ce type, mais je ne cite ici que celles qui me semblent les plus performantes et abouties.

Ces solutions intégrées sont très performantes (implémentation multi-cœur CPU, calcul GPU⁶⁷, algorithmes dédiés) et ergonomiques (ce qui est un point très important dont je discuterai plus loin), mais souffrent inévitablement d'une limitation d'expressivité car elles imposent fortement une modélisation (et non un développement) dans le cadre prévu par la solution. Dans ce cadre, systèmes hybrides alliant représentations cellulaires et tissulaires, intra-cellulaires, systémiques (régulation génétique et métabolique), biomécaniques, alliant systèmes multi-agents, calculs sur grille ou éléments finis, réseaux Booléens et multi-valués, sont difficiles à implémenter.

De plus, leur extensibilité est particulièrement difficile, les codes sources de ces solutions n'étant pas toujours ouverts et restant en tout cas complexes à comprendre pour ceux qui chercheraient à en augmenter les capacités. Dès lors, leur éventuelle extension est laissée à la discrétion du concepteur de la solution.

Concernant l'expressivité et l'extensibilité de ces solutions, prenons d'abord l'exemple de *SimCells* [Logiciels :SimCells ; Ballet, 2018 ; Douillet, 2020] qui offre un solution d'implémentation de modèles et leur simulation sous une forme multi-agents sur grille pour la simulation de tissus cellulaires 3D. Ce logiciel permet de simuler des ensembles de cellules (des meta-agents composés d'agents plus petits que l'utilisateur ne contrôle pas et qui permettent de simuler leurs membranes et leur centre actif, pouvant représenter le noyau cellulaire), dotées de pseudo-membranes, dont on peut contrôler certaines propriétés de rigidité, d'âge, de sécrétion ou ab/adsorption moléculaire, dont la biomécanique (rigidité cellulaire, contacts inter-cellulaires, motilité) est phénoménologique. *SimCells* offre pour *programmer* le comportement des composants disponibles, des plus petits (des molécules) aux plus gros (des cellules), un fort astucieux langage de combinaison graphique des propriétés et actions semblable au *Blueprint Visual Scripting* de l'*Unreal Engine* [Logiciels :BVS] : les propriétés et autres paramètres issus des agents ou du milieu sont combinés par des fonctions logiques (et, ou ...) dans des schémas conditionnels (if – then) et des boucles d'actions, pour créer des comportements.

67 CPU : *Central Processing Unit*, autrement dit le processeur classique d'un ordinateur. GPU : *Graphical Processing Unit*, c'est-à-dire le processeur de la carte graphique. Ces processeurs fonctionnent très différemment des CPU ; en particulier ils comportent de très nombreuses unités de calcul (plusieurs centaines à milliers) moins puissantes qu'un CPU, mises en parallèle.

Ce système est impressionnant d'efficacité : le modélisateur le prend en main en quelques minutes et peut très rapidement modéliser un système biologique s'apparentant à un tissu et pouvant générer des comportements phénoménologiques émergents et complexes. Son exécution est performante car fondée sur une implémentation exploitant parfaitement la puissance des unités de calcul massivement parallèles des GPUs, en particulier dans la représentation des composants de la simulation (les agents) et dans les algorithmes, notamment celui du calcul des membranes cellulaires [Douillet, 2020].

SimCells ne permet cependant pas d'intégrer ni biomécanique réaliste tant au niveau des membranes que du milieu extracellulaire (notamment la matrice extra-cellulaire) ou intra-cellulaire (représentation du cytosquelette et son interaction avec la membrane), ni simuler le milieu intra-cellulaire, en particulier des réseaux de génétiques et métaboliques, et encore moins des compartiments intra-cellulaires comme le noyau et ses prolongements vers le réticulum endoplasmique ou vers l'appareil de Golgi (dont on sait maintenant qu'ils sont contigus), ou les mitochondries, avec leurs dynamiques internes ... sauf à modifier profondément le code du logiciel. Il n'est pas prévu pour tout cela ! *SimCells* ne dispose pas non plus de la possibilité d'encoder de façon exacte des dynamiques réactionnelles, biochimiques par exemple. Le modèle sous-jacent est fondé sur une description probabiliste phénoménologique des événements intervenant dans le système biologique, non pas sur la transcription probabiliste des lois d'action de masse décrivant biochimiquement ou biophysiquement l'évolution des processus biologiques. Ce dernier point est critique pour celui qui souhaite aller plus loin, vers des modèles davantage quantitatifs, fondés sur des modèles de dynamiques chimiques et biomécaniques éprouvés.

Étant en contact étroit avec Pascal Ballet, le concepteur de *SimCells*, l'idée de procéder ainsi pour augmenter les capacités de *SimCells* est envisageable, mais cela reste complexe, le code source étant en *OpenCL*⁶⁸ et pas particulièrement pensé pour une intégration à la volée de nouveaux modules. De tels rajouts ne seraient par ailleurs que limités en nombre et faits au cas par cas. Le modèle d'extensibilité de *SimCells* ne me paraît pas facilement viable. Il faut, je le pense (et Pascal Ballet le pense aussi), rester avec ce logiciel sur son esprit initial qui est celui de la solution de modélisation semi-phénoménologique de systèmes biologiques cellulaires, ce qu'il fait très bien, et ne pas chercher à aller au-delà ; ce n'est pas son objectif premier.

HSIM [Logiciels :HSIM], développé par Patrick Amar est un simulateur dédié à la modélisation des réactions biochimiques à l'intérieur d'une cellule. Il implémente une solution multi-agent permettant de représenter les molécules en interaction dans la cellule sous forme d'agents. Extrêmement rapide, le système de calcul sous-jacent permet l'actuation des équations biochimiques via une implémentation de l'algorithme de Gillespie en *tau leaping* (au lieu de fonctionner en temps continu comme l'algorithme de Gillespie exact qui évalue l'instant de la prochaine réaction et son type à partir de la transcription probabiliste de l'équation maître intégrant l'ensemble des réactions et leurs paramètres cinétiques, le *tau leaping* détermine dans des laps de temps discret le nombre d'événements de chaque type réactionnel se produisant ; cet algorithme oblige à prendre en compte

68 Un langage semble à Objective C (une sorte de langage C) mais dédié au calcul sur GPU.

des contraintes d'intégrité sur la loi globale d'action de masse, à savoir les problèmes d'interdépendances de certaines réactions comme la conjonction de la transformation d'une substance et de sa diffusion qui peuvent amener à la génération de quantités négatives) [Gillespie, 1976 & 1977]. La modélisation sous *HSIM* se résume, tout aussi efficacement, ergonomiquement devrait-on dire, à écrire de façon scriptée les réactions (transformations) existant dans le milieu et leurs paramètres cinétiques dans un fichier.

Le cadre de modélisation de *HSIM* est également strict. Il n'est pas envisageable de modéliser autre chose que les réactions intra-cellulaires. Il n'est pas possible de modéliser des compartiments internes en interaction avec le milieu cytosolique, ni de donner un comportement biomécanique aux assemblages supra-moléculaire (ex : actine, microtubules) qu'il est possible de programmer de façon simplifiée. Il n'est pas envisageable non plus de passer à l'échelle supra-cellulaire (tissus cellulaires).

PhysiCell [Logiciels:PhysiCell] et son dérivé⁶⁹ *PhysiBOSS* [Logiciels :PhysiBOSS] semblent être les solutions de modélisation de ce type les plus ouvertes actuellement car, organisées autour d'une application type (un *core*) permettant de simuler rapidement des tissus cellulaires, il est possible de développer ses propres algorithmes. Néanmoins, les remarques que l'on peut faire concernant ces logiciels-frameworks sont assez semblables à celles que j'ai faites pour *SimCells*. Essayer d'aller plus loin que ce qui est proposé par défaut dans *PhysiCell* ou *PhysiBOSS* impose de descendre bas niveau dans la structure existante du programme et de le modifier pour permettre de nouvelles fonctionnalités comme simuler une biomécanique réaliste ou décrire un comportement intracellulaire. C'est en réalité tout à fait faisable avec *Physicell*, mais il faut pour cela développer une culture de ce logiciel dans le laboratoire et la maintenir.

Dans ces solutions, beaucoup de choses ne sont pas permises ; en tout cas, pas facilement. Souvent les algorithmes mis en œuvre sont des boîtes noires auxquelles il est difficile de se fier si l'on souhaite développer les aspects quantitatifs d'un modèle : il est en effet peu envisageable de publier un article sur une modélisation de biomécanique tissulaire sans savoir comment celle-ci est gérée par le logiciel ; même chose concernant la manière exacte dont les entités simulées interagissent ou se comportent individuellement. De plus, il est souvent difficile de superposer des modèles fondés sur des équations différentielles et des modèles multi-agents, de coupler des réseaux de régulation Booléens avec des systèmes de réaction-diffusion, de mixer éléments finis et systèmes multi-agents ... Sans compter que rien n'est prévu particulièrement pour simuler des dispositifs (et actions) de mesure et de manipulation biologique (ponctions, découpes, injection de réactifs, greffes ...) et dans ces mêmes simulations de systèmes biologiques, ni même de simuler des dispositifs d'imagerie biologique (ex : images de microscopie simulées ...).

Ce paysage, qui s'étend des bibliothèques bas niveau aux solutions dédiées haut niveau est certes nécessaire au modélisateur ; chacun y trouvera d'ailleurs tôt ou tard une utilité se manifestant

69 fork

dans le besoin d'intégrer des structures informatiques et algorithmes performants dans ses propres algorithmes, ou dans le besoin de réaliser un prototype plus ou moins fin pour se faire une image mentale du problème étudié ou simplement pour faciliter la communication entre partenaires d'un projet de modélisation au sein d'une équipe interdisciplinaire ou dans le cadre d'enseignements (citons à ce titre le logiciel *NetBioDyn* [Logiciels:NetBioDyn qui remplit précisément ces rôles et que nous utilisons intensément pour l'enseignement de la modélisation), voire d'implémenter un modèle complet, dans un cadre strictement délimité par les possibilités d'une solution comme *HSIM*, *SimCells* ou *PhysiCell*.

Entre ces 2 niveaux d'abstraction, il faut compter sur les frameworks ouverts couvrant parfois plusieurs niveaux logiciels (UI, structures, algos, meta-structures, ...). C'est un peu le cas de *PhysiCell* avec les limites que nous venons d'évoquer. Un autre exemple est le framework *CamiTK* [Logiciels:CamiTIK] développé au laboratoire TIMC. Ce framework est très ouvert mais il est principalement pensé pour l'usage des GMCAO⁷⁰, intégrant manipulation d'images médicales, contrôle de périphériques utilisés en GMCAO (caméras, dispositifs d'acquisition 3D ...). J'ai étudié un moment la possibilité d'utiliser *CamiTK* comme framework pour la modélisation hybride et multi-échelles en biologie, mais la façon dont est organisée *CamiTK* facilite principalement l'interactivité mais ne met pas immédiatement l'accent sur le calcul intensif, orienté simulation donc. En revanche, *CamiTK* dispose d'un atout majeur dans sa conception : il est pensé pour la modularité !

Enfin, je n'ai pas encore vraiment parlé de l'expressivité et de l'utilisabilité du code : certains logiciels comme *SimCells* ou *NetBioDyn*, par exemple, mais aussi des framework livrés avec un système d'interface par défaut comme *CamiTK*, mettent l'accent sur cela : il est facile et rapide d'exprimer ce que l'on souhaite simuler ou développer dans ces systèmes logiciels. C'est un point très important car, d'abord, cela permet d'aller le plus directement possible à l'objectif qui est la modélisation (l'implémentation de modèles) mais aussi car cela évite de jongler avec des astuces de programmation. Enfin, une structuration propre, parfaitement cadrée du logiciel ou du framework évite les erreurs de développement.

J'ai dessiné ci-dessus une analyse somme toute assez froide et fort critique des solutions existantes. Je tiens à ce stade à préciser mon admiration pour ces logiciels et surtout leurs concepteurs, d'une part parce qu'ils ont fait des choix parfois courageux sachant les limitations de leur produit, parce qu'ils ont pris du temps de recherche pour faire du développement logiciel, activité fort mal reconnue est pourtant nécessaire aux équipes de modélisation et indissociable de leur activité, enfin parce que ces solutions sont abouties, fonctionnelles !

Comment puis-je alors proposer quelque chose de valable qui puisse s'intégrer dans ce contexte et qui réponde à un besoin dans le paysage de la modélisation en biologie que j'ai décrit tout en se démarquant des solutions existantes ? Et pourquoi ?

70 Gestes Médicaux et Chirurgicaux Assistés par Ordinateur. <https://www-timc.imag.fr/GMCAO>

C'est dans le positionnement de la solution que je souhaite développer, MESS, intermédiaire entre bibliothèque bas niveau et solution intégrée que je vais situer ma contribution. Pour le 2ème logiciel que je développe, *Network Designer*, c'est l'absence d'alternatives logicielles et le besoin d'un tel outil pour la modélisation et l'analyse des réseaux de régulation qui motive son développement.

Voici ma proposition pour ces deux solutions logicielles, *MESS* et *Network Designer*, que je développe et qui continueront de s'intégrer naturellement à mon activité de recherche des années à venir.

2.2. Network Designer 2

2.2.1. Historique

La première version de *Network Designer* (figure A2.1) a été écrite en 2008 en C++03 et Qt 4 par Hedi Ben Amor, alors étudiant en thèse sous ma co-direction et celle de J. Demongeot. Ce logiciel répondait au besoin de pouvoir modéliser graphiquement des réseaux Booléens à seuil par des actions de la souris (ajout de nœuds, d'arcs, ...) et une configuration de leurs paramètres dans un panneau latéral, et de simuler leur comportement (les nœuds actifs étaient en vert ; les inactifs en rouge) pas à pas ou d'une traite (exportation dans un fichier de données). Rapidement, le logiciel s'est en plus doté de quelques fonctionnalités supplémentaires comme le calcul, pour un réseau donné, des attracteurs et des bassins d'attractions correspondants, l'estimation de la robustesse des comportements par rapport à des changements d'états probabilistes, ... mais également une possibilité d'exporter les designs de réseaux sous la forme de leur description logique en Prolog, de manière à pouvoir être utilisés dans des opérations d'inférence logique par contrainte, comme nous le faisons actuellement en ASP.

Cette première version était un premier pas vers ce que je souhaite développer maintenant : un véritable cahier de laboratoire intelligent dédié aux réseaux d'interaction biologiques, aux réseaux de régulation génétique en particulier.

Elle souffrait cependant de nombreuses limitations, d'abord dans ses fonctionnalités, ensuite dans son ergonomie qui n'était pas la plus intuitive qui soit, enfin et surtout dans son extensibilité. En effet, ce logiciel a été conçu exclusivement pour un modèle précis de réseaux (les TBNs). Il était très difficile de modifier le programme existant pour prendre en compte d'autres formalismes (réseaux Booléens propositionnels PBNs, réseaux Booléens à signe SBNs, réseaux multi-valués propositionnels ou à seuil ...). Les systèmes de simulation et d'analyse eux même étaient pensés pour les TBNs et difficiles à repenser pour d'autres types de réseaux. Ces systèmes de simulation, dans une bonne logique de programmation objet, étaient intégrés aux objets sur lesquels ils s'appliquaient, les réseaux. De même, les différentes propriétés (plan de mise à jour, température, *etc*) que pouvaient prendre un réseau étaient encodées dans l'objet réseau lui même ; difficile dans ces conditions d'adapter le logiciel à tout type de réseaux formels. Enfin, et c'était probablement la contrainte la plus forte pour améliorer ce logiciel, le noyau fonctionnel n'était pas indépendant du système d'interface Qt, rendant presque impossible de repenser cette interface.

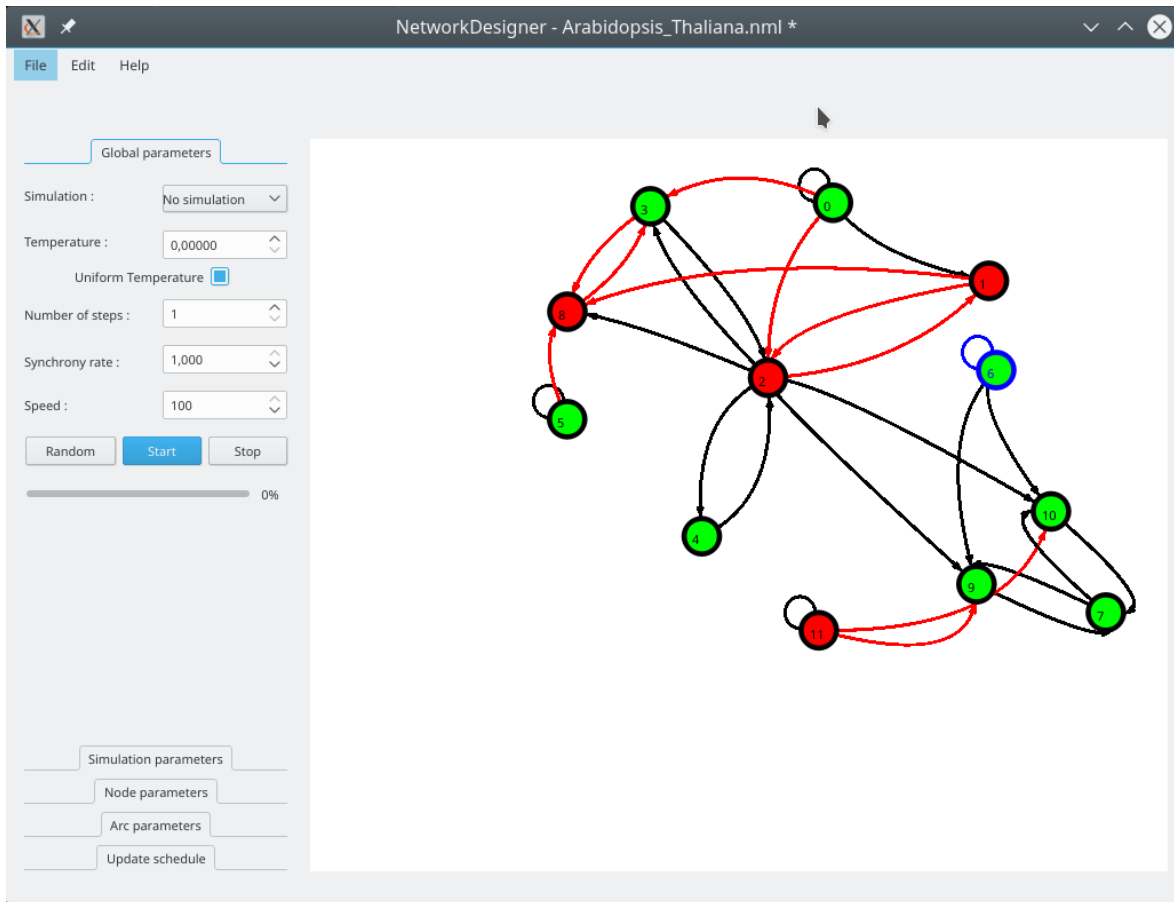


Figure A2.1. Network Designer version 1 (2008). Dans cette version, l'utilisateur pouvait concevoir un réseau de type TBN dans une fenêtre graphique unique, le paramétrer dans un bandeau latéral et simuler son comportement.

Après un essai infructueux dans ce sens que j'ai fait il y a quelques années, j'ai décidé de réécrire intégralement *Network Designer* dans une deuxième version. De plus, j'ai décidé de repenser le logiciel dans son ensemble en l'axant sur les besoins du modélisateur, théoricien ou biologiste. Le choix a aussi été fait de changer de langage de programmation et de système d'interface. Java et JavaFX ont été choisis pour 2 raisons : d'abord, Java, par sa simplicité (il s'agit ni plus ni moins que de C avec des objets, contrairement à C++ qui va très loin dans les possibilités de typage) évite de se perdre dans le typage complexe et les fonctionnalités poussées et permet en conséquence un développement rapide. Ensuite, JavaFX est un système d'interface utilisateur (UI) très bien pensé, très léger et facile à intégrer dans un programme. Sa lecture est en outre très claire, ce qui permet des modifications aisées *a posteriori*. Enfin, la possibilité de coupler, peut-être un jour, *Network Designer* à *SimCells*, m'apparaît très séduisante. *SimCells* est aussi écrit en Java et JavaFX. Un couplage des logiciels sous forme de modules pourrait être envisageable et aurait du sens : le modélisateur pourrait alors concevoir des réseaux, en inférer, les analyser, les simuler, puis les injecter comme comportements moteurs dans les cellules simulées dans *SimCells*. Dans ce contexte, *SimCells* bénéficierait alors du noyau fonctionnel et des moteurs développés pour *Network Designer*.

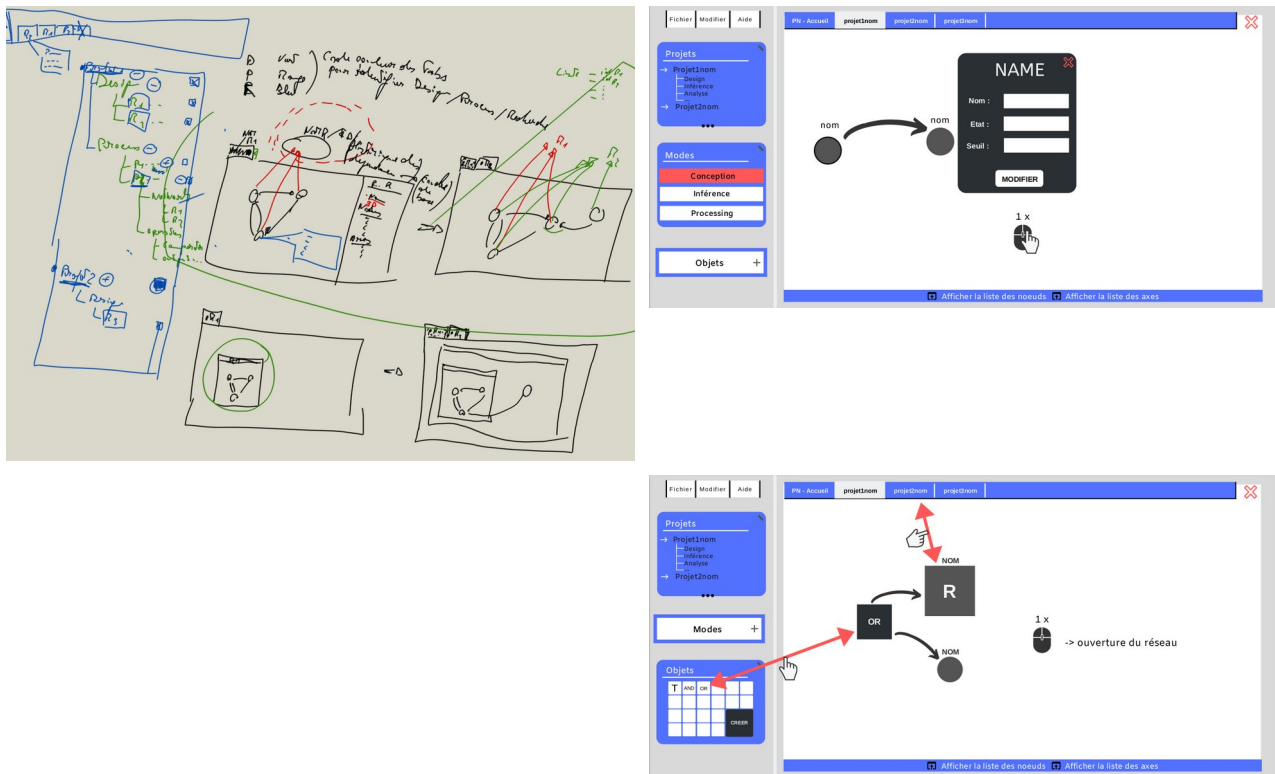


Figure A2.2. Maquettages de Network Designer version 2 (2018-20210). Maquettes papier et infographiques de Network Designer faites entre 2018 et 2020.

Depuis environ 5 ans, je réfléchis à la forme que pourra prendre cette nouvelle version. Ce qui est long, c'est d'une part d'imaginer une interface et une ergonomie adaptée au public visé, aussi bien des théoriciens des réseaux d'interaction que des biologistes peu enclins à la modélisation, des systèmes inter-opérationnels (le design de connaissances, les outils d'analyse et les outils de recherche de réseaux et d'inférence. Cela a impliqué de nombreuses phases de maquettage sur papier et numériques (figure A.1 et A.2.3), beaucoup d'essais d'interfaces, beaucoup de dialogue. La programmation de la nouvelle version a pris un tournant pendant la période de la crise Covid en mai-juin 2020, sous la forme d'un projet réalisé par 4 étudiants de Polytech en TIS4, projet se substituant à leur stage de fin d'année annulé. Ces étudiants ont produit une première version (très incomplète) de ce qu'allait devenir *Network Designer*. Leur version du système d'interface était encore trop imparfaite : le logiciel ne comportait alors que 2 couches : un noyau fonctionnel autonome et une interface utilisateur. Début 2021, 2 élèves de BTS informatique ont repris ce projet et nous avons encore repensé son architecture générale qui comprend maintenant 3 couches : un noyau fonctionnel autonome, une couche d'interface et une couche intermédiaire, faisant le lien entre les deux autres. Enfin, pour leurs stages respectifs de L3 et de M1, deux très bons étudiants, N. Marinho et F. Birkner ont rendu une partie du logiciel fonctionnelle et on surtout contribué à en améliorer l'architecture. C'est cette version en cours d'élaboration que je vais présenter.

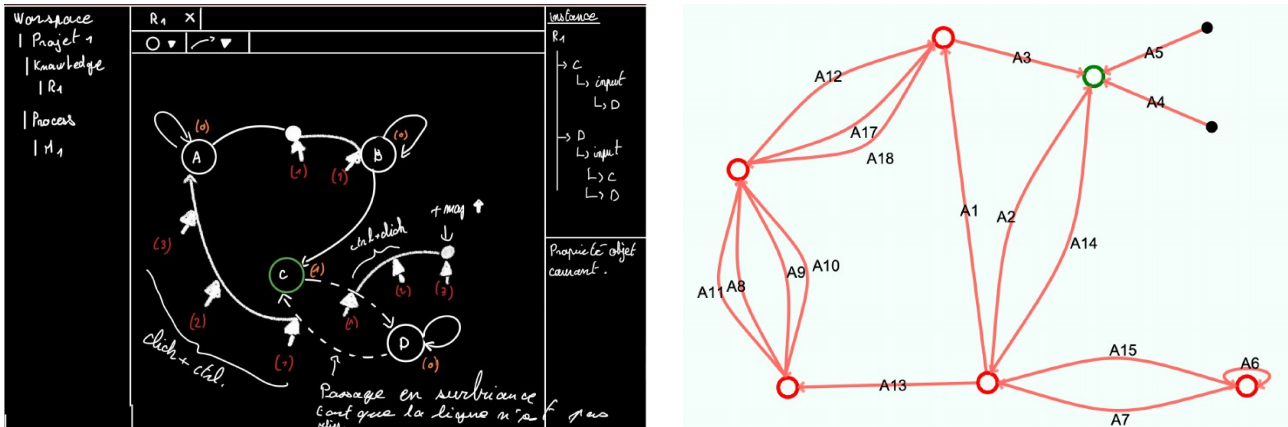


Figure A2.3. Maquettage et interface implémentée de Network Designer version 2. (gauche) Maquette de Network Designer et (droite) test fonctionnel d'une interface avec ses nombreuses opérations, réalisés par N. Marinho en 2021.

2.2.2. Fonctionnalités

Le logiciel se présente (et est pensé) comme une IDE (interface de développement intégrée) avec laquelle les programmeurs Java , C++ ... développent et gèrent des projets de programmation. De la même manière, l'utilisateur travaillera dans un espace de travail (*workspace*) pouvant accueillir 1 à plusieurs projets (*project*) (voir figure A2.4). Chaque projet sera composé d'une collection de travaux (*work items*), l'équivalent de fichiers de programmes *.java* ou *.h* et *.cpp* dans des projets Java ou C++, de types variés.

Le logiciel comprend 2 grands modes de fonctionnement produisant des *work items* différents :

- la conception (*design*) de connaissances (*knowledges*), incluant :
 - les connaissances sur la structure des réseaux (*network*) incluant les types de fonctions, les types de relations les unissant (arcs) mais aussi des meta-informations comme des motifs.
 - les connaissances sur la dynamique des réseaux (*dynamics*)
 - les ensemble de réseaux (*network sets*)
- le calcul (*processing*) impliquant ou générant ces connaissances et comprenant :
 - l'analyse des connaissances et la simulation de réseaux (*analysis process*)
 - la recherche de réseaux ou de dynamiques par inférence (*inference process*)

Chaque *work item* est accessible depuis un arbre des projets dans une colonne latérale et peut être développé dans un espace graphique ouvert dans un nouvel onglet. Les onglets sont identifiés par type (réseau, dynamique, analyse, inférence) par un code couleur. De plus, une description complète des composants d'un *work item* (ex : un réseau) est disponible dans une autre fenêtre latérale.

Dans ces onglets graphiques, l'utilisateur édite les *work items* de façon principalement graphique, en manipulant des formes géométriques (ronds, carrés ... flèches) associées aux objets à modéliser et aux actions à mettre en place. Il peut bien sûr les paramétrer via des boîtes de dialogue accessibles par des menus contextuels par exemple.

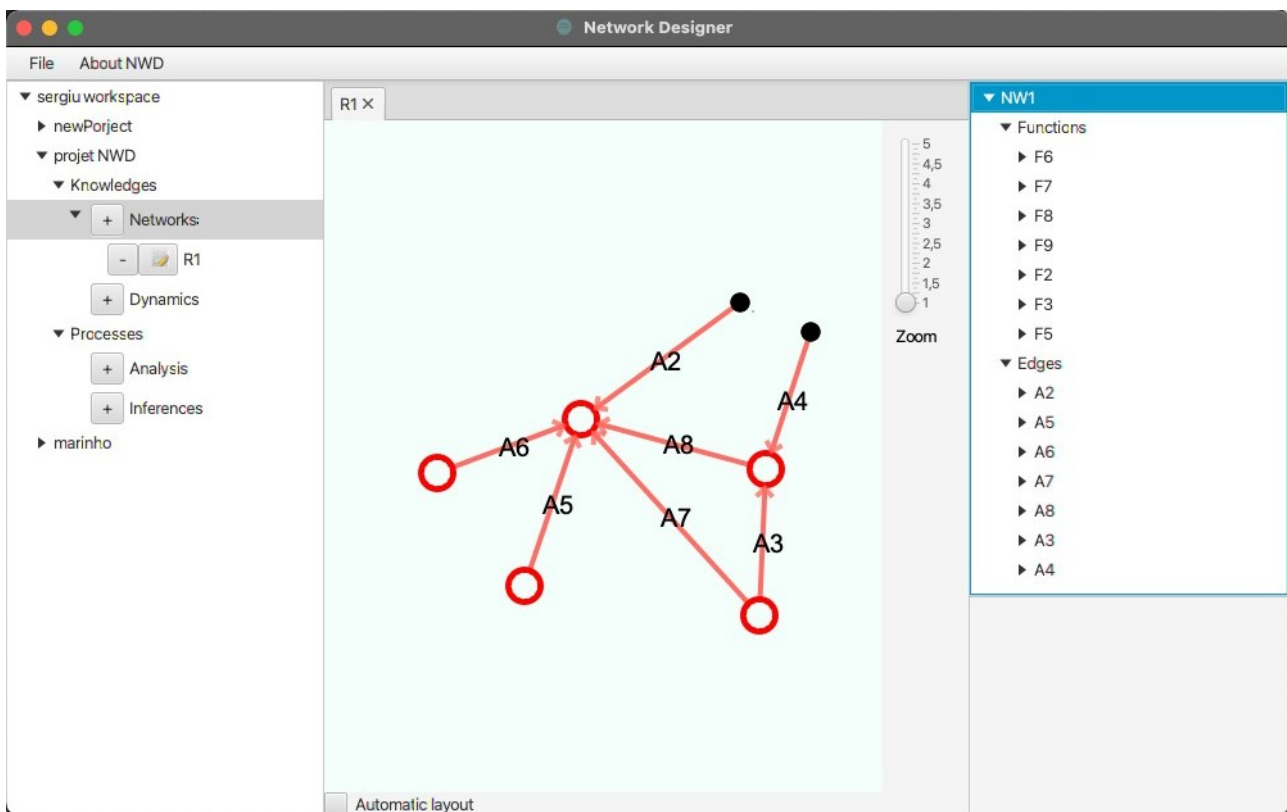


Figure A2.4. Version implémentée de Network Designer version 2. On voit à gauche l'arborescence du workspace contenant 2 projets (*newProject* et *projet NWD*). Le projet *projet NWD* est détaillé et on voit qu'il contient notamment un réseau nommé *R1*.

Au sein d'un projet, tout objet (un réseau, une dynamique, un set, un calcul, des clauses ...) doit être vu comme une sorte de programme susceptible d'être utilisé dans plusieurs situations (dans d'autres *work items*), si tant est que l'action est valide, et pouvant être relié à d'autres objets. Par exemple, un utilisateur qui aurait conçu un réseau complet *N1* (le nom choisi ici est arbitraire et pourrait tout aussi bien être *Arabidopsis_thaliana*) de dimension 3, commencé à élaborer une structure partielle de réseau *N2* de dimension 3 aussi, et défini une dynamique incomplète *D* sous la forme d'une table de vérité ou d'un graphe de transition, pourrait alors d'une part créer un process d'analyse et simulation dans lequel il glisserait le réseau *N1* depuis le bandeau de projet vers l'écran graphique, ce réseau apparaissant alors comme une boîte pouvant éventuellement être connectée à d'autres objets, ajouterait un moteur de simulation particulier pioché parmi plusieurs disponibles, ce moteur apparaissant comme une autre boîte, relierait ce moteur à la boîte du réseau *N1*, et ajouterait trois vues, c'est-à-dire deux nouveaux objets graphiques pouvant lire les données contenues ou émises par divers objets, par exemple une vue *V1* ciblant *N1* et montrant ses caractéristiques générales dans une petite fenêtre, une vue *V2* et une vue *V3* ciblant toutes deux le moteur de simulation, la première produisant un graphique d'activité du réseau, la deuxième écrivant les données de simulation dans un fichier. L'utilisateur pourrait ensuite définir un autre process de type inférence dans lequel il utiliserait le réseau *N2* défini partiellement (par exemple un réseau de dimension 3 dans lequel seuls quelques arcs sont définis et de paramètres connus et certains arcs sont connus mais de paramètres

inconnus) et la dynamique partielle D (définie par exemple pour seulement 2 nœuds), toutes deux des connaissances partielles qu'il relierait à un moteur d'inférence, doté de clauses qu'il aurait également définies comme des objets graphiques, et qui chercherait l'ensemble des réseaux satisfiables de dimension 4 possédant au moins la structure partiellement définie de $N2$ et dont au moins 2 nœuds réaliseraient la dynamique D . Le moteur d'inférence, traduisant ces connaissances et clauses en ASP et, en faisant des appels ASP, générerait cet ensemble de réseaux satisfiables dans une nouvelle entité, un *set* auquel l'utilisateur aurait donné un nom, $S1$. Ce *set* constituant une nouvelle connaissance, se retrouverait d'office associé au projet comme nouveau *work item*. Enfin, pour continuer sur cet exemple de projet, l'utilisateur pourrait décider d'analyser la complexité de l'ensemble des réseaux du *set*, et l'ensemble de leurs comportements dynamiques, comme nous l'avons fait dans [Christen, 2019 ; Segretain, 2020a et 2020b], par l'intermédiaire d'un autre processus d'analyse et, dans un quatrième processus, d'inférence cette fois, il pourrait rechercher dans ce *set* $S1$, le sous-ensemble de $S1$ des réseaux contenant la structure de $N1$ à l'aide d'un autre moteur d'inférence. Il obtiendrait alors un autre *set* $S2$, pouvant à son tour être utilisé comme base de connaissances.

Cet exemple permet de se faire une idée de l'objectif recherché pour cette nouvelle version de *Network Designer*. C'est ici un exemple de projet de taille limitée, mais la façon dont seront gérés les bases de connaissances (données stockées des réseaux et dynamiques, *etc*) et les processus (parallélisation via la lib *Java Parallel Pipeline*, ...) devrait permettre des projets d'ampleur importante (plusieurs milliers à milliards de réseaux et dynamiques).

Beaucoup d'autres fonctionnalités sont d'ores et déjà prises en charge, notamment la possibilité de travail collaboratif, chaque utilisateur étant répertorié et les actions réalisées associées à un utilisateur, ou encore la possibilité de dupliquer des projets, de copier ou déplacer des objets entre projets d'un même *workspace*, *etc*.

2.2.3. Architecture

Le logiciel est organisé en 3 couches successives :

- la couche *Core* contenant le noyau fonctionnel, à savoir la gestion des réseaux et des différents types de fonctions, arcs et paramètres associés, des meta-informations comme les motifs (sous-réseaux) ou des patterns (sous-réseaux aux fonctions et arcs indéfinis), des dynamiques, des ensembles de réseaux (sets), des schémas de mise à jour, des opérateurs dont les moteurs de simulation, d'inférence, d'analyse, les filtres et les propriétés. Cette couche contient en outre un système de simulation parallèle doté de son constructeur de simulation, un travail réalisé par F. Birkner.

Cette couche se présentant comme une API formée actuellement d'une 50aine de classes Java, est indépendante de tout logiciel. Elle pourrait être exploitée telle quelle dans un autre logiciel que *Network Designer*.

- La couche *NWD* permet la gestion des projets et espaces de travail de *Network Designer* et la configuration du logiciel. Elle ne contient aucun élément graphique (pas d'interface utilisateur). Elle permet notamment la mise en œuvre du *Core* en le rendant accessible par l'interface et les représentations graphiques. C'est cette couche également qui assure la gestion des sauvegardes et chargements des données dans des fichiers. L'ensemble des données (données sur les réseaux, les dynamiques, les sets, mais aussi leurs représentations graphiques, les différents work items implémentés, ..., est stocké dans un ensemble arborescent et interdépendant de fichiers au format *xml*.
- La couche *NWDUI* correspond à tout le système d'interface et de représentations graphiques. Cette couche n'interagit pas directement avec le *Core* mais passe exclusivement par la couche *NWD*. Elle comprend là aussi de nombreux éléments (classes Java) comme l'arborescence des projets dans un *workspace*, les onglets (*tabs*), les fenêtres de dialogue, les menu (principal et contextuels), mais aussi les représentations graphiques des *work items* dans les différents onglets ouverts.

Actuellement, les 3 couches logicielles existent. La couche *Core* est très avancée. La gestion des *workspaces*, *projects* et *work items* fonctionne aussi. L'utilisateur peut créer un réseau, l'éditer, partiellement le paramétrer. En revanche, le système de gestion graphique de spécification des *process* n'est pas encore programmé.

2.3. MESS

MESS, pour *Mundi et Entitates in Simulacro Systematae* (Monde et Entités dans un Système Modèle) est un framework de programmation écrit en C++14 et dédié à la simulation numérique, principalement de modèles spatialisés 2D ou 3D. C'est un très gros projet que j'ai initié il y a longtemps (10 ans) qui est partiellement utilisable et qui croît régulièrement. Ce projet a pour objectif de répondre à des besoins de capitalisation du code d'une part, d'accélération de la programmation de modèles numériques spatialisés notamment, hybrides potentiellement, comme décrit dans la partie introductive de cette annexe (voir figure A2.5).

MESS comporte principalement 2 aspects :

- C'est un moteur de simulations intégré comprenant une gestion multi-thread des processus de calcul, graphiques, d'interface utilisateur, d'entrées-sorties.
- C'est un ensemble (ouvert) de structures de données et algorithmes adaptés à la modélisation des systèmes biologiques, en particulier comprenant de très nombreux éléments en interaction dans un environnement spatial.

Mais, MESS est surtout un framework structurant l'écriture de programmes C++ dédiés à la simulation, en vue de leur réutilisation. Le principe est que MESS ne décharge le modélisateur que

d'une partie de la programmation de son modèle. Une bonne partie de la programmation de son modèle reste cependant à sa discrétion. Le programmeur est néanmoins incité, via l'utilisation d'objets déjà programmés (structures de données et algorithmes) et via un *coding style* très présent, à produire du code modulaire, performant et réutilisable. Comme de nombreux systèmes habituellement pénibles à programmer sont implémentés et doivent juste être « connectés », le modélisateur peut se concentrer sur l'implémentation de son modèle.

La programmation se fait au sein d'objets pris en charge par les *process* de MESS en tant qu'unités calculable, affichables, *etc.* Ces *process* organisent l'exécution concurrente des tâches exécutées de façon asynchrone. Les différents processus (*compute process, graphical process, io process ...*) communiquent entre eux via une queue de messages concurrente. Les données de simulation sont protégées et contrôlées par le processus maître dans MESS qui est le processus de calcul (*compute process*). C'est lui seul qui peut rendre les données disponibles aux autres processus. Les autres processus (par exemple le processus graphique) ne peuvent faire que des demandes de mise à disposition des données de simulation.

Il est rare que l'intégralité des données d'une simulation soit requise pour faire des sauvegardes ou destinées à être affichées ; il ne s'agit souvent que d'une petite partie. De plus, l'objectif est d'avoir un système donnant priorité au calcul et non à l'affichage. Il s'agit cependant de conserver une très bonne interactivité dans le cas de simulations sur lesquelles on souhaite interagir ou que l'on souhaite analyser en cours de route. Pour cette raison, le processus graphique dispose de ses propres données à partir desquelles vues et interactions sont faites. La mise à jour de ces données peut-être demandée au processus maître (processus de calcul) avec une certaine fréquence définie à l'avance, cette fréquence pouvant être ajustée par le processus maître lui-même. La copie en mémoire vive de données de simulation en données destinées à l'affichage d'informations (textuelles ou graphiques) est en effet rapide et ne requiert pas d'être faite à chaque itération.

Côté graphique (dont la présence est facultative, le processus pouvant être purement computationnel), le choix d'un gestionnaire de fenêtres (GLUT, SDL, SFML, GLFW, ...) ainsi que d'un système graphique d'interface utilisateur (GUI, tel AntTweakBar ou MiniGUI) est laissé à l'utilisateur. Plusieurs systèmes pré-programmés sont disponibles.

Pour gagner du temps de programmation et surtout pour structurer son programme, le programmeur dispose d'un ensemble de structures de données et algorithmes, ensemble potentiellement croissant avec le développement de nouveaux projets. Parmi ces objets vus comme des briques de programmation haut niveau, on trouve des environnements 2D ou 3D en grille, des systèmes de définition du voisinage (maillage), des systèmes capable de gérer de grandes collections de particules ainsi que leur détection dans un environnement, leurs collisions, leurs déplacements dans l'environnement, *etc.* On trouve aussi des structures plus bas niveau comme des vecteurs mathématiques, des matrices ou des tenseurs et les opérations associées, des systèmes de tamponnage mémoire, un système de gestion du temps et des évènements, un système d'unités

internationales et des opérations associées qui permet de typer les données, *etc.*

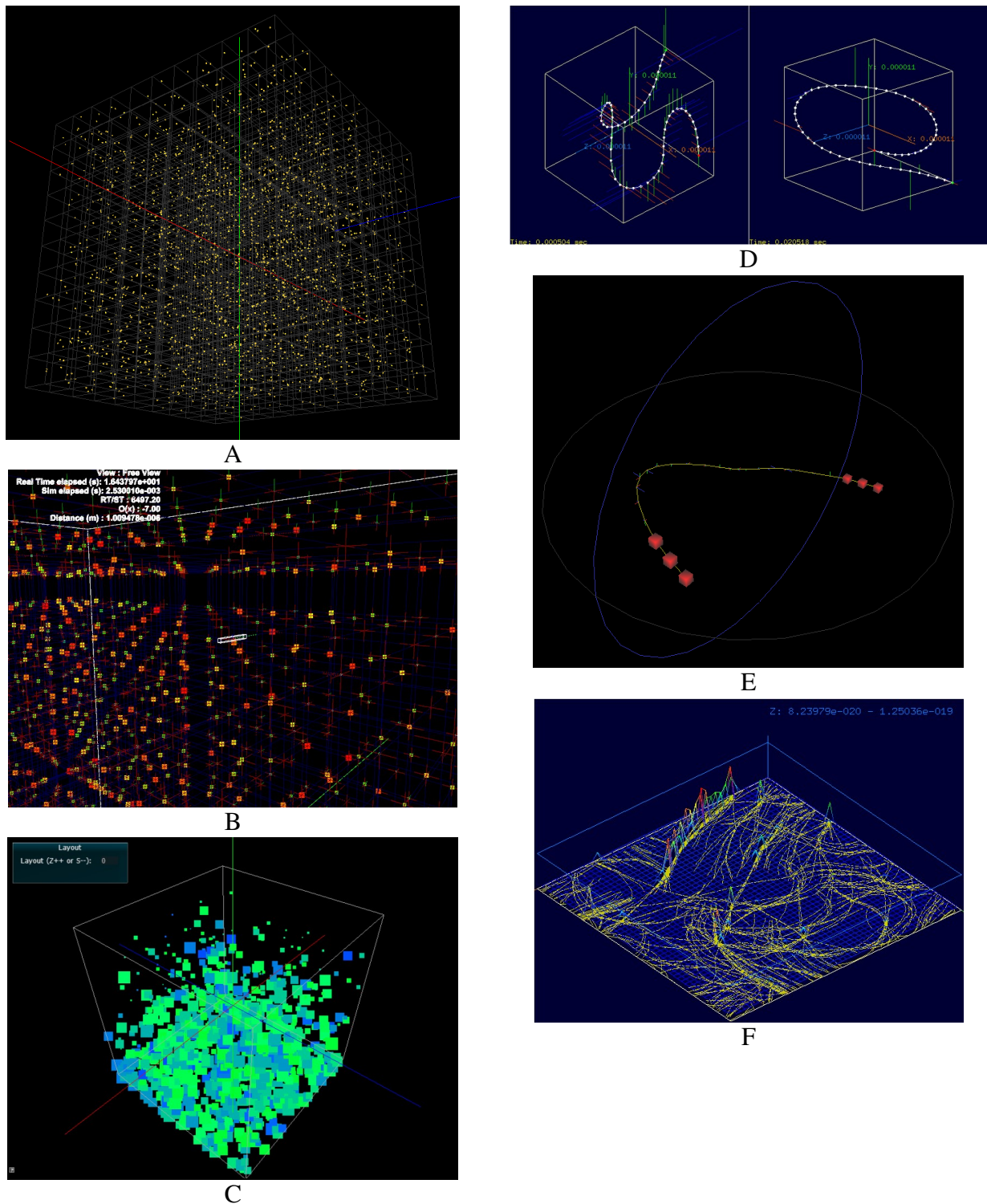


Figure A2.5. Quelques sorties écrans de simulations réalisées sous MESS. (A) Une grille 3D et 30000 particules présentant chacune une position et une orientation dans l'espace. Cette grille est optimisée pour la gestion des systèmes multi-agents. (B) Une grille 3D affichant les relations de voisinage et des concentrations (taille et couleur des points). (C) Une simulation multi-agent de vie artificielle (chaque point est une créature, la taille représentant sa masse et la couleur la taille de son génome). On voit la présence d'une fenêtre interne, un mini-GUI (ici AntTweakBar). (D) et (E) des fibres visco-élastiques dans un environnement 3D grid-free. Les forces appliquées sont visualisés, ainsi que des points d'attache (carrés rouges en (E)). (F) Des fibres visco-élastiques dotées de capacités de réactions de

croissance ou décroissance à leurs extrémités, simulées en grid-free et projetées dans une grille 2D stockant des concentrations en particules libres. L'affichage montre les fibres, et la densité en fibres.

MESS, comme son nom l'indique, est un chantier important mais compliqué sur lequel je reviens régulièrement. Je devrais m'y atteler davantage dans les temps à venir pour implémenter les systèmes de vie artificielle que j'ai évoqués dans ce manuscrit, mais aussi pour porter certains modèles de morphogenèse sur lesquels l'équipe a travaillé ces dernières années, en particulier un modèle biomécanique de cellule dont le modèle a été publié récemment [Manificier, 2020], dans l'optique d'en faire un modèle hybride couplant régulation, métabolisme et biomécanique cellulaire et biomécanique du milieu environnant.