Modélisation de la dynamique des réseaux biologiques : applications en génétique et immunologie

Hana Hazgui

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **MBS-Modèles, méthodes et algorithmes en biologie, santé et environnement**

Arrêté ministériel : 7 août 2006

Présentée par

Hana HAZGUI

Thèse dirigée par **Jacques DEMONGEOT**

Préparée au sein du **Laboratoire AGIM (AGe-Imagerie-Modélisation)**
Dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Modélisation de la dynamique des réseaux biologiques : applications en génétique et immunologie

Thèse soutenue publiquement le **11 Décembre 2015**,
devant le jury composé de :

Monsieur, Christian DROUET
Professeur, Université Joseph Fourier„ Président
Madame, Georgia BARLOVATZ-MEIMON
Professeur, Université d'Evry, Rapporteur
Monsieur, Sylvain SENE
Professeur, Université d'Aix-Marseille, Rapporteur
Madame, Alexandra HENRION-CAUDE
Directrice de recherche à l'INSERM, Hôpital Necker, Examinatrice
Monsieur, Mustapha RACHDI
Professeur, Université Pierre Mendès France, Examinateur
Monsieur, Nicolas VUILLERME
Maître de conférences, Université Joseph Fourier, Invité
Monsieur, Jacques DEMONGEOT
Professeur, Université Joseph Fourier, Directeur de thèse

Remerciements

Quelle meilleure façon, pour avouer sa reconnaissance, pour témoigner de sa gratitude envers mon encadreur, Jacques Demongeot, avec toute sa rigueur scientifique, son savoir, sa générosité et son altruisme, que de le hisser au rang d'un père spirituel qui m'a tout donné et j'ai tout gardé, tant il est à mes yeux une source intarissable de bienveillance et de bravoure. De telles qualités et d'autres encore, furent pour ce modeste travail de recherche ce que la cheville de roue est au chariot qui roule....

Merci Jacques

Je tiens à remercier les professeurs Georgia Barlovatz-Meimon et Sylvain Sené pour avoir accepté d'examiner mon travail de thèse en tant que rapporteurs. Je remercie également Christian Drouet, Alexandra Henrion-Caude, Mustapha Rachdi et Nicolas Vuillerme dont la présence pour discuter de ce travail me fait honneur.

Mes remerciements pour leur soutien et leurs collaborations durant mes années de thèse vont à tous mes collègues et amis : Sarah Boughanmi, Malika Dakir, Ali Hamie, Philippe Cousin, Islem Mefteh, Mohamed Ghassani, Dhaou Ghoul, Olivier Hansen, Caroline Messina Dos Santos et Xavier Ronot.

Je remercie aussi mes professeurs Mohamed El Methni et Taghi Barumandzadeh.

Je remercie également Danielle Dhouailly pour son soutien précieux ainsi que ses encouragements.

Enfin, je ne passerai pas sous silence la profonde gratitude que je désire exprimer à mes frères Hichem et Haykel, à mes sœurs Nedra et Habiba, ainsi qu'à mes grands-parents.

A mon père Mongi,
A ma mère Najet,
Vous êtes la lumière de ma vie.

Résumé

Dans cette thèse, nous nous intéressons à la modélisation statistique de données biologiques, et plus particulièrement à l'étude de l'information génétique et protéique.

Dans un premier volet, nous avons amélioré un modèle statistique des données immunologiques existant chez la souris, que nous avons transposé à l'homme, afin d'étudier les différentes recombinaisons qui apparaissent au sein du thymus, à la fin de la vie embryonnaire, entre les segments des gènes de la portion V(D)J du chromosome 14 humain, appelées **recombinaisons V(D)J**.

Dans un deuxième volet, nous avons étudié l'information génétique par le biais des **réseaux de régulations génétique**, celui de la maladie familiale « atrésie biliaire », ainsi que dans les réseaux de contrôle du système immunitaire, que nous avons appelés « Immunetworks ».

Dans un troisième volet, nous proposons une nouvelle approche de la compression des données biologiques, qui intègre une étape de modélisation des processus dynamiques qui leur ont donné naissance : nous avons appelé cette approche la transformée **Dynalet** et nous l'appliquons, entre autres, à des signaux de spectrométrie RMN (Résonance Magnétique Nucléaire) des protéines et acides nucléiques. Cette méthode consiste à convertir les signaux de spectrométrie en sons, afin de construire une lutherie anharmonique permettant de reproduire les pics de relaxation périodisés, issus des spectres RMN des 20 acides aminés, ainsi que de ceux des 4 bases nucléiques azotées.

Abstract

In this thesis, we focus on statistical modeling of biological data, and more particularly on the study of genetic and protein information.

First, we have improved a statistical model of existing immunological data in mice, we have transposed it to human, in order to study the various recombinations that occur within the thymus, at the end of the embryonic life, between segments of genes from the portion V (D) J of the human chromosome 14, called **recombinations V(D)J**.

Secondly, we studied the genetic information through **genetic regulatory networks**, for example in the case of a family illness called the "biliary atresia," as well as in the immune system control networks, which we have called "Immunetworks".

In a last part, we propose a new approach for biological data compression, which includes a step of modeling the dynamic processes that gave rise to them : we call this approach the **"Dynalet"** transform and we apply it, among others, to NMR spectrometry signals, i.e., Nuclear Magnetic Resonance spectra of proteins and nucleic acids. This method consists in converting the peaks of the spectrometric signals into sounds, in order to construct an anharmonic instrument capable to reproduce periodized relaxation peaks from the NMR spectra of the 20 amino acids, as well as those of the 4 nucleic bases.

Table des matières

Table des figures

Liste des tableaux

Introduction générale

PREAMBULE

Dès à présent, la complexité des données est telle que le biologiste ne peut plus raisonnablement faire de la biologie sans collaborer avec un mathématicien. L'intérêt principal de la modélisation est d'influencer la manière dont le biologiste réfléchit et donc planifie ses expériences. Un modèle n'a pas pour unique but d'extraire des informations des données expérimentales : un bon modèle permet de générer des prédictions et donc de nouvelles expérimentations.

Du gène aux réseaux de régulation

Un gène correspond à un segment d'A.D.N. conditionnant la synthèse d'une ou de plusieurs protéines et, donc, la manifestation et la transmission d'un caractère héréditaire déterminé. Le gène est donc une "unité d'hérédité", ainsi qu'un "message" (Demongeot et Tracque dans Eléments de Biologie, 2003). Si on veut comprendre le vivant au niveau cellulaire, on doit comprendre le fonctionnement des gènes, ou plutôt leur expression à travers les protéines synthétisées. Les gènes exercent leurs fonctions sous la forme de protéines. Ils contiennent un message qui doit donc être traduit en protéines : d'où la notion de l'expression des gènes. Cette expression se déroule en trois phases :

1. **La transcription :** correspond à la synthèse d'un copie d'ARNm à partir d'une séquence d'ADN. Elle comporte plusieurs étapes, telles que l'initiation, l'élongation et la terminaison
2. **La maturation :** correspond à une série de modifications du ARNm. Elle se déroule en deux étapes (la fixation d'un coiffe méthylée et la polyadénylation)
3. **La traduction :** correspond à la phase de synthèse des protéines à partir des ARNm. Elle se déroule en trois étapes (l'initiation, l'élongation et la terminaison).

Durant la dernière décennie, dans la foulée du séquençage des génomes complets, les informations moléculaires issues de la post-génomique n'ont pas seulement augmenté en quantité mais aussi en complexité. Chez les organismes eucaryotes pluricellulaires, tels que les mammifères, qui sont dotés de régions non codantes étendues, de nombreuses études montrent que plusieurs facteurs de transcription différents peuvent être impliqués dans la régulation d'un gène. En se fixant à l'ADN et en formant des complexes les uns avec les

autres, les facteurs de transcription ont deux modes d'action : à un niveau qualitatif, ils peuvent activer (effet positif) ou inhiber (effet négatif) la transcription de leur(s) gène(s) cible(s). Les régulateurs peuvent eux-mêmes être régulés, dans ce cas, ils participent à une voie de régulation génétique. Un gène cible est généralement régulé par une combinaison de facteurs de transcription, et un facteur de transcription peut réguler plusieurs gènes cibles. Il est donc possible de représenter qualitativement un réseau de régulation génétique comme un graphe dont les nœuds sont des gènes (facteurs de transcription. et gènes cibles) et les arêtes orientées représentent des effets transcriptionnels ou traductionnels activateurs ou inhibiteurs.

Outre la complexité des réseaux de régulation,génique, la séquence même d'un gène peut être modifiée par recombinaison, un phénomène qui est la base même de la réaction immunitaire. Le système immunitaire représente le système protecteur de l'organisme contre les virus, les bactéries ... Pour s'adapter aux différentes "attaques" de pathogènes, il utilise plusieurs récepteurs, afin de reconnaître les divers antigènes. La diversité de ses récepteurs provient de la diversité combinatoire des gènes qui se recombinent entre eux, de la diversité jonctionnelle et de la diversité d'appariement.

Buts des travaux de thèse

Les travaux entrepris ont portés sur trois grands axes, la recombinaison génétique dans le cadre de la réaction immunitaire, les réseaux de régulation génétique et la restitution du signal protéique sous forme sonore.

La recombinaison génétique

Nous présenterons, dans le premier chapitre, la modélisation de la recombinaison V(D)J. Un premier modèle a été proposé en 2006 par l'équipe américaine de Warmflash. Ce modèle mathématique décrit la fenêtre d'accessibilité des gènes.

Un modèle de recombinaison des gènes V(D)J a été validé, chez la souris, puis chez l'homme, au cours de deux thèses précédentes (M. Simonet 2008 et F. Thuderoz 2010). Lors de ces travaux, ont été confrontées les simulations de ce modèle avec les données expérimentales de l'équipe d'E. et P. Marche et avec celles d'autres équipes, comme celle de M.S. Krangel, à Duke U School of Medicine. Nous avons montré en particulier que les fréquences de réarrangement obtenues par le modèle étaient en accord avec ces données expérimentales chez la souris.

De plus, les travaux de l'équipe de Krangel ont montré que le promoteur TEA, qui contrôle l'accessibilité de la chromatine de la région J, permet l'ouverture de la chromatine du gène J61 au gène J49 et inhibe l'ouverture au-delà de J49. Suite à l'activation de J49, les J en aval deviennent accessibles (Abarrategui and Krangel 2007). Si nous confrontons nos résultats de simulation du modèle aux observations expérimentales de l'équipe de Krangel, nous constatons qu'ils correspondent à la première fenêtre d'accessibilité décrite dans notre modèle.

Les réseaux de régulation génétique

La notion de réseau de régulation a vu le jour très tôt dans des réflexions de M. Delbrück, vers 1949 : pour obtenir une différenciation cellulaire, le réseau doit avoir plusieurs états stationnaires. Ces notions ont été reprises et développées par S. Kauffman et R. Thomas entre la fin des années 1960 et le début des années 1970.

En fait, l'étude des comportements des gènes d'un réseau de régulation biologique, s'avère complexe. D'où la caractérisation des ces réseaux comme étant des systèmes dynamiques « complexes ». D'une manière générale, la plupart des systèmes complexes sont caractérisés par leur composition : système complexe est caractérisé par un nombre élevé d'éléments en interaction. De plus, le système lui-même est conditionné par des influences extérieures à prendre en compte. Nous avons étudié ces systèmes de régulation génétique, afin de déterminer les comportements des gènes dans le cas du réseau de contrôle de la morphogénèse des voies biliaires, pour la pathologie familiale "l'atrésie biliaire". Nous avons étudié également les réseaux de régulation dans le cadre de système immunitaire qui portent le nom "immunetworks".

Ces réseaux, afin d'être exploités, doivent passer par une étape primordiale : la modélisation. Il s'agit de déterminer les gènes ainsi que les interactions critiques, autrement dit, l'absence ou les changements des valeurs qui peuvent modifier considérablement le paysage du réseau (on parle du nombre, de la nature ou de la taille de bassins d'attraction).

Nous avons utilisé le concept de réseau d'automate booléen à seuil afin d'étudier les réseaux de régulation génétique. Ces réseaux d'automates ont été introduits vers les années 1940 par McCulloch et Pitts. Plus tard, Hopfield a exploité ces travaux vers 1982, en utilisant la fonction de transition locale qui sera expliquée dans le chapitre 2.

La méthode Dynalet

Dans le but d'exploiter les informations protéiques, nous avons utilisé la méthode Dynalet (Demongeot et al 2007). Cette méthode permet une exploitation sonore de l'information visuelle. En utilisant le signal protéique, nous proposons de convertir en sons des pics du spectre RMN, de manière à pouvoir distinguer à l'oreille des tracés indistinguables à la vue, dans les modes de visualisation classiques. Nous appelons cette approche le stéthoscope protéique et la restitution de l'information protéique sous forme sonore pourrait constituer une "première" importante en compression et rendu de l'information médicale d'origine biologique.

Le décryptage des réseaux de régulation, celui de la recombinaison génétique, ainsi que le stéthoscope protéique, outre leur intérêt théorique en biologie, ont des applications très pratiques dans le domaine de la santé.

Organisation de manuscrit :

Dans le premier chapitre, nous étudions la recombinaison V(D)J, en détaillant ce mécanisme. Les travaux relatifs à ce chapitre sont rapportés dans l'article soumis.

-H. HAZGUI, F. THUDEROZ, O. HANSEN, M.A. SIMONET, J.F. DYON, P.N. MARCHE, E. MARCHE and J. DEMONGEOT
Numerical Model for the Vα-Jα Gene Use in Human TRA/TRD locus : Recombination Dynamical Rules (submitted)

Dans le deuxième chapitre, nous étudions les réseaux de régulation génétique, ainsi que la robustesse de ces réseaux. Ce chapitre a donné lieu à plusieurs publications, suite à des conférences internationales ou dans des journaux scientifiques.

1. J. DEMONGEOT, H. HAZGUI , S. BANDIERA, O. COHEN and A. HENRION-CAUDE
   MitomiRs, ChloromiRs and general modelling of the microRNA inhibition.
   Acta Biotheoretica, 61, 367-383 (2013).

2. J. DEMONGEOT, H. HAZGUI & N. VUILLERME
   MicroRNAs : unspecific inhibitory regulation in immunologic control and in mitochondrial respiration.
   In : IEEE AINA' 13 & BLSMC' 13, IEEE Proceedings, Piscataway, 1509-1516 (2013).

3. J. DEMONGEOT, H. HAZGUI, J. ESCOFFIER and C. ARNOULT
   Inhibitory regulation by microRNAs and circular RNAs.
   In : Medicon'13, IFBME Proceedings 41, Springer Verlag, New York, 722-725 (2014).

4. J. DEMONGEOT, H. BEN AMOR, H. HAZGUI and J. WAKU
   Stability, Complexity and Robustness in Population Dynamics.
   Acta Biotheoretica, 62, 243-284 (2014).

5. J. DEMONGEOT, H. HAZGUI and A. HENRION-CAUDE
   Genetic regulatory networks : focus on attractors of their dynamics.
   In : Computational Biology, Bioinformatics and Systems Biology, eds. Q.N. Tran and H.R. Arabnia, Elsevier, New York

6. J. DEMONGEOT, H. BEN AMOR, H. HAZGUI and A. LONTOS
   La simplexité, dernier avatar de la complexité.
   Collège de France, Paris and OpenEdition, Marseille. http ://books.openedition.org/cdf/3393 (2014).

7. J. DEMONGEOT and H. HAZGUI
   Entropy for genetic networks. Role of RNA effectors (submitted).

Dans le dernier chapitre, nous présentons une autre approche, afin d'exploiter les données protéiques ; l'application de la méthode Dynalet. Cette méthode fait l'objet d'un travail en cours. Une publication est liée à ce chapitre :

-J. DEMONGEOT, O. HANSEN, A. HAMIE, H. HAZGUI, G. VIRONE and N. VUILLERME
Actimetry@home : actimetric tele-surveillance and tailored to the signal data compression.
In : ICOST'14, Lecture Notes in Comp. Sci., 8456, 59-70 (2015).

Chapitre 1

Recombinaison V(D)J

Sommaire

1.1 Introduction et Problématique

Nous allons étudier essentiellement les réarrangements V(D)J, qui sont des réarrangements de l'ADN spécifiques aux cellules immunitaires thymiques, dans la période pré- et post-natale.
Avant d'évoquer la notion de modélisation en immunologie, je ferai un bref rappel du système immunitaire. Dans ce chapitre, nous avons étudié plusieurs aspects de recombinaison. L'étude a été focalisée sur les axes suivants :

— Le modèle de recombinaison de Thuderoz-Simonet 2010 a été ré-analysé à la lumière des derniers résultats expérimentaux

— Ce modèle étant validé en rétro-prédiction chez l'homme, nous avons ensuite recherché sa trace, dans la variabilité des gènes recombinés encore présents en périphérie dans les lymphocytes T

— Ce modèle a été ensuite raffiné, en tenant compte de l'existence et de la valeur de paramètres tels que vitesse d'ouverture, largeur des fenêtres de recombinaison„ chez l'homme et chez la souris. Nous avons ensuite comparé les versions validées dans ces deux espèces.

— Les mécanismes de régulation (en particulier par microARNs), qui permettent de contrôler la concentration de la principale protéine RAG, impliquée dans le mécanisme de recombinaison, seront ensuite comparés dans ces deux espèces.

1.2 Le système immunitaire

L'organisme humain (composé 10^{13} à 10^{14} cellules de plus de 300 types, ainsi que de 10^{14} à 10^{15} bactéries), a besoin d'un "pilote" qui veille au bon fonctionnement entre ces différents éléments afin de garantir la meilleure santé possible : il s'agit du système immunitaire. Ce système est composé d'organes, de cellules et de molécules qui interviennent dans les réactions immunitaires.

1.2.1 Les organes du système immunitaire

Ces organes sont dispersés dans tout le corps et comprennent des organes centraux et des organes périphériques.

1.2.1.1 Les organes lymphoïdes centraux

Il s'agit de la moelle osseuse, qui produit les lymphocytes B, et du thymus, qui produit les lymphocytes T. Elle représente le lieu de maturation, appelée lymphopoïèse. Au cours de leur développement dans les organes lymphoïdes centraux, les lymphocytes se différencient et vont acquérir leur compétence. Les lymphocytes acquièrent des marqueurs de surface spécifiques (CD19 pour les lymphocytes B et CD3 pour les lymphocytes T), ainsi qu'un récepteur de spécificité (TCR ou BCR). A l'issue de leur maturation, les lymphocytes sélectionnés sont libérés dans la circulation sanguine.

1.2.1.2 Les organes et tissus lymphoïdes périphériques

Les organes lymphoïdes périphériques comprennent les ganglions lymphatiques et la rate, qui jouent le même rôle dans les réponses immunitaires. Ils sont colonisés par les lymphocytes immunocompétents produits dans les organes centraux. Les organes lymphoïdes secondaires assurent une partie du renouvellement des lymphocytes au cours des divisions cellulaires qui sont déclenchées par la reconnaissance de l'antigène et ont pour but d'amplifier la réponse immunitaire, une fois qu'elle a été initiée.

1.2.2 Les molécules du système immunitaire

Les cellules du système immunitaire communiquent entre elles en sécrétant toute une série de molécules.

1.2.2.1 Les anticorps

Les anticorps ou immunoglobulines sont produits par des cellules du système immunitaire (les lymphocytes B). Ce sont des molécules symétriques, composées de quatre chaînes polypeptidiques : deux chaînes lourdes et deux chaînes légères). Il existe 5 types de chaînes lourdes (γ, α, μ, δ et ϵ) qui définissent respectivement les 5 classes des Ig (IgC, IgA, IgM, IgD et IgE) et deux types de chaînes légères (kappa et lamda). Les gènes codant pour ces chaînes se sont localisés respectivement au niveau des chromosomes 2 et 22, chez l'homme, et ces chaînes peuvent se combiner avec n'importe quel type de chaîne lourde (gène localisé au niveau de chromosome 14). Ces anticorps ont ainsi la propriété ainsi de reconnaître de manière spécifique une molécule étrangère à l'organisme.

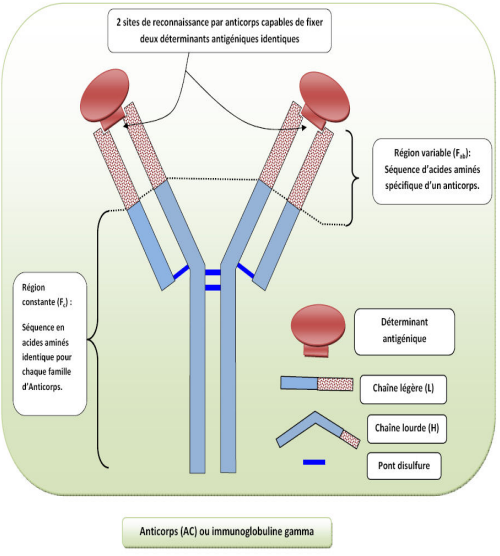


FIGURE 1.1 – Représentation d'un anticorps.

1.2.2.2 Les autres molécules

On peut citer aussi :

1. Les molécules d'adhésion Les molécules d'adhésion cellulaire interviennent dans la migration, l'activation et les fonctions effectrices des lymphocytes. Elles appartiennent à diverses classes moléculaires : les sélectines, les intégrines...

2. Les cytokines sont des substances solubles de signalisation cellulaire. Elles sont constituées des interlkeukines, ainsi que des interférons.

3. Le système du complément est composé d'un ensemble de protéines plasmatiques, ainsi que de protéines membranaires à fonctions régulatrices.

1.2.3 Les cellules du système immunitaire

L'essentiel des cellules du système immunitaire provient de cellules hématopoïétiques. Le graphique suivant nous montre les différentes étapes de leur différenciation.

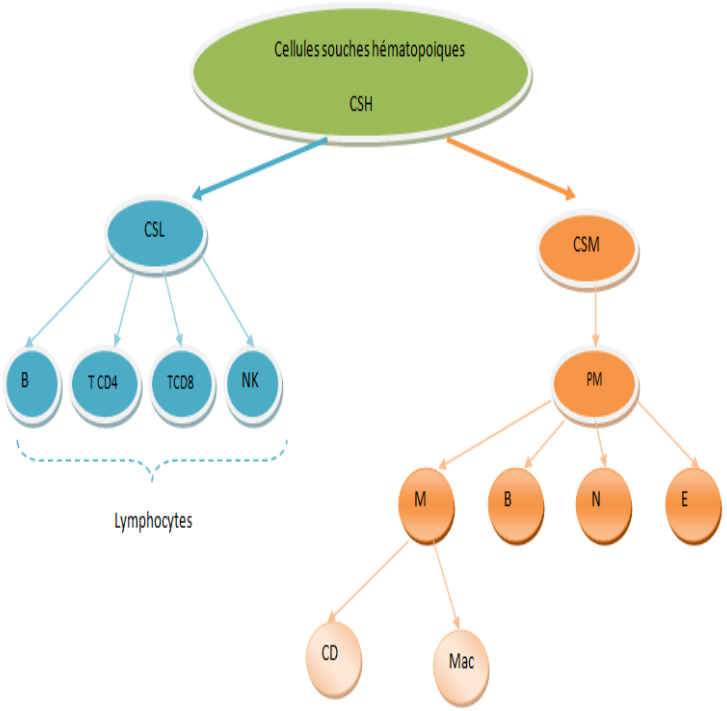


FIGURE 1.2 – Les cellules Souches Lymphoïdes **(CSL)** et les Cellules Souches Myéloïdes **(CSM)** sont issues des Cellules Souches Hématopoïétiques**(CSH)**. A partir des CSL, sont générés les lymphocytes B **(B)**,les lymphocytes **T CD4** ou **TCD8** et les cellules NK **(NK)**. A partir des CSM, sont générés les neutrophiles **(N)**, éosinophiles **(E)**, basophiles **(B)** et les monocytes **(M)**, qui donnent lieu aux macrophages **(Mac)** et aux cellules dentriques **(CD)**, par le biais du mécanisme de différenciation.

Les cellules immunitaires sont les globules blancs ou leucocytes. Elles déclenchent une réaction inflammatoire, quand elles rencontrent un élément étranger. On distingue trois types de globules blancs. Les phagocytes comprennent les monocytes, les macrophages, les polynucléaires (neutrophiles, éosinophiles et basophiles), ainsi que les cellules dendritiques. Les mastocytes représentent une variété des leucocytes. Ils interviennent principalement dans les réactions d'allergie, grâce à l'expression des récepteurs membranaires. A côté des basophiles, avec lesquels ils ont une ressemblance physiologique, mais une différence morphologique, ils sont les réacteurs de l'hypersensibilité immédiate.

1.2.3.1 Les cellules NK

Les cellules NK sont des cellules caractérisées par leur capacité à lyser des cellules tumorales ou infectées, d'où leur nom " Natural Killer". Elles représentent en moyenne 7.5 % des lymphocytes dans le sang. On les retrouve aussi dans les organes lymphoïdes secondaires, ainsi que dans certains organes, comme le foie ou les poumons. Contrairement aux autres types de lymphocytes (T et B, voir dans le paragraphe suivant), elles ne prolifèrent pas, d'où la mise en place rapide de leur fonction, grâce à un équilibre entre les signaux activateurs et inhibiteurs.

1.2.3.2 Les lymphocytes B et T

Les lymphocytes sont les seules cellules équipées des récepteurs spécifiques d'antigène. Ils peuvent être de type B ou T. Bien qu'il soit impossible de distinguer morphologiquement un lymphocyte T d'un lymphocyte B, leurs fonctions, ainsi que leurs phénotypes, sont extrêmement hétérogènes. La différence primordiale consiste dans le fait que les lymphocytes T agissent directement en cas de danger, ou autrement dit de pénétration des corps étrangers, tandis que les lymphocytes B agissent par des molécules appelées anticorps. Ces cellules sont réparties selon le type de réponse immunitaire.

1.2.4 Les types de réponse

L'organisme possède deux systèmes de défense : l'immunité innée et l'immunité adaptative (Tableau 1).

CRII	Cellules à l'interface entre les deux systèmes	CRIA
• Les phagocytes • Les cellules NK • Les mastocytes • Les cellules résidentes	• cellule NKT • Lymphocyte T γ	• Lymphocyte B • Lymphocyte T

Table 1.1 – Distribution des cellules du système immunitaire selon le type de réponse (Cellules Réponse Immunitaire Innée (CRII)et Cellules Réponse Immunitaire Adaptée (CRIA)).
Par la suite, nous décrivons ces 2 types de réponse.

1.2.4.1 La réponse immunitaire innée

Plusieurs termes désignent cette réponse comme naturelle, non spécifique, immédiate ou réponse sans mémoire.Elle représente la première ligne de défense et de protection. Elle existe chez tous les animaux pluricellulaires. Pour accomplir sa tâche de défense, le système inné repose sur différentes barrières :
• barrières physiques ou anatomiques : peau, muqueuses...
• barrières biologiques / biochimiques : pH, enzymes...
• barrière écologique : flore bactérienne

1.2.4.2 La réponse immunitaire adaptative

Lorsqu'un virus ou bactérie a réussi à franchir la première ligne de défense, une deuxième se met en action : c'est la réponse immunitaire adaptative, qui est appelée aussi réponse spécifique (elle est spécifique d'un antigène donné) ou acquise ou "pas immédiate". On ne la trouve que chez les vertébrés. Il en existe deux types : immunité humorale (gérée par les lymphocytes B) et immunité à médiation cellulaire (gérée par les lymphocytes T). Contrairement à la première, cette seconde immunité est caractérisée par un effet de mémoire. Elle fait intervenir des lymphocytes provenant d'un immense répertoire (10^{12} récepteurs de cellules B et 10^{15} récepteurs de cellules T). Chaque cellule T ou B subit des recombinaisons génomiques aléatoires. Leurs récepteurs antigéniques font l'objet d'une recombinaison somatique et aléatoire de l'ADN, appelée recombinaison V(D)J, qui fait l'objet de notre étude.

1.3 La régulation des réarrangements chez l'homme

1.3.1 Principe des réarrangements

Le mécanisme donne naissance à un gène, dans lequel différents fragments de génome sont réunis. On peut citer la recombinaison VJ ou la recombinaison V(D)J. Pour cette recombinaison, les segments suivants sont mis en jeu :

1.3.1.1 Les segments impliqués

— V : Il s'agit du segment "variable" du gène.
— J : Il s'agit du segment "jonction" du gène.
— D : Il s'agit du segment "diversité" du gène.

A partir de ces différents segments, on a une grande diversité combinatoire, car une même chaîne lourde peut s'associer à plusieurs chaînes légères (Tableau N°2). Suite à l'assemblage aléatoire, on obtient un répertoire immense. Le nombre de combinaisons des chaînes est estimé à plus de 1 million. Il est calculé de la manière suivante (41*23*6)*((41*5)+(33*5)).

segment	ENSCL	ENSCK	ENSCA
V	41	41	33
D	23		
J	6	5	5

Table 1.2 – Diversité combinatoire des chaînes d'anticorps chez l'homme

Dans notre exemple, nous avons utilisé 10 segments de J (TRAJ) et 11 segments de V (TRAV).

1.3.1.2 Les étapes

La recombinaison se déroule selon plusieurs étapes :

1) Reconnaissance de la séquence SSR
2) Clivage entre la séquence codante et la SSR
3) Formation d'extrémités en épingle à cheveux pour les régions V et J
4) Ligation des extrémités signal
5) Clivage des structures en épingle à cheveux
6) Extrémités sortantes et origine des nucléotides palindromiques
7) Ligation des segments V et J de chaîne légère
8) Elagage par exonucléase
9) Addition des nucléotides N
10) Ligation et réparation des gènes de chaîne lourde

1.3.2 Les facteurs impliqués

Plusieurs facteurs sont impliqués dans le mécanisme étudié (Tableau N°3).

facteur	rôle
• complexe RAG1 et RAG2	catalyse le clivage de l'ADN et la formation des récepteurs
• HMGB1 et HMGB2	stabilise la liaison de RAG aux SSR
• DNA-PK	résolution du réarrangement
• XRCC4 et Cernunnos	activation de l'ADN ligase IV
• Tdt	ajoute les nucléotides N aux exons
• ADN polymérases	insertion des nucléotides ajoutés par Tdt
• La ligase IV	ligation des extrémités d'ADN
• ATM et p53	régulation de l'expression des RAG

Table 1.3 – Liste des facteurs impliqués dans la recombinaison, ainsi que leur rôle.

1.3.3 Rôle des SSR

La recombinaison est catalysée par des enzymes qui sont aussi impliqués dans la réparation d'ADN. Elle est ciblée sur les sites appropriés, grâce aux motifs appelés Séquence Signal de Recombinaison (SSR). Cette séquence est constituée de deux blocs de séquence (Fig °3) : un nonamère et un heptamère. Ces deux séquences sont espacées, soit par 12 pb, soit par 23 pb (correspondant à un ou deux tours de la double hélice). Ces espaceurs placent les séquences heptamère et nonamère du même côté de la molécule d'ADN : un 12RSS reconnaîtra préférentiellement un 23RSS à 2 tours d'hélice, évitant ainsi des réarrangements inutiles de type [V-V] ou [J-J]. Cette règle est appelée règle 12/23.

FIGURE 1.3 – séquence nucléotidique des SSR

1.4 Modélisation en immunologie

Un phénomène biologique peut être traduit par une équation mathématique : d'où la modélisation. Il peut être aussi étudié par une approche "bio-informatique". Cette dernière méthode consiste à utiliser une interface informatique, afin de tester les différentes hypothèses. Une interface a été développée avec le langage Python, au sein du laboratoire, variante d'une interface développée pour étudier les recombinaisons chez la souris (Simonet 2010). Elle permet de rendre compte des différentes recombinaisons possibles ou, autrement dit, des "mariages" entre des segments de différents gènes, en partant d'un modèle mathématique.

1.5 Résultats et conclusions

Nous avons expliqué, par une extension du modèle, les données thymiques obtenues chez le nouveau-né humain. Cependant, ces données humaines sont en faible nombre et l'étude statistique sera à valider sur des échantillons thymiques humains plus importants. Cela permettra de confirmer le fait que le mécanisme de recombinaison V(D)J est différent chez l'homme et chez la souris, en particulier que les fenêtres d'ouverture de la portion V(D)J permettant la recombinaison sont de moins grande ampleur chez la souris que chez l'homme.

Numerical Model for the Vα-Jα Gene Use in Human TRA/TRD locus: Recombination Dynamical Rules

October 9, 2015

Hana Hazgui* , Florence Thuderoz* , Olivier Hansen*, Maria-Ana Simonet*, J.F.Dyon***, Patrice Noël Marche**, Evelyne Jouvin-Marche**, Jacques Demongeot*.

Abstract

The V(D)J recombination is the process by which the variable exons encoding the antigen recognition sites of receptors expressed on B and T lymphocytes are generated during early development via somatic assembly of gene segments. Concerning T-Cell receptor alpha-chains, V and J genes are used from the inside out of the TRA locus during successive rearrangements, with no allelic exclusion. Beside extensive quantifications of V-J associations from human thymic genomic DNA, a model approach is proposed. Comparison between model and experimental results enhances knowledge about kinetics and dynamical rules controlling human V-J segment use. Predictions are made about parameters not accessible through experiments, like the number of successive rearrangements (4 or 5) or the successive accessibility window sizes and locations. Finally, the precise progression of gene accessibility to rearrangements, according to non- constant opening speeds, together with a synchronized opening of the J regions between both alleles, are sufficient to fully explain both the experimental V-J frequencies currently available and the interallelic J usage. Model and experimental results provide a coherent representation of Vα-Jα combinatorial repertoire. Comparison with previous studies led on mouse draws a fine conservation of Vα-Jα rearrangement dynamics between both species.

1 Introduction

Infectious agent multiplicity and dysfunctional host cells, like tumor cells, pushed jawed vertebrates to develop mechanisms permitting the production of an extensive variety of antigenic receptors (Rast et al, 1997). In humans and rodents, T cells, responsible of cell mediated immune response, mostly express clonotypic alpha beta T Receptors (TR) on their surface. The alpha and beta protein chains are coded by loci being non-functional in the germ-line configuration: the human TRA locus consists of 54 TRAV genes belonging to 41 subgroups, 61 TRAJ segments localized on 71 kb, and a unique TRAC gene; TRB locus at 7q35 spans 620 kb. It consists of 64-67 TRBV genes belonging to 32 subgroups and encompasses Diversity

[0]Corresponding author: Jacques.Demongeot@agim.eu
*Team AGIM, Jean-Raoul Scherrer Laboratory, UniGe UJF, Faculté de Médecine, University Joseph Fourier, 38700 La Tronche, France.
**U823 Institut Albert Bonniot, INSERM-UJF, University Joseph Fourier, 38700 La Tronche, France.
***Chirurgie infantile, University Hospital CHU Grenoble, 38700 La Tronche, France.

(D) genes as well. Several combinations of V(D)J genes are generated in developing lymphocytes by means of somatic site-specific DNA rearrangements (Bassing et al, 2002). RAG enzymes work as a multi-subunit complex to induce cleavage of a single double stranded DNA (dsDNA) molecule between the antigen receptor coding segment and a flanking recombination signal sequence (RSS). RSSs are composed of seven conserved nucleotides (a heptamer) that reside next to the gene encoding sequence followed by a spacer (containing either 12 or 23 unconserved nucleotides) followed by a conserved nonamer. Only a pair of dissimilar spacer RSSs are efficiently recombined, referred to as the 12/23 rule. The RSSs are present on the 3 of a V region and the 5 of the J region. RAG-induced DNA breaks are repaired by ubiquitously expressed nonhomologous end-joining proteins, forming precise signal end joints and imprecise coding end joints (Bassing et al ., 2002). Due to the non frank coding joints, a maximum of 1/3 of the performed rearrangements produce in-frame rearrangements, which conserve the proper translation reading frame of the gene (Coleclough et al, 1983). A last factor of TRα repertoire diversity consists in heterodimer pairing. If this step is successful, after surface expression of the TR, the T cells will undergo thymic selection, which guarantees only functional and non-autoreactive T cells will migrate to the periphery. At last, V(D)J rearrangements constitute a somatic neo-gene creation giving the organism the possibility to build a vast immune repertoire diversity. Although profiles of TRα repertoire are well established, knowledge related to TR alpha chain diversity remains restricted because few anti-VAD antibodies are available. Vα to Jα rearrangements take place during the CD4 + CD8 + Double-Positive (DP) stage of T cell intrathymic development. The TRA locus is known to go through multiple rearrangement rounds (Petrie et al, 1993), without strict allelic exclusion (Krangel, 2009). When a Vα-Jα in-frame rearrangement is achieved in a cell, pairing with the beta chain is assayed, leading to surface expression and positive selection that are both needed to allow the TR-mediated signals stopping further rearrangements. Vα and Jα genes are used sequentially from inside the locus toward distal genes. First rearrangements use Jα genes proximal to the Vα region (5) and Vα genes proximal to the Jα region (3); successive secondary rearrangements implicate progressively more distal Jα and more distal Vα genes (Thompson et al,1990; Jouvin-Marche et al, 1998; Aude- Garcia et al, 2001; Pasqual et al, 2002; Krangel et al, 2004). This inside out use depends on Cis regulating elements conserved between mouse and human: enhancers and promoters constrain accessibility over Jα and Vα regions. The TRα enhancer (Ealpha), located at the 3 end of the C gene, firstly activates two promoters that control the use of proximal J genes: T early alpha promoter (TEA) and J49 (Villey et al, 1996; Hawwari et al, 2005). The structure of the Jα region (TRAJ) is remarkably conserved between human and mouse genomes: 61 J genes (49 functional) are spread over a 71Kb J region and 60 J genes (44 functional) are spread over 64Kb respectively (see www.imgt.org for locus representation). The human Vα region includes 54 V genes (44-46 functional) spread over 800 kb, whereas in mouse, several duplications during evolution formed a longer V region (1400kb) encompassing 70 to more than 100 V genes depending on the haplotype (Gahéry-Ségard et al, 1996). The random V-J association model, taking into account non-functional genes, gave 2156- 2254 V-J association possibilities for the human species. This value was reduced to roughly 2000 associations because of the inside out use of the TRA locus that leads to improbable associations of proximal V to distal J genes and of distal V to proximal J genes (Fuschiotti et al, 2007; Jouvin-Marche et al, 2009). Within these 2000 associations, knowledge about each individual V-J association frequency remains to debate; V-J association frequencies constitute a further limitation of the combinatorial diversity that reduces the available TRα repertoire.

After recent advances in the field of TRA recombination modeling (Warmflash et al,

2006; Thuderoz et al, 2010), the present study addresses precisely Jα and Vα gene use in the human TRA locus on the basis of both experimental and model approaches. A number of 110 Vα-Jα associations were quantified in the human thymus using genomic real time PCRs. The corresponding 11 Vα and 10 Jα genes examined were chosen regularly placed along the Vα and Jα regions in order to offer a first experimental sampling of the human TRα computational repertoire in the thymus. Our successive windowing model (Figure 1), accounting for the Vα to Jα rearrangement process in mouse (Thuderoz et al, 2010) was adapted to the human locus structure particularities. Human model simulation results were compared with our experimental V-J quantifications and with interallelic J usage (Davodeau et al, 2001); the quality of the fit tells a specific and non constant progression of gene accessibility to rearrangements over the V and J regions along with a synchronized opening of the J region between both alleles is sufficient to fully explain V-J combinatorial repertoire features. Model and experiment results provide a complete representation of Vα-Jα combinatorial repertoire, which displays unbalanced V-J frequencies. The study showed that the dynamic rules and kinetics governing Vα to Jα rearrangements are practically conserved between human and mouse species. Interestingly, the model approach allows making some predictions about parameters, which cannot be determined through experimentation, like the number of successive rearrangements or the size and position of successive DNA accessible windows. Among other results, model studies predict the human locus undergoes 4-5 successive rearrangement rounds versus 3-4 rearrangements rounds in mouse. A supplementary rearrangement would give the human species one additional chance per allele to generate an in-frame rearrangement, which constitutes a major selective advantage compared to the mouse, compensating for the lower number of human V genes compared to that in mouse (cf. Figure 1).

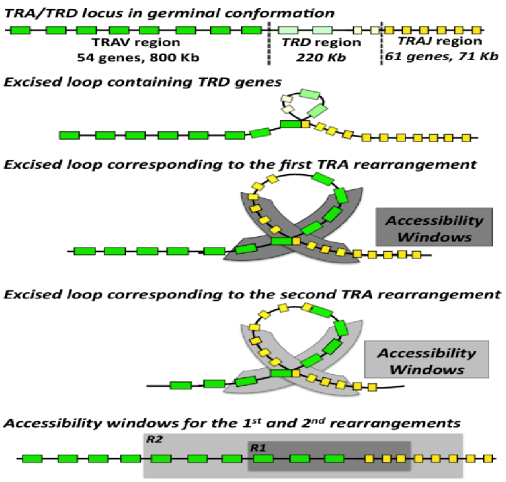


Figure 1: Successive windowing mechanism. After a first maturation allowing the excision of TRD genes, accessibility windows are calculated for each rearrangement, progressing over the V and J regions according to an inside out manner. Only the schemes corresponding to the first and second rounds of TRA rearrangements are represented and the last diagram provides a distinctive representation of the window progression.

2 Material and Methods

2.1 Experimental approach: quantification of Vα-Jα associations from human thymus extracts

2.1.1 Gene Nomenclature

Gene names correspond to the ImMunoGeneTics (IMGT) nomenclature (http://imgt.cines.fr). NCBI (National Center for Biotechnology Information) accession numbers are AE000658-AE000662 for the V region and M94081 for the J region. Positions of the V and J genes in the TRA/TRD locus were determined from the first nucleotide of the TRAC gene as previously described (Baum et al, 2004).

2.1.2 Provenance of the samples

The experimental dataset presented in this issue was generated from genomic DNA extracted from thymi of 4 healthy children, between 1 month and 1 year old. The thymi were from surgical waste and the committee approving the experiments, and checking that informed consent was obtained from all subjects, has been the same than those noticed in the previously published paper (Fuschiotti et al, 2007).

2.1.3 Selection of V J segments

We quantified the rearrangements for 110 associations. The V and J segments were selected for the following reasons :

1. relation to those analyzed in mice (for comparison)
2. location at the windows of accessibility
3. localization in the areas of hot spot of recombination
4. conservation of the RSS, kept or almost identical to the consensus sequence.

2.1.4 Selection and design of primers

The Vα and Jα primers were selected in order to globally cover the V and J regions and were spaced regularly between 50 to 60 kb and 5 to 10 kb for the V and J regions respectively. The specificity of each primer was checked using the BLAST sequence alignment program (Altschul et al, 2009). The selection criteria for the primers were :

1. The Recombination Signal Sequence (RSS) of the V or J genes were quite identical to the consensus sequence
2. The optimal amplification temperatures were in the same range of order for all the primers, allowing the use of identical amplification conditions for all the PCRs
3. The amplification yield of each couple of primer was 90%, thus allowing a direct relative comparison
4. Each couple of primers gave only one band on agarose gel and size of each amplification product was similar, allowing comparing PCR results.

2.1.5 PCR conditions

The polymerase Chain Reaction (PCR) is a powerful and sensitive technique of DNA amplification (40 cycles). Real-Time PCR had become the gold-standard method for accurate quantification in Light Cycler (Roche Diagnostics,Meylan, France). The table 1 contains the conditions of quantitative PCR.

Step	Time	Temperature
Denaturation	10mn	95°
Denaturation	15sc	94°
Hybridation	7sc	67°
Elongation	7sc	72°

Table 1: Steps of PCR method at several conditions.

2.1.6 Standardization of the experimental quantifications

A number of 110 V-J combinations were quantified for each of the 4 human thymic samples among a total of 2250-2350 potential associations (considering that 45 to 47 V genes and 50 J genes are functional). The quantifications of the V-J associations (Fig.4) were standardized among the PCR experiences according to the house keepinggenes quantification values, used as a normalizer :

$$\boxed{\text{V-J quantif. = Ct(housekeeping gene amplification product)-Ct(V-J association),}}$$

where Ct is the Cycle threshold i.e., the cycle at which the sample reaches the threshold line (level of detection). In order to be plotted on the model surface (Fig. 3), averages of PCR results were expressed as 1.8 Ct, due to the rate of amplification efficiency of the PCRs (90%):

$$\boxed{1{,}8^{\,V-Jquanti}}$$

2.2 Modeling approach: description of the stochastic model

2.2.1 Opening mechanism

The sequential windowing model we developed is a stochastic model consisting in two windows of accessibility progressing on the TRA/TRD locus over Vα and Jα regions. On the Figure 7, showing the diagram flow of the simulation program, the arrows represent the windows of accessibility moving over the V region from proximal V (3) to distal V genes (5) and over the J region from proximal J (5) to distal J genes (3). For the first rearrangement round, the status of the V and J region accessibility window is calculated from the V and J opening speed parameters for both alleles. For the subsequent rearrangement rounds, the new sizes and positions of the V and J region accessibility windows result from the opening mechanism and from the DNA sequence(s) deleted by the previous rearrangement(s). The lengths of windows of accessible DNA over the V and J regions, LVk and LJk verify the equations:

$$\boxed{LV_1 = S_V(t_0 + \tau_1, LJ_1 = S(t_0 + \tau_1) \text{and, for k} \geq 2, \text{LV}_k = LV_{k-1} + S_V\tau_k, LJ_k = LJ_{k-1} + S_j\tau_k,}$$

where k refers to the rearrangement round number. The opening offset time t0 denotes a minimal time of opening and τ_k ($k \geq 1$) are random variables uniformly distributed between t0 and the end of the opening process. The V and J opening speed parameters, consist in varying parameters; their sensitivity was studied by simulations. The sequential windowing model was first developed in mouse. In this species, the V and J opening speed parameters were determined from the time of first detection of rearranged genes during ontogeny. They were described as non-constant speeds varying within an interval. Based on this experimental result, we used non-constant speeds as well for the opening speed parameters in the model developed for human TRA rearrangements. The quality of the fit of model results with experimental V-J profiles allowed evaluating the opening speeds values in human. The opening duration before each rearrangement is a constant parameter whose value was determined through simulations as well, by varying values between 2 hours and 50 hours.

2.2.2 Accessible genes and genes to rearrange

In the model, the position and size of the accessibility windows at the very time of a rearrangement event directly determine the sequential V and J genes being accessible for this rearrangement based on their physical positions. At each rearrangement event and for each allele the model settles on a number of sequential V genes and J genes considered as accessible genes for the rearrangement. For each rearrangement round k, and for each such accessible Vik and Jjk genes, Boolean variables named BVik and BJjk were defined: they equal to 1 if the gene is accessible, and to 0 if it is non-accessible or deleted during a previous rearrangement of order less than k. Among the accessible genes, one V gene and one J gene are chosen according to the probability distribution given by their RSS scores. The RSS scores (KVi and KJj) represent, for each Vi or Jj genes, the homology percentage of the gene sequence compared to the consensus sequence. The RSS scores take a value between 0 and 1, the 1 value corresponding to a fully consensus RSS. The RSS score we determined from homology percentages fell in agreement with the functional/pseudo gene status of all the V or J genes in human (Glusman et al, 2001; Lefranc et al, 2009; see www.imgt.org for RSS sequences). According to its non-functional status, a pseudo recombination gene is never found rearranged and consequently the corresponding RSS score is assimilated to zero in simulations. We denote by FVk (resp. FJk) the cumulative distribution function of the relative window length obtained after the kth rearrangement by adding the BVmk (resp. BJmk) variables weighted by their score KVm:

$$\begin{aligned} \text{FV}_k(i) &= \textstyle\sum_{m=1,\ldots,i} KV_m BV_{mk} / (\sum_{m=1,\ldots,104} KV_m BV_{mk}) \\ \text{FJ}_k(j) &= \textstyle\sum_{m=1,\ldots,j} KV_m BV_{mk} / (\sum_{m=1,\ldots,60} KJ_m BJ_{mk}) \end{aligned}$$

At step k, we use the cumulative distribution functions FVk and (resp. FJk) corresponding to the random variables RVk (resp. RJk) uniform on [0,1] and we calculate a number NVk (resp.NJk) equal to:

$$\boxed{inf(FV_k^{-1}(RV_k)), \text{ and respectively } inf(FJ_k^{-1}(RJ_k)),}$$

where NVk and NJk correspond to the V and J genes to rearrange at a step k.

2.2.3 In-frame/out-of-frame rearrangements

The conservation of the proper reading frame of the genes is equiprobably determined. The in-frame rearrangement frequency parameter (if) gives the probability to perform a productive (in-frame) rearrangement. The frequency of in-frame rearrangements is known to be at maximum of 1/3 (Coleclough et al, 1983). This upper limit was used in simulations. Effects of lower values for the if parameter were tested by simulations as well.

2.2.4 Selection

When an in-frame rearrangement is generated on one or on both alleles, it/they undergo selection. The selection coefficient, s, included in our model refers to a global selection coefficient corresponding to the successive steps of selection: alpha pairing, surface expression efficiency and thymic selection. The sensibility of the s parameter was studied by simulations. We consider s as a constant parameter applied to every in-frame rearrangement. This homogenous re-shaping of the T cell repertoire gives statistically an evaluation of the frequency of each V-J association and was already successfully used in the Warmflash model (Warmflash et al, 2006). In fact, the distinction between CD4+ and CD8+ subsets is not supported in the experimental data.

2.2.5 Subsequent rearrangement rounds

If an out-of-frame rearrangement is selected on both alleles, another rearrangement round occurs in the program. The new sizes and positions of the V and J region windows are calculated as a result of the opening mechanism and of the DNA sequences deleted by the previous rearrangement. The successive rearrangement mechanism continues until an in-frame rearrangement is selected or until k, the index of rearrangement round Rk, reaches the number maximal of successive rearrangements, Rmax. Different values of Rmax were tested by simulations.

2.2.6 Stop of the successive rearrangement mechanism

Once an in-frame rearrangement is selected, it will block the successive rearrangement mechanism within the cell. A parameter named Stopdelay is defined which takes integer values. If Stopdelay=0, the rearrangements will stop within the same rearrangement round as for the other allele. For example, if an in-frame rearrangement is performed at R1 on the allele A and successfully selected (Fig.1), it will block rearrangements on the other allele B at R1, and the data corresponding to the R1 rearrangement round for allele B (i.e., Vi and Jj genes implicated, and the in-frame or out-of-frame status of the new gene built from this (Vi, Jj) rearrangement) will be stored in the simulation results. If Stopdelay=1, the rearrangements will stop within the subsequent rearrangement round for the other allele B, namely at $R2 = R1 + 1$. More generally, we have:

$$\boxed{\mathrm{R}_{kalleleB} = R_{kalleleA} + Stop_{delay}}$$

If the first in-frame rearrangement being selected occurs over the B allele locus, we have:

$$\boxed{\mathrm{R}_{kalleleA} = R_{kalleleB} + Stop_{delay}}$$

The effect of the Stopdelay parameter value on simulated results was studied using different parameter values and in a second time, using Stopdelay=0 and Stopdelay=1 inside different fractions of the entire simulated population. Experimental interallelic distance distributions in the Jα usage were employed to set the intervals of variation of the parameter values (Davodeau et al, 2001).

3 Results

3.1 Experimental Results

Figure 2 presents relative measures of 110 V-J associations determined by quantitative genomic PCR analysis from 4 human thymic samples. Corresponding 11 V and 10 J genes appear ordered according to their physical position over the locus. Amplification of G3PDH gene was used as a normalizer and results were expressed in arbitrary units, indicating the differences in cycle numbers at which PCR products were first detected. Average values and standard deviations are indicated for each V-J association and a color code illustrates V-J average levels. Results are consistent among the individual samples, with a maximum standard deviation value of 2,61 and an average standard deviation of 0,73. In detail, V41 and V40 genes, located next to the J region (3 end of the translated strand), rearranged predominantly with J genes from J61 to J24. The V38, V35, V30, and V 21 genes, located in the central part of the V region, mainly rearranged with J genes from J61 to J10. V16 used from J56 to J10 segments and eventually, V2 gene, situated in the distal part of the V region (3 end), rearranged mostly with distal J genes (3 end of the J region), from J33 to J5. All in all, Figure 2 clearly demonstrates a higher abundance of associations implicating V and J genes symmetrically placed in the germinal configuration of the TRA locus, namely proximal V to proximal J, centered V to centered J, and distal V to distal J genes. Finally, the random variable Mi equal to the number of rearrangements involving a given pair Vi-Ji can follow either a log-normal or a normal but we do not have a sufficient data base for testing it, hence we simulate the model successively with these two hypotheses and we discuss the differences between the corresponding two set of results..

PCR results		TRAJ61		TRAJ58		TRAJ53		TRAJ48		TRAJ41		TRAJ33		TRAJ24		TRAJ16		TRAJ10		TRAJ5	
TRAV41	Sample 1	6.2	μ	7.1	μ	7.3	μ	7.5	μ	6	μ	6.3	μ	3.3	μ	0	μ	0	μ	0	μ
	Sample 2	6.3	6.6	6.5	6.83	7.35	7.13	6.3	7.33	5.95	6.14	6.9	6.09	2.55	2.1	0	0	0	0	0	0
	Sample 3	5.25	σ	6.2	σ	6.3	σ	7.5	σ	6.1	σ	5	σ	1.1	σ	0	σ	0	σ	0	σ
	Sample 4	8.65	1.45	7.5	0.59	7.55	0.56	8	0.72	6.5	0.25	6.15	0.79	1.45	1.01	0	0	0	0	0	0
TRAV40	Sample 1	7.75	μ	6.3	μ	7.4	μ	7.65	μ	7.2	μ	6.65	μ	3.15	μ	0	μ	0	μ	0	μ
	Sample 2	7.45	7.25	6.1	6.06	6.65	6.51	8.25	7.83	5.35	6	5.95	5.8	4.35	4.86	0	0	0	0	0	0
	Sample 3	5.4	σ	5.3	σ	5	σ	7.4	σ .	5.25	σ	4.3	σ	5.45	σ	0	σ	0	σ	0	σ
	Sample 4	8.4	1.3	6.55	0.54	7	1.05	8	0.38	6.2	0.91	6.3	1.04	6.5	1.44	0	0	0	0	0	0
TRAV38	Sample 1	8	μ	9.5	μ	9	μ	10	μ	8.35	μ	8	μ	6.5	μ	6.5	μ	7	μ	4	μ
	Sample 2	7.4	7.18	7.7	8.3	8	8.25	8	8.25	6.8	7.29	7	7.08	5	6.13	5	6.08	5.8	6.33	0	1.5
	Sample 3	6	σ	8	σ	8	σ	8	σ	7	σ	6	σ	6	σ	6.5	σ	7	σ	2	σ
	Sample 4	7.3	0.84	8	0.81	8	0.5	7	1.26	7	0.71	7.3	0.83	7	0.85	6.3	0.72	5.5	0.79	0	1.91
TRAV35	Sample 1	5.1	μ	6.45	μ	8.7	μ	6.8	μ	6.75	μ	6.48	μ	6.55	μ	6.8	μ	4.65	μ	0	μ
	Sample 2	4.5	5.65	5.9	6.34	8.1	8.05	6.35	6.79	6.15	6.3	5.85	5.82	4.9	6.78	5.2	6.13	3.35	4.25	0	0.11
	Sample 3	6	σ	6	σ	7	σ	7	σ	6	σ	4.5	σ	8.25	σ	5.7	σ	4	σ	0	σ
	Sample 4	7	1.09	7	0.5	8.4	0.74	7	0.31	6.3	0.32	6.45	0.93	7.4	1.43	6.8	0.81	5	0.73	0.45	0.23
TRAV30	Sample 1	3.75	μ	4.75	μ	6.8	μ	6.8	μ	5.4	μ	5.1	μ	5.8	μ	6.1	μ	5.4	μ	2.75	μ
	Sample 2	4.15	4.43	4.15	5.05	5.9	6.11	6.9	6.03	4.05	5.14	3.6	4.9	5.05	5.56	4.8	5.7	4.3	5.2	0.7	3.86
	Sample 3	4.3	σ	5.3	σ	5.5	σ	5.4	σ	5.5	σ	5.5	σ	6.4	σ	6.3	σ	6.1	σ	6	σ
	Sample 4	5.5	0.75	6	0.79	6.25	0.55	5	0.97	5.6	0.73	5.4	0.88	5	0.67	5.6	0.67	5	0.75	6	2.61
TRAV21	Sample 1	6.6	μ	7.2	μ	7	μ	7.5	μ	6.5	μ	6.6	μ	6.1	μ	6.7	μ	5.9	μ	5.1	μ
	Sample 2	6.2	5.79	6.9	6.53	6.95	6.36	6.7	6.05	6.15	5.78	5.8	5.44	5.4	5.21	5.7	5.93	5.05	5.61	3.65	4.81
	Sample 3	5	σ	7	σ	6.5	σ	5	σ	6	σ	5	σ	5	σ	6.2	σ	5	σ	4	σ
	Sample 4	5.35	0.74	5	1.02	5	0.94	5	1.26	4.45	0.91	4.35	0.98	4.35	0.73	5.1	0.68	6.5	0.72	6.5	1.28
TRAV16	Sample 1	4.8	μ	6.2	μ	7.1	μ	6.4	μ	5.3	μ	5.5	μ	6.2	μ	6.8	μ	6.9	μ	3.2	μ
	Sample 2	6.35	4.93	6.8	5.75	7.5	6.76	8.1	6.13	6.4	5.73	6.05	5.76	6.35	6.75	7.1	6.98	7.1	7.84	3.5	3.3
	Sample 3	4	σ	4.5	σ	6	σ	5	σ	6.2	σ	6.1	σ	7	σ	7.7	σ	8	σ	3	σ
	Sample 4	4.55	1.01	5.5	0.99	6.45	0.67	5	1.47	5	0.68	5.4	0.36	7.45	0.58	6.3	0.59	9.35	1.12	3.5	0.24
TRAV5	Sample 1	0	μ	0	μ	0	μ	1.4	μ	4.8	μ	6	μ	6.5	μ	5.2	μ	4.4	μ	5.1	μ
	Sample 2	0	0	0	0	0	0	1	2.7	4.5	4.88	5.8	5.9	6.7	6.38	5.3	5.6	4.1	5.38	4.8	5.28
	Sample 3	0	σ	0	σ	0	σ	4.4	σ	6	σ	5.5	σ	7	σ	5.4	σ	6.7	σ	4.9	σ
	Sample 4	0	0	0	0	0	0	4	1.75	4.2	0.79	6.3	0.34	5.3	0.75	6.5	0.61	6.3	1.31	6.3	0.69
TRAV3	Sample 1	0	μ	0	μ	0	μ	1.2	μ	4.3	μ	6	μ	5.7	μ	6	μ	4.8	μ	5.7	μ
	Sample 2	0	0	0	0	0	0	1.5	2.23	4.2	4.18	5.9	5.88	4.5	5.1	5.6	5.53	4.4	5.18	5.3	5.58
	Sample 3	0	σ	0	σ	0	σ	2.2	σ	3	σ	5.1	σ	5.7	σ	4.5	σ	5.5	σ	5.8	σ
	Sample 4	0	0	0	0	0	0	4	1.26	5.2	0.9	6.5	0.58	4.5	0.69	6	0.71	6	0.71	5.5	0.22
TRAV2	Sample 1	0	μ	0	μ	0	μ	0	μ	3.15	μ	3.75	μ	6.45	μ	6.2	μ	5	μ	3.9	μ
	Sample 2	0	0	0	0	0	0.7	0	0.75	4.4	3.76	4.2	4.74	6.7	6.83	5.6	5.73	5.3	6.63	4	5.43
	Sample 3	0	σ	0	σ	2	σ	2	σ	4	σ	4	σ	6.25	σ	5.1	σ	8.4	σ	6	σ
	Sample 4	0	0	0	0	0.79	0.94	1	0.96	3.5	0.55	7	1.52	7.9	0.74	6	0.49	7.8	1.73	7.8	1.86
TRAV1	Sample 1	0	μ	0	μ	0	μ	0	μ	2.75	μ	3.75	μ	5.1	μ	6.3	μ	5.3	μ	6.2	μ
	Sample 2	0	0	0	0	0	0	0	0.57	3.4	3.26	4.2	4.54	4.8	5.63	6	5.58	5.7	5.25	5.8	7.48
	Sample 3	0	σ	0	σ	0	σ	0.8	σ	3.2	σ	5	σ	5.8	σ	5.3	σ	5.4	σ	9.9	σ
	Sample 4	0	0	0	0	0	0	1.50	0.72	3.70	0.40	5.20	0.68	6.80	0.89	4.70	0.72	4.60	0.47	8.00	1.88

Table 2: Real time PCR quantifications of 110 V-J associations from 4 human thymic samples. Results are expressed in arbitrary units indicating the difference in cycle number at which products were first detected. Amplifications of the G3DPH gene were used as normalizer. Quantifications were performed from 4 thymic genomic DNA samples corresponding to 4 children aged between 1 month and 1 year. .

3.2 Model Approach

3.2.1 Model interface and generated graphs

The aim of the simulation program based on the successive windowing model was to present a suitable interface (Fig.3) in order to test different scenarios for V-J recombination mechanisms. Simulation results are displayed on a 3-dimensional histogram of V-J rearrangements representing the whole TRα chain combinatorial repertoire. The programs offer as well a graphic representation designed to plot the J region use by some V genes by clicking on the V gene axis directly on an interactive graph. For distinctive usages, an Excelalpha file is generated giving direct access to the simulated results: frequencies FVi-Jj of all the Vi-Jj associations, details about the associations generated through each rearrangement round, couple of J genes used at both allelic loci for each simulated cell and, as a reminder, the entire values of parameters corresponding to the simulation.

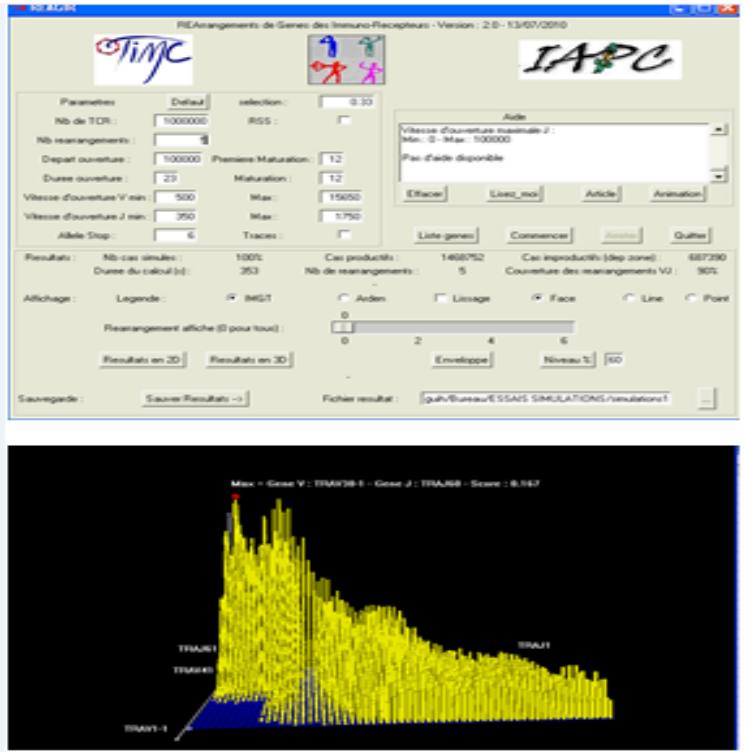


Figure 2: Program interface (values used for simulations) and generated graphs.

3.2.2 Parameter values, robustness of the human model, and comparison with a priori study conducted on mouse

The stochastic successive windowing model was firstly developed for the Vα-Jα use during mouse TRA locus successive rearrangements. For this species, the ontogeny days when V and J genes were first seen rearranged in conjunction with physical gene positions and gave speeds of progression of the gene accessibility to rearrangements (Pasqual, 2002; Thuderoz, 2010). These speeds, denoted V and J region opening speeds, were successfully used in mouse model as simulation results and fitted thymic quantifications from experiments only by using two opening speeds chosen within intervals closed to the experimental 99.9 % confidence intervals. In fact, the V speed (denoted SV) belonged to the speed interval [0.35 Kb/h, 34

Kb/h] and the J speed (SJ) to [0.4 Kb/h, 1.55 Kb/h] with a mean opening speed of roughly 18 Kb/h for the V region and 1 Kb/h for the J region. In the transition to human, structural specificities of the human TRA/TRD locus were integrated to the model. Concerning opening speeds, values of the J region opening speed interval determined from experience in mouse were used in the human model in order to be consistent with the high conservation of both the J region structure and the J gene sequences between the two species (Uenishi et al, 2003). On the other hand, the V opening speed interval was determined by simulations: the interval [0.5 Kb/h, 15.65 Kb/h] (average 8 Kb/h) provided simulation results that fitted the best human experimental distributions (least squares method). The opening location of the simulation was fixed between the V and J genes in order to access directly to the TRAV and TRAJ genes after the first maturation, which was set to allow the elimination of the TRD genes in the model (Fig. 1). For humans, as well as in mice, issues obtained from the modeling that fit the best the experimental data, indicated the duration of the first maturation step had a mean value of 5 hours and the opening duration before each rearrangement was 24 hours. Interestingly, parametric study gave distinct numbers of successive rearrangements between the two species with 3-4 for mouse and 4-5 for human. To make certain the sampling size used in simulations was large enough, the representativeness of the repertoire was tested by making sets of simulations of increasing size. Diversity became constant when the population size was higher than 5×105 T cells, showing the pertinence of a repertoire calculation based on a 106 alpha chain population (Arstila et al, 1999). Tests of robustness performed on the model indicated that variations of about 5 to 10% in the values of the parameters provided simulation results statistically coherent with experimental data, though larger variations induced major deviations on the modeling simulation results inconsistent with (i.e., significantly different from) experimental data.

3.2.3 Stochastic successive windowing model results

In order to compare simulation results to experiments, the very 110 V-J associations experimentally examined were extracted from the simulated results computed for the totality of the V-J association frequencies in the framework of the successive windowing model. Results from simulations were transformed. The values are calculated using the number of simulated recombinations (n_{Vi-Jj}) as follows :

$$\boxed{\mathrm{F}_{ViJj} = c * n_{ViJj}, c = \sum(\log 1{,}8(m))/\sum(n_{ViJj}).}$$

Where, m= mean of 4 human thymus values.
Table 3 showed simulated results generated using the set of parameters fully described above; the Table 2 indicates that results obtained with 4 and 5 successive rearrangements (R=4 and R=5) for the two simulated populations fit experimental results. Numbers in bold indicated that simulated values lie within the two standard deviations confidence interval of the experimental PCR quantifications (under Gausian hypothesis for the distrubition of experimental data), corresponding to the fact that the alpha risk of rejection of the hypothesis of a difference between predicted and observed levels of gene rearrangements is more than 0.25. At last, the consistency between simulated data and frequencies determined from thymic genomic DNA validates our model as a relevant tool accounting for the dynamical building of the TRα combinatorial repertoire in human thymus.

(a: R=4)

genes	TRAJ61	TRAJ58	TRAJ53	TRAJ48	TRAJ41	TRAJ33	TRAJ24	TRAJ16	TRAJ10	TRAJ5
TRAV41	59,67	73,60	74,01	60,70	22,87	0,41	0,00	0,00	0,00	0,00
TRAV40	92,31	111,49	108,49	89,03	35,03	1,64	0,00	0,00	0,00	0,00
TRAV38	71,89	119,00	130,68	117,57	42,60	10,72	2,32	0,41	0,00	0,00
TRAV35	91,56	108,15	120,98	121,60	77,49	52,23	25,88	5,94	6,14	1,84
TRAV30	53,53	71,89	72,44	68,75	56,26	36,25	27,38	12,15	5,39	6,62
TRAV21	65,82	82,48	78,04	99,07	61,58	57,56	34,27	25,33	16,45	14,34
TRAV16	16,80	17,75	18,98	39,53	40,97	37,96	26,90	13,18	13,72	10,24
TRAV5	0,00	0,00	0,00	8,94	9,35	20,69	16,52	16,66	14,41	18,43
TRAV3	0,00	0,00	0,00	4,64	4,44	14,20	18,64	18,64	20,55	19,12
TRAV2	0,00	0,00	0,00	1,16	3,89	15,57	15,16	16,11	17,41	16,52
TRAV1	0,00	0,00	0,00	0,00	1,43	21,23	20,28	23,01	20,35	17,27

(b:R=5)

genes	TRAJ61	TRAJ58	TRAJ53	TRAJ48	TRAJ41	TRAJ33	TRAJ24	TRAJ16	TRAJ10	TRAJ5
TRAV41	52,90	78,13	67,95	63,41	25,43	1,58	0,00	0,00	0,00	0,00
TRAV40	83,78	111,38	116,24	90,16	30,56	0,85	0,00	0,00	0,00	0,00
TRAV38	89,43	114,53	110,72	94,95	42,45	12,81	2,69	0,00	0,39	0,00
TRAV35	86,61	124,52	119,07	125,51	87,13	56,84	26,42	7,36	4,07	4,73
TRAV30	59,60	65,71	64,00	84,64	46,79	43,57	26,74	13,27	6,97	8,87
TRAV21	63,02	70,31	72,35	96,66	69,65	56,97	40,15	20,44	22,80	24,18
TRAV16	14,98	17,48	17,87	33,51	38,51	40,41	21,88	16,95	11,37	12,88
TRAV5	0,00	0,00	0,00	6,11	9,13	20,83	21,62	15,51	16,10	17,15
TRAV3	0,00	0,00	0,00	4,14	4,93	22,01	23,98	14,72	21,88	16,82
TRAV2	0,00	0,00	0,00	1,84	2,30	13,87	14,13	20,17	16,62	13,93
TRAV1	0,00	0,00	0,00	0,00	1,12	16,36	16,89	20,76	18,73	25,76

Table 3: V-J association frequencies F_{Vi-Jj} from simulations converted in theoretical PCR cycle numbers (log1.8 (F_{Vi-Jj}). Parameter values used to generate these simulated results are fully presented in the Results section. The Figure indicates that results obtained with 4 and 5 successive rearrangements (R=4 and R=5) for the two simulated populations fit experimental results. Numbers in bold indicate that simulated values lie within the two standard deviation confidence interval of experimental PCR quantifications(under the Gaussian hypothesis for the distrubtion of experimental data), corresponding to the fact that the alpha risk of rejection of the hypothesis of a difference between predicted and observed levels of gene rearrangements $\geq 25\%$.

Comparing the experimental distribution of recombination to its simulated counterpart, we find that the trend of the distribution is the same. For example, if we take two V (TRAV1 and TRAV30), you can see this this trend :

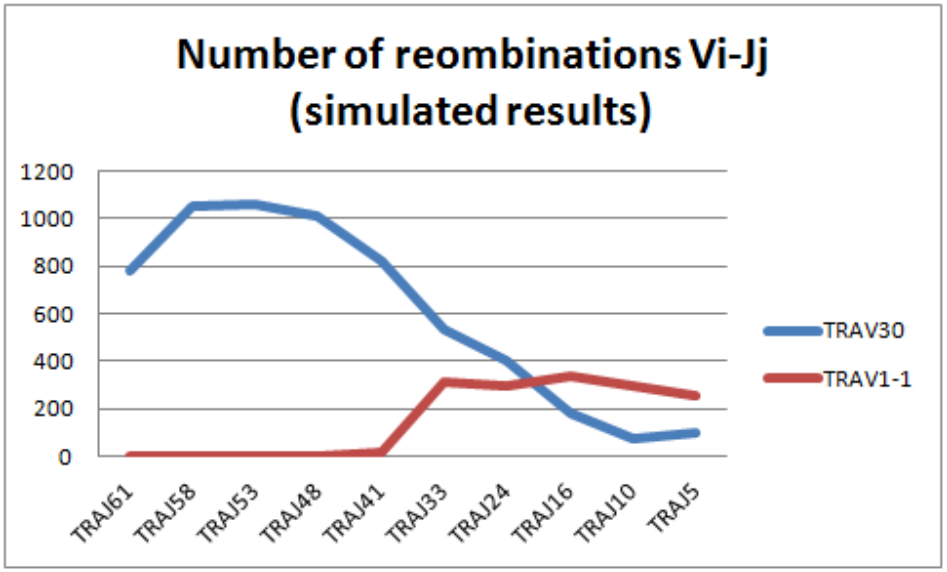


Figure 3: Distribution of some simulated recombinations of TRAV1 and TRAV30.

3.2.4 Experimental V J quantifications from human thymi

Simulation results were tested in parallel for their coherence with V-J quantifications and for their ability to fit an interallelic J distance distribution issued from the literature (Davodeau et al, 2001). The involvement of a parameter denoted Stopdelay within the model allowed computing the interallelic behavior at the very moment that TR-mediated signals stopped successive rearrangements in a given simulated cell (see MM section). Figure 5 shows that fixing Stopdelay=0 for all simulated cells gave an interallelic usage of J genes too close compared to experimental distribution (Fig. 5.A). On the contrary, a systematic stop of successive rearrangements within the subsequent rearrangement round (Stopdelay=1) gave an interallelic usage of J genes, which are too distant compared to the experimental distribution (Fig. 5.D). Simulations using Stopdelay=0 and Stopdelay=1 in proportions 25%/75% or 75%/25% of the cells gave two distributions that fit the experimental distribution (Fig. 5.B and 5.C). Beyond a satisfactory result concerning successive windowing model validation, the consequences of this fit regarding the putative interallelic behavior during the stop of the successive rearrangements were addressed in the Discussion section. Briefly, to complete these results, an identical analysis was conducted on mouse (additional Fig. 1). In this species, four studies addressing Jα usage at both TRA alleles available in the literature were compiled in order to get a more significant database (n=110), giving an empirical distribution of the interallelic distances in J usage statistically more confident than in the human case (n=29) (Davodeau et al, 2001; Heath et al, 1995; Malissen et al,1992; Casanova et al, 1991). The same conclusions emerge in mouse as in human, rejecting simulations performed using Stopdelay=0 or Stopdelay=1 for the whole cells (additional Fig. 1.A and 1.D). However, the more precise experimental distribution allows stating between the B and D graphs: simulations using Stopdelay=0 and Stopdelay=1 in the respective proportions of 75% and 25% fit the best the experimental distribution (additional Fig. 1.B)

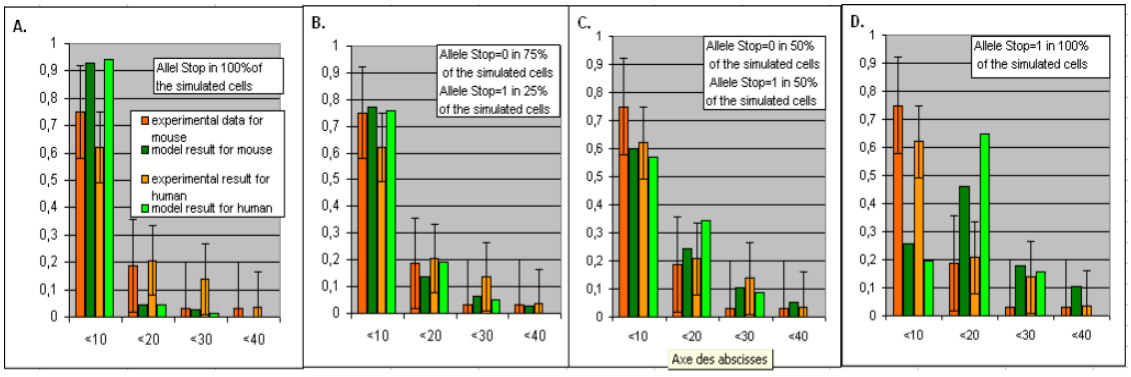


Figure 4: Experimental J interallelic distance distribution presented along with model results generated using different values for the Stopdelay parameter. The experimental distribution reproduced on the four graphs was taken from Figure 2 in (Davodeau, 2001), which presented the distribution of the differences in rank of Jα segments used at both alleles in several clones (authors denoted these values interallelic J distances). Error bars were added on the distribution for n=29; the formula (freq.*(1-freq.))/n was used for calculating each frequency (freq.) variance in histogram classes. The four graphs compare the experimental distribution with parametric study results generated varying the Stopdelay parameter value (described in MM). Comparison between experimental distribution and model simulation results showed that a stop of successive rearrangements within the same rearrangement round (Stopdelay=0) gave an interallelic usage of J genes too close compared to experimental distribution (A). On the contrary, a systematic stop of successive rearrangements within the subsequent rearrangement round (Stopdelay=1) gave interallelic usage of J genes too distant compared to the experimental distribution (D). Finally, parametric studies predict that when an in-frame rearrangement is successfully selected in a cell, the intracellular signaling will stop further rearrangements over the other allele during the same rearrangement round for 50% to 75% of the cells and during the subsequent rearrangement round for the rest of the cells (B or C).

3.2.5 Successive windowing model results fit interallelic distance for J segment use

Simulation results were tested in parallel for their coherence with V-J quantifications and for their ability to fit an interallelic distance distribution issued from the literature (Davodeau et al, 2001). The involvement of a parameter denoted Stopdelay within the model allowed computing the interallelic behavior at the very moment that TR-mediated signals stopped successive rearrangements in a given simulated cell (see MM section). Figure 5 shows that fixing Stopdelay=0 for all simulated cells gave an interallelic usage of J genes too close compared to experimental distribution (Fig. 5.A). On the contrary, a systematic stop of successive rearrangements within the subsequent rearrangement round (Stopdelay=1) gave an interallelic usage of J genes, which are too distant compared to the experimental distribution (Fig.5.D). Simulations using Stopdelay=0 and Stopdelay=1 in proportions 25%/75% or 75%/25% of the cells gave two distributions that fit the experimental distribution (Fig.5.B and 5.C). Beyond a satisfactory result concerning successive windowing model validation, the consequences of this fit regarding the putative interallelic behavior during the stop of the successive rearrangements were addressed in the Discussion section. Briefly, to complete these results, an identical analysis was conducted on mouse (additional Fig.1). In this species, four studies addressing Jα usage at both TRA alleles available in the literature were compiled in order to get a more significant database (n=110), giving an empirical distribution of the interallelic distances in J usage statistically more confident than in the human case (n=29) (Davodeau et al, 2001; Heath et al, 1995; Malissen et al, 1992; Casanova et al, 1991). The same conclusions emerge in mouse as in human, rejecting simulations performed using Stopdelay=0 or Stopdelay=1 for the whole cells (additional Fig.1.A and 1.D). However, the more precise experimental distribution allows stating between the B and D graphs: simulations using Stopdelay=0 and Stopdelay=1 in the respective proportions of 75% and 25% fit the best the experimental distribution (additional Fig.1.B).

4 Discussion

4.1 Human thymic TRα combinatorial repertoire from extensive real time PCR analyses

In spite of the human TRA gene polymorphism and the use of an elevated number of primers within the same PCR amplifications, the 110 V-J quantifications from thymic genomic DNA displayed a good inter-individual consistence among the 4 samples tested (maximal standard deviation over the entire experiments: 2.61; average standard deviation: 0.73) and the same qualitative speed of rearrangement over the V-J plane for both empirical distrubition (Gaussian and lognormal) of experimental PCR quantificatins (Tables 2 et 3). Quantitative PCR analysis results (Fig. 2) undoubtedly demonstrated each V gene tested used a subsequent subset of J genes according to their position over the locus, as we previously observed on a fewer number of V and J genes, using multiplex PCR analyses (Jouvin-Marche et al, 2009). It is worth to notice that the outside parts of these subsets formed two areas of non-observed or infrequent associations, corresponding to distal V / proximal J and proximal V / distal J (appearing in blue on tables 2 and 3, corresponding to the same experimental region of Fig.2) and that within each subset, frequencies differ highly. Finally, the corresponding 11 Vα and 10 Jα genes investigated, chosen regularly spread along the Vα and Jα regions, offered a first experimental wide-ranging sampling of the human TRα computational repertoire in the thymus.

4.2 Explanatory power of inside out gene use

If the inside out use of the genes was presumed to cause unequal frequencies of V-J gene associations (Krangel, 2009), the successive windowing model definitely demonstrated that a precise progression of gene accessibility to rearrangements together with an opening of the J region synchronized between both TRA alleles was sufficient to fully explain the experimental V-J frequencies currently available as well as the experimental interallelic J usage. Given the prominent conservation of J region structure between mouse and human species, the accessibility progression over the J region experimentally determined from ontogeny analyses in mice was used for the human model. This J region opening speed consisted in a variable speed (SJ) belonging to the [0.4 Kb/h, 1.55 Kb/h] variability interval, corresponding to a mean opening speed of about 1 Kb/h. Parametric studies performed for human models showed that this experimental J speed variability interval was the unique one allowing the generation of V-J association profiles fitting experimental data (a bias superior of 10% of the mean speed generated data incoherent with biological results), as it was already demonstrated for mouse (Thuderoz et al, 2010). Hence, model studies, performed for both the human and mouse species, predicted J regions, highly conserved in terms of genes number, length, and regulatory Cis elements, would proceed in a similar progression of accessibility through the successive rearrangement process. Concerning the V region accessibility progression, experiments performed during mouse ontogeny allowed the calculation of a mean speed SVmouse= 18Kb/h. Mouse model showed also the necessity to introduce a variability interval for the opening V speed equal to [0.35 Kb/h, 34 Kb/h] and centered on this mean value, in order to get distributions in accordance with experimental data (with an acceptable bias of 10% on the interval). Regarding the human V opening speed, major differences between mouse and human V regions avoided making preliminary assumptions and we have retained the opening speed variability interval [0.5 kb/h, 15.650 kb/h] after extensive parametric studies. The models developed for both species predict that the recombination centers (constituted over the J region throughout the successive V-J recombinations process)are the same in mouse and human V regions.
These distinct opening speed variability intervals, specific of the two species, seem to be proportional to the V region length, hence by the totality of the V region length, hence allowing to take advantage of the diversity offered by the totality of the V genes, independently of the length of the V region. In addition, simulations performed within the distinctive frameworks of human and mouse models showed the use of speeds variable inside an interval was important for the diversity of the repertoire generated. In fact, the use of constant opening speeds in simulations gave a narrower use of V-J associations, increasing considerably the two zones of improbable associations observed from experiments (depicted in blue cells on Fig2). Human model shows that progression of gene accessibility to rearrangements over TRAJ and TRAV regions (with human V speed different from the mouse one) constitutes a mechanism sufficient to explain the relative abundances of the V-J association frequencies determined from real time PCR assays on human thymic genomic DNA.

4.3 Stop of successive rearrangements: putative interallelic behavior

Regarding the J region, it is well accepted both J alleles perform a synchronized opening, which provokes an interallelic use of J segments separated by less than 10 genes for the majority of the lymphocytes (Villey et al, 1996). Nevertheless, minor occurrences of more distant interallelic J genes use are observed in experiments (Davodeau et al, 2001). When a Vα-Jα in-frame rearrangement is performed over an allele, beta pairing, surface expression

and positive selection are all needed to allow the TR-mediated signals stopping further rearrangements in the cell. As commonly accepted, the feedback loop due to positive selection of a TRα chain may logically inhibit further rearrangements on the very TRA allele that originated its synthesis. Concerning the other allele, simulations showed a stop of successive rearrangements within the same rearrangement round (Stopdelay=0) gave interallelic usage of J genes too close compared to experimental distribution (Fig. 5.A). On the contrary, a systematic stop of successive rearrangements within the subsequent rearrangement round (Stopdelay=1) gave interallelic usage of J genes too distant compared to the experimental distribution (Fig. 5.D). Hence, the involvement of the Allele Stop parameter inside the model suggests a putative scenario for the other allele behavior: in a majority of cells (supposedly from 50% to 75% of the cells, Fig.5.B and 5.C), the feedback loop would inhibit further rearrangements on this other allele within the same rearrangement round, but in the rest of the cells, a slight out of phase between two allele rearrangements would allow the other allele to perform an extra rearrangement round before TR-mediated signals step in and stop the mechanism. To complete this analysis, additional results in mouse, based on a larger experimental data base proposed the feedback loop would inhibit further rearrangements on the other allele during the same rearrangement round in the clear majority of the cells (75%, as on the additional Fig. 1.B), the extra rearrangement being performed only in a minority of cells (25%).

4.4 RSS: refining local frequencies on the repertoire shape

The human model included Recombination Signal Sequence (RSS) diversity effect through RSS scores, which were calculated from the percentage of homology of each RSS sequence with the consensus. The facultative use of RSS scores in simulations allowed observing their influence on results. RSS sequences appeared to change not the global repertoire shape, but only local specificities. This is in good accordance with mouse TRB locus observations that demonstrated V gene RSSs neither correlate with any specific restriction in use of the subset of J genes nor with any elevated V-J rearrangement frequencies (Wilson et al, 2001). Eventually, the RSSs would hold less impact in terms of gene use frequencies than in bi-directional use of the TRA genes during the successive rearrangements process.

4.5 Conclusions for the human TRα repertoire size

The potential TRα repertoire was first estimated to 0.5*106 chains in human blood through CDR3 heterogeneity analysis and considering that every V-J association was obtained from independent events consisting in tossing a couple of genes V and J with their marginal frequency (Arstila et al, 1999). In this estimation, every one of the 54 V genes was supposed to rearrange every one of the 61 J genes, giving a combinatorial diversity of 3294 associations. Identification of non-functional genes first limited this evaluation. In a previous work, experimental quantifications performed from blood and thymic material stated any one of the three V genes tested (V41, V40 and V1) rearranged not the entire J region but a subset of subsequent J genes, according to the V position in the locus. Consequently, combinations outside every J subsets, corresponding to non-observed or rare V-J associations, were presumed to lower the potential TRα combinatorial repertoire and a preliminary estimation of approximately 2000 associations in periphery and thymus was announced (Jouvin-Marche et al, 2009). In the present study, the J region use was tested through the use of ten J genes by eight V genes, using genomic DNA quantifications from human thymic samples. This further comprehensive experimental study defines more precisely the limits of the non-

observed or rare V-J association areas (appearing in blue cells on Fig.2). At a mechanistic level, the successive windowing model proposed a dynamical explanation for the occurrence of the proximal V / distal J and distal V / proximal J genes rare associations: they would originate from a non-synchronized placement. More, this stochastic model, validated for the human thymic V-J quantifications, allows calculating the V-J frequencies for the entire V-J associations. Knowledge about these frequencies allows proposing a new update on the potential TRα combinatorial repertoire evaluation. In fact, if about 2100 V-J associations would correspond to the totality of the possible combinations, some of these associations are considerably more abundant than others. Table 4 indicates the size of the most abundant V-J associations along with their frequency. If roughly 1700 associations would correspond to 95% of the alpha chains, about 1000 associations would correspond to 75%, and approximately, the 500 most abundant V-J associations would represent a half of the total. These observations point out the alpha chain combinatorial repertoire diversity supporting immunocompetence was highly over-estimated in thymic cells: the diversity may consist essentially in roughly 500 strongly represented V-J associations accounting for about the half of the V-J associations found in alpha chains, along with 1500 other less represented V-J associations.

Number of the most abundant V-J associations	Percentage of the chains corresponding to the most abundant V-J associations
2125	100%
1721	95%
954	75%
453	50%

Table 4: Number of the most abundant V-J associations along with the percentage of the alpha chains they represent. Table data were compiled using the entire V-J associations frequencies computed from the human successive windowing model.

4.6 Number of rearrangement rounds expected in human TRA loci

More than offering information on the alpha repertoire shape and diversity, the combined experimental and model approaches allow making predictions on the values of intrinsic parameters not accessible through experience like the very number of rearrangement rounds performed through the TRA rearrangement mechanism. Whereas the mouse model predicted a total of 3 to 4 successive rearrangements performed in mouse TRA locus (Thuderoz et al, 2010), in human however, the model parametric studies gave an estimation of 4 to 5 rearrangement rounds. Simulations showing the occurrence of four or five rearrangements would favor the use of the distal genes of V and J regions in human. In fact, the use of a fewer number of rearrangements in simulations gave an under-representation of distal J / distal V associations. At last, the incidence of a supplementary rearrangement would give the human species one additional chance per allele to generate an in-frame rearrangement, which may constitute a major selective advantage compared to the mouse, compensating for the lower number of human V genes compared to that in mouse.

5 Conclusion

The stochastic successive windowing model presented throughout this paper stands that the precise progression of gene accessibility to rearrangements according to non-constant opening speeds along with an interallelic synchronized J opening within each cell, constitutes a sufficient mechanism to explain experimental V-J frequencies and interallelic J usage. Eventually, the modeling step, using a multi-level systemic approach followed by a simulation phase, offered a clear understanding of the dynamics building of the human alpha repertoire, in order to propose predictions on this repertoire diversity richness and to dispose of a simulated theoretical repertoire showing the frequencies of the entire V-J associations. Knowledge about the human thymic repertoire shape (Figure 3) constitutes indeed a key issue in the immune system development, and thus a crucial requirement in therapeutic interventions aiming to reconstitute or to control immune responses.

References

[1] J. P. Rast, M. K. Anderson, S. J. Strong, C. Luer, R. T. Litman, and G. W. Litman, "α, β, γ, and δ t cell antigen receptor genes arose early in vertebrate phylogeny," *Immunity*, vol. 6, no. 1, pp. 1–11, 1997.

[2] C. H. Bassing, W. Swat, and F. W. Alt, "The mechanism and regulation of chromosomal v (d) j recombination," *Cell*, vol. 109, no. 2, pp. S45–S55, 2002.

[3] C. Aude-Garcia, M. Gallagher, P. N. Marche, and E. Jouvin-Marche, "Preferential adv-aj association during recombination in the mouse t-cell receptor alpha/delta locus," *Immunogenetics*, vol. 52, no. 3-4, pp. 224–230, 2001.

[4] S. F. Altschul, E. M. Gertz, R. Agarwala, A. A. Schäffer, and Y.-K. Yu, "Psi-blast pseudocounts and the minimum description length principle," *Nucleic acids research*, vol. 37, no. 3, pp. 815–824, 2009.

[5] T. P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky, "A direct estimate of the human $\alpha\beta$ t cell receptor diversity," *Science*, vol. 286, no. 5441, pp. 958–961, 1999.

[6] T. Azuma, V. Igras, E. B. Reilly, and H. N. Eisen, "Diversity at the variable-joining region boundary of lambda light chains has a pronounced effect on immunoglobulin ligand-binding activity," *Proceedings of the National Academy of Sciences*, vol. 81, no. 19, pp. 6139–6143, 1984.

[7] T.-P. Baum, N. Pasqual, F. Thuderoz, V. Hierle, D. Chaume, M.-P. Lefranc, E. Jouvin-Marche, P.-N. Marche, and J. Demongeot, "Imgt/geneinfo: enhancing v (d) j recombination database accessibility," *Nucleic acids research*, vol. 32, no. suppl 1, pp. D51–D54, 2004.

[8] K. Bleakley, M.-P. Lefranc, and G. Biau, "Recovering probabilities for nucleotide trimming processes for t cell receptor tra and trg vj junctions analyzed with imgt tools," *BMC bioinformatics*, vol. 9, no. 1, p. 408, 2008.

[9] C. Coleclough, "Chance, necessity and antibody gene dynamics." *Nature*, vol. 303, no. 5912, pp. 23–26, 1982.

[10] J.-L. Casanova, P. Romero, C. Widmann, P. Kourilsky, and J. Maryanski, "T cell receptor genes in a series of class i major histocompatibility complex-restricted cytotoxic t lymphocyte clones specific for a plasmodium berghei nonapeptide: implications for t cell allelic exclusion and antigen-specific repertoire." *The Journal of experimental medicine*, vol. 174, no. 6, pp. 1371–1383, 1991.

[11] F. Davodeau, M. Difilippantonio, E. Roldan, M. Malissen, J.-L. Casanova, C. Couedel, J.-F. Morcet, M. Merkenschlager, A. Nussenzweig, M. Bonneville *et al.*, "The tight interallelic positional coincidence that distinguishes t-cell receptor jα usage does not result from homologous chromosomal pairing during vαjα rearrangement," *The EMBO journal*, vol. 20, no. 17, pp. 4717–4729, 2001.

[12] P. Fuschiotti, N. Pasqual, V. Hierle, E. Borel, J. London, P. N. Marche, and E. Jouvin-Marche, "Analysis of the tcr α-chain rearrangement profile in human t lymphocytes," *Molecular immunology*, vol. 44, no. 13, pp. 3380–3388, 2007.

[13] H. Gahéry-Ségard, E. Jouvin-Marche, A. Six, C. Gris-Liebe, M. Malissen, B. Malissen, P.-A. Cazenave, and P. N. Marche, "Germline genomic structure of the b10. a mousetcra-v2 gene subfamily," *Immunogenetics*, vol. 44, no. 4, pp. 298–305, 1996.

[14] S. Gilfillan, A. Dierich, M. Lemeur, C. Benoist, and D. Mathis, "Mice lacking tdt: mature animals with an immature lymphocyte repertoire," *Science*, vol. 261, no. 5125, pp. 1175–1178, 1993.

[15] P. N. el Marche, "Differential chronology of tcradv2 gene use by> and p chains of the mouse tcr," *Eur. J. Immunol*, vol. 28, pp. 818–827, 1998.

[16] A. Hawwari, C. Bock, and M. S. Krangel, "Regulation of t cell receptor α gene assembly by a complex hierarchy of germline jα promoters," *Nature immunology*, vol. 6, no. 5, pp. 481–489, 2005.

[17] M. S. Krangel, "Mechanics of t cell receptor gene rearrangement," *Current opinion in immunology*, vol. 21, no. 2, pp. 133–139, 2009.

[18] M. S. Krangel, J. Carabana, I. Abbarategui, R. Schlimgen, and A. Hawwari, "Enforcing order within a complex locus: current perspectives on the control of v (d) j recombination at the murine t-cell receptor α/δ locus," *Immunological reviews*, vol. 200, no. 1, pp. 224–232, 2004.

[19] M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, J. Lane *et al.*, "Imgt®, the international immunogenetics information system®," *Nucleic acids research*, vol. 37, no. suppl 1, pp. D1006–D1012, 2009.

[20] E. Jouvin-Marche, P. Fuschiotti, and P. N. Marche, "Dynamic aspects of tcrα gene recombination: qualitative and quantitative assessments of the tcrα chain repertoire in man and mouse," in *V (D) J Recombination*. Springer, 2009, pp. 82–92.

[21] M. Malissen, J. Trucy, E. Jouvin-Marche, P.-A. Cazenave, R. Scollay, and B. Malissen, "Regulation of tcr α and β gene allelic exclusion during t-cell development," *Immunology today*, vol. 13, no. 8, pp. 315–322, 1992.

[22] M. A. Oettinger, D. G. Schatz, C. Gorka, and D. Baltimore, "Rag-1 and rag-2, adjacent genes that synergistically activate v (d) j recombination," *Science*, vol. 248, no. 4962, pp. 1517–1523, 1990.

[23] N. Pasqual, M. Gallagher, C. Aude-Garcia, M. Loiodice, F. Thuderoz, J. Demongeot, R. Ceredig, P. N. Marche, and E. Jouvin-Marche, "Quantitative and qualitative changes in vj α rearrangements during mouse thymocytes differentiation implication for a limited t cell receptor α chain repertoire," *The Journal of experimental medicine*, vol. 196, no. 9, pp. 1163–1174, 2002.

[24] H. Petrie, F. Livak, D. Schatz, A. Strasser, I. N. Crispe, and K. Shortman, "Multiple rearrangements in t cell receptor alpha chain genes maximize the production of useful thymocytes." *The Journal of experimental medicine*, vol. 178, no. 2, pp. 615–622, 1993.

[25] S. Takeshita, M. Toda, and H. Yamagishi, "Excision products of the t cell receptor gene support a progressive rearrangement model of the alpha/delta locus." *The EMBO journal*, vol. 8, no. 11, p. 3261, 1989.

[26] S. D. Thompson, J. Pelkonen, M. Rytkönen, J. Samaridis, and J. L. Hurwitz, "Nonrandom rearrangement of t cell receptor j alpha genes in bone marrow t cell differentiation cultures." *The Journal of Immunology*, vol. 144, no. 7, pp. 2829–2834, 1990.

[27] F. Thuderoz, M.-A. Simonet, O. Hansen, N. Pasqual, A. Dariz, T. P. Baum, V. Hierle, J. Demongeot, P. N. Marche, and E. Jouvin-Marche, "Numerical modelling of the vj combinations of the t cell receptor tra/trd locus," *PLoS Comput Biol*, vol. 6, no. 2, pp. e1 000 682–e1 000 682, 2010.

[28] H. Uenishi, H. Hiraiwa, R. Yamamoto, H. Yasue, Y. Takagaki, T. Shiina, E. Kikkawa, H. Inoko, and T. Awata, "Genomic structure around joining segments and constant regions of swine t-cell receptor α/δ (tra/trd) locus," *Immunology*, vol. 109, no. 4, pp. 515–526, 2003.

[29] I. Villey, D. Caillol, F. Selz, P. Ferrier, and J.-P. de Villartay, "Defect in rearrangement of the most 5? tcr–jα following targeted deletion of t early α (tea): Implications for tcr α locus accessibility," *Immunity*, vol. 5, no. 4, pp. 331–342, 1996.

[30] A. Warmflash and A. R. Dinner, "A model for tcr gene segment use," *The Journal of Immunology*, vol. 177, no. 6, pp. 3857–3864, 2006.

[31] A. Wilson, H. R. MacDonald, and F. Radtke, "Notch 1–deficient common lymphoid precursors adopt a b cell fate in the thymus," *The Journal of experimental medicine*, vol. 194, no. 7, pp. 1003–1012, 2001.

Chapitre 2

Réseau de régulation biologique

Sommaire

2.1 Introduction et Problématique

Un réseau biologique est une représentation de la circulation d'un certain type d'information dans la cellule. Il en existe plusieurs types :

1. réseau d'interaction protéine-protéine : la protéine A interagit physiquement avec la protéine B.
2. réseau de signalisation : la protéine A transmet un signal informatif à la protéine B.
3. réseau génétique ou de régulation : le gène A régule l'expression du gène B, par l'intermédiaire d'un produit direct ou indirect de son expression (protéine ou complexe ARN-peptide) servant de facteur de transcription ou de traduction du gène B .
4. réseau métabolique : il régule les réactions enzymatiques reliées à une fonction précise d'un organite, d'une cellule ou d'un tissu.

Nous nous intéressons ici à l'étude des réseaux de régulation génétique. Nous avons traité plusieurs réseaux, pour lesquels la question primordiale était : quels sont les gènes critiques, ainsi que les interactions qui sont susceptibles de changer la configuration du réseau ?

2.2 Modèles de régulation biologique

2.2.1 Présentation des modèles : réseau d'automates booléens à seuil

A la fin des années 60 et au début des années 70, S. Kauffman, puis R. Thomas, ont proposé une modélisation sous forme de réseau d'automates : chaque gène est modélisé par un automate, dont le but est de donner un sens « biologique » aux attracteurs de la dynamique de ces automates. Dans nos analyses, on s'intéresse aux réseaux booléens à seuil introduits par McCulloch (1943), dans le contexte des réseaux de neurones. Ces travaux ont été mis en valeur plus tard par Goles (1981) et par Hopfield (1982). Dans la suite, nous définissons la fonction de transition utilisée :

Définition 1 (Fonction de transition locale)**.**

$x_i(t+1) = H(\sum_{j=1}^{n} w_{ij} * x_j(t) - \theta_i)$, où $x_i(t)$ est l'état du gène i à l'itération t (égal à 1, si le gène i s'exprime, et à 0, sinon), avec
$j \in \{1, ..., n\}, w_{ij} \in \mathbb{R}, \theta_i \in \mathbb{R}$ et H est la fonction de Heaviside définie sur $\mathbb{R}$ et à valeurs dans {0,1} égale à 0 (resp. 1), si la variable dans H est négative ou nulle (resp. strictement positive).

Quelles informations portent les paramètres ?
La fonction de Heaviside est la fonction échelon ; elle vaut 1, si son argument est strictement positif, et 0, sinon. Elle est utilisée dans la modélisation de nos réseaux de régulation. Son argument est, dans ce cas, caractérisé par :

— w_{ij} : ce paramètre représente le poids d'influence que le gène j exerce sur le gène i. Il peut être positif s'il s'agit d'une activation et négatif dans le cas d'une inhibition.

— θ_i : ce paramètre représente le seuil d'activation pour un gène quelconque i.

2.2.2 Architecture d'un réseau de contrôle génétique

Un réseau de contrôle génétique possède, comme sommets, des gènes ou ARN de contrôle (micro-ARN, si-ARN, ARN circulaires,...) et, comme arcs, des interactions d'inhibition ou d'activation. Ces interactions sont exercées par des répresseurs ou des inducteurs de l'expression génique. Concernant son architecture, elle est caractérisée par une structure de graphe orienté, dont la matrice d'incidence est la matrice des signes des poids wij (ce signe est égal à +1, si $w_{ij} > 0$, à -1, si $w_{ij} < 0$ et à 0 si $w_{ij} = 0$). Ce graphe comporte trois parties :

1. Une arborescence initiale représentée par des arbres sources (**« up-trees »**), partant des gènes sources qui n'ont pas d'antécédents dans le réseau.
2. Une arborescence terminale (**« down-trees »**) représentée par des arbres puits, aboutissant à des gènes puits qui n'ont pas de successeurs dans le réseau.
3. Entre les deux, des composantes fortement connexes, c'est-à-dire connexes et maximales : connexes, car comportant un chemin d'interaction reliant les sommets de

tout couple de leurs sommets. Ils comportent donc en particulier des circuits (chemins d'interaction fermés, dont les deux extrémités sont identiques, chacun de leurs sommets appartient à un circuit) ; et maximales, car non contenues dans une composante connexe strictement plus grande. Elles sont faites d'un circuit isolé, ou de grappes de circuits tangents (un seul gène commun) ou intersectés (plus d'un gène commun), pouvant être une arborescence de grappes dépendantes.

La question qui se pose : qu'est ce qu'un circuit ?
Avant de définir un circuit en terme de théorie de graphes, on doit répondre à la question : qu'est ce qu'un graphe ?

Définition 2 (Définition d'un graphe)**.**

Un graphe est défini par un couple G= (Sommet, Ensemble sommet) tel que :
-Sommet est un ensemble fini de sommets noté S.
-Ensemble sommet : est un ensemble de couples de sommets $(s_i, s_j) \in S^2$ noté E.

Un graphe peut être orienté ou non. Par la suite, nous définissons ces deux types.

Définition 3 (Un graphe non orienté)**.**

Les couples $(s_i, s_j) \in E$ sont appelés arêtes et sont représentés de la manière suivante $s_i - s_j$. Ils ne sont pas orientés, c'est-à-dire que : $(s_i, s_j) \Leftrightarrow (s_j, s_i)$.

Définition 4 (Un graphe orienté)**.**

Les couples $(s_i, s_j) \in E$ sont appelés arcs et sont représentés de la manière suivante $s_i \rightarrow s_j$. Ils sont orientés.

La différence vient du fait que les couples (s_i, s_j) sont orientés ou pas. Dans notre cas, on s'intéresse aux graphes orientés. On peut définir un circuit de la manière suivante :

Définition 5 (Un circuit)**.**

Un circuit est le chemin d'un graphe orienté dont les sommets de départ et de fin sont les mêmes.

Dans le cas d'un réseau de régulation biologique, on utilise la notion de circuit, afin de définir deux types de circuits :

— C+ : circuit ayant un nombre pair d'arcs négatifs.

— C- : circuit ayant un nombre impair d'arcs négatifs.

Les circuits sont utilisés par la suite, afin d'identifier les attracteurs, qui seront détaillés dans la section suivante. Un circuit est défini aussi bien par sa longueur que par son signe (voir le graphique ci-dessous).

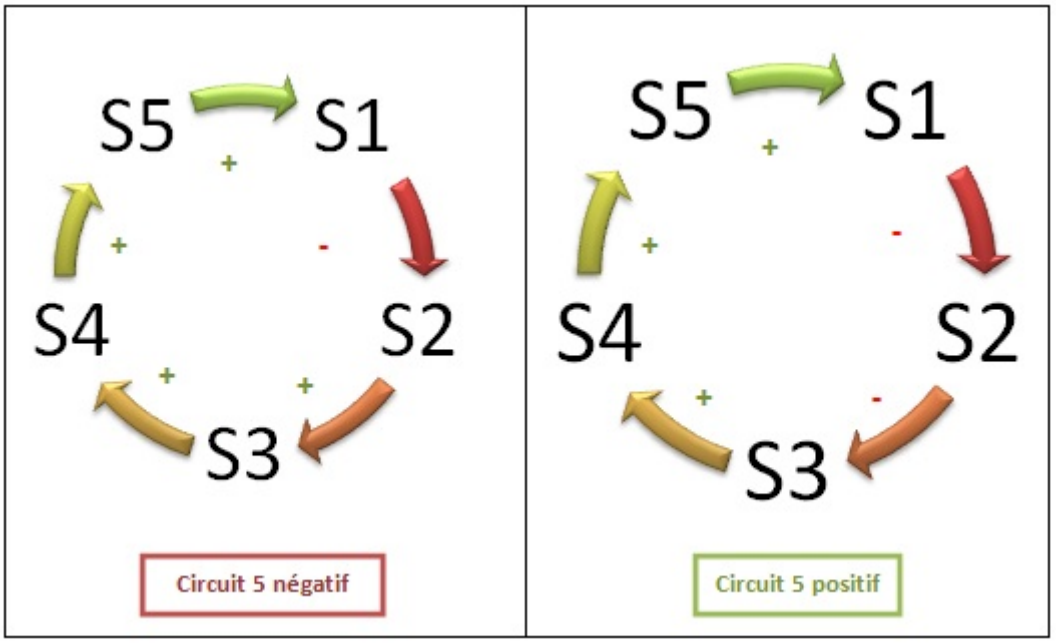


FIGURE 2.1 – Notion de circuit

Une boucle est un circuit du longueur égale à 1.

2.2.3 Les attracteurs

2.2.3.1 Définition

Le nombre total des points fixes et des cycles limites définit le nombre d'attracteurs d'un réseau.
La définition mathématique utilisée pour les calculs est, par exemple, celle définie par Demongeot et Cosnard en 1985 : nous appellerons configuration tout point de l'ensemble d'état E={0,1}n du réseau de n automates correspondant au réseau de régulation de n gènes étudié. Soit x(0) une configuration initiale et x(t) la configuration obtenue après t itérations du réseau. Chaque itération peut-être définie par une mise à jour parallèle (tous les sommets du réseau sont mis à jour en même temps, par application de la règle de transition ci-dessus), séquentielle (les sommets sont mis à jour l'un après l'autre, dans une séquence déterminée) ou bloc-séquentielle (des blocs, composés de sommets itérés parallèlement, étant itérés de manière séquentielle). Soit A une partie de E et soit B(A) (bassin de A) l'ensemble des configurations initiales x, telles que l'ensemble L(x) des configurations visitées une infinité de fois lors de leurs itérations successives d'ordre t (t tendant vers l'infini), soit dans A. Alors, A est un attracteur si et seulement si :
i) L(B(A) = A
ii) il n'existe pas d'ensemble strictement plus grand que A qui vérifie i).
Il est à noter que cette définition est valable aussi bien pour des itérations discrètes que pour des système à temps continus (à une sophistication près concernant les trajectoires « fantômes » au sens de R. Bowen). Dans le cas discret, une configuration a de A vérifie : il existe un nombre minimum d'itérations p(a), tel que a soit obtenu de nouveau par application à a de p(a) itérations du réseau. Ce nombre est appelé la période de a.

2.2.3.2 Types

Les attracteurs sont divisés en deux types :
-un point fixe : c'est un attracteur dont la période est égale à 1.
-un cycle limite : c'est un attracteur dont la période est supérieure ou égale à 2.

Les attracteurs sont caractérisés par la taille de leurs bassins d'attraction, qui résume le nombre des configurations qui évoluent vers l'attracteur. L'ensemble {attracteur, bassin d'attraction} forme le paysage dynamique du réseau. Le graphe d'interaction forme l'architecture du réseau.

2.2.3.3 Règle de calcul

Selon une règle développée dans la thèse de Mathilde Noual 2012, on peut calculer le nombre d'attracteurs d'un circuit. Selon le signe du circuit, on obtient :

ℓ \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	2	2	3	3	5	5	8	10	15	19	31	41	64
2	1	1	2	3	3	4	5	8	10	14	19	31	41	63
3	1	2	1	3	3	6	5	8	8	15	19	33	41	64
4	1	1	2	1	3	4	5	11	10	14	19	24	41	63
5	1	2	2	3	1	5	5	8	10	26	19	31	41	64
6	1	1	1	3	3	1	5	8	8	14	19	63	41	63
7	1	2	2	3	3	5	1	8	10	15	19	31	41	158
8	1	1	2	1	3	4	5	1	10	14	19	24	41	63
9	1	2	1	3	3	6	5	8	1	15	19	33	41	64
10	1	1	2	3	1	4	5	8	10	1	19	31	41	63
11	1	2	2	3	3	5	5	8	10	15	1	31	41	64
12	1	1	1	1	3	1	5	11	8	14	19	1	41	63
13	1	2	2	3	3	5	5	8	10	15	19	31	1	64
14	1	1	2	3	3	4	1	8	10	14	19	31	41	1

a.

ℓ \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	–	–	–	–	–	–	–	–	–	–	–	–	–
2	1	1	–	–	–	–	–	–	–	–	–	–	–	–
3	1	1	2	–	–	–	–	–	–	–	–	–	–	–
4	1	2	1	2	–	–	–	–	–	–	–	–	–	–
5	2	1	2	2	4	–	–	–	–	–	–	–	–	–
6	1	1	3	3	2	6	–	–	–	–	–	–	–	–
7	2	2	3	2	4	3	10	–	–	–	–	–	–	–
8	2	3	2	8	3	4	6	16	–	–	–	–	–	–
9	3	2	2	3	5	9	7	7	30	–	–	–	–	–
10	2	4	3	4	17	7	7	10	11	52	–	–	–	–
11	4	3	5	6	7	7	11	11	16	19	94	–	–	–
12	3	4	9	2	7	42	11	33	17	23	28	172	–	–
13	5	6	7	7	11	11	16	19	24	28	39	46	316	–
14	6	7	7	10	11	17	105	23	28	38	46	60	75	586

b.

FIGURE 2.2 – Les tableaux ci-dessus, extraits de Demongeot et al. 2012 (J. DEMONGEOT, M. NOUAL and S. SENE, Combinatorics of Boolean automata circuits dynamics, Discrete Applied Mathematics, 160, 398-415 (2012)) donnent le nombre d'attracteurs de couples de circuits tangents, dans deux cas : au-dessus un des deux circuits est positif de longueur r, l'autre étant négatif de longueur l ; en-dessous, les deux circuits sont négatifs, de longueur respective r et l. Dans les deux cas, le nombre d'attracteurs du couple de circuits tangents est donné à l'intersection de la colonne r et de la ligne l.

2.3 Résultats expérimentaux

Plusieurs réseaux ont été étudiés, mais nous allons présenter seulement les réseaux « immunetworks » et le réseau « biliary atresia » . Les autres sont détaillés dans les articles ci-joints.

2.3.1 Exemple 1 : Les « Immunetworks »

Les réseaux de régulation génétique impliqués dans le contrôle du système immunitaire sont appelés « immunetworks ». Ils sont caractérisés par leur comportement dynamique asymptotique, défini par leurs attracteurs (Demongeot et al., 2011). Dans le cas des «

immunetworks », le nombre d'attracteurs est directement lié aux possibilités de différenciation des cellules immuno-compétentes. L'activation des Natural Killer (NK), impliqués dans la réponse immunitaire innée, est contrôlée par les ligands de récepteurs de type Toll (TLR) et par RP105. Le réseau de contrôle de la Toll et d'expression de ICAM1 contient 2 circuits 4-positifs.

2.3.2 Exemple 2 : L'atrésie biliaire (« biliary atresia »)

L'atrésie des voies biliaires est caractérisée par une obstruction des voies biliaires d'origine inconnue, survenant en période périnatale. Par la suite, nous étudions le réseau de régulation relatif à cette maladie.

2.3.2.1 Présentation générale du réseau

Le réseau de départ contient 31 nœuds. Afin de le simplifier, nous avons enlevé les puits et les sources. Nous gardons seulement 22 nœuds.

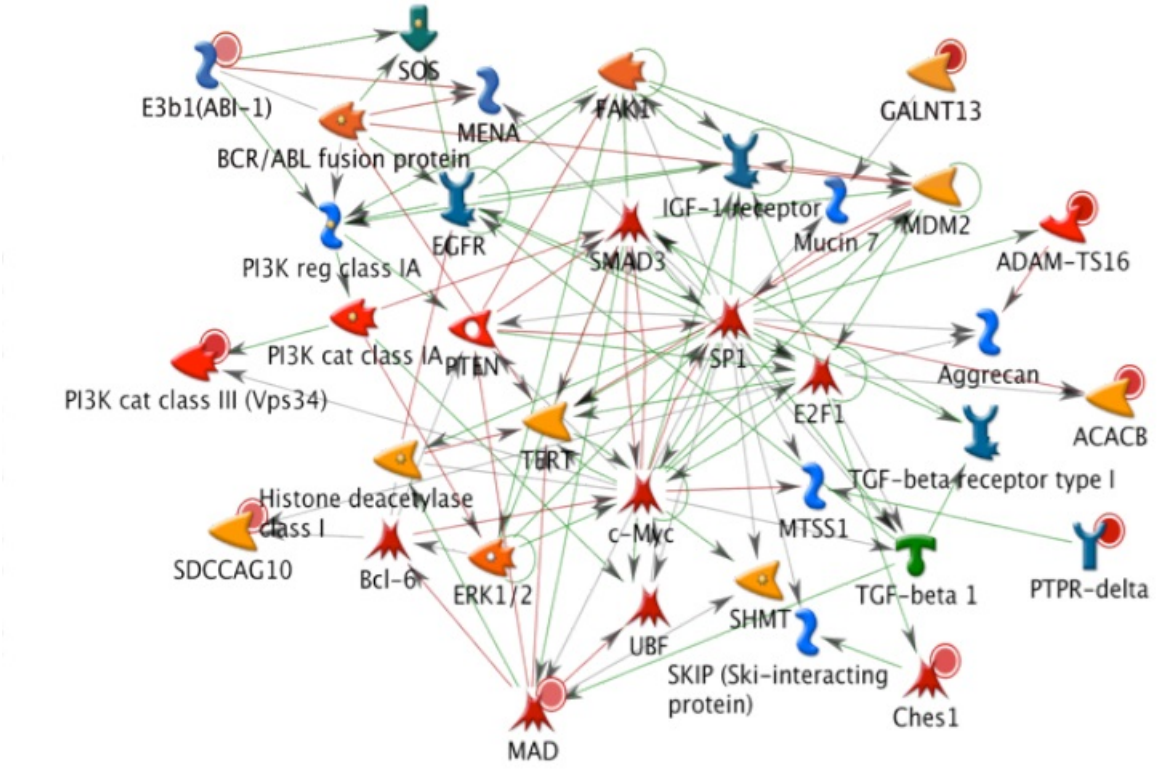


FIGURE 2.3 – Représentation du réseau de régulation pour l'atrésie biliaire.

2.3.2.2 Stratégie d'analyse

Toujours avec les réseaux booléens de type Hopfield, nous avons étudié le réseau en question. Le point de départ était un réseau conçu par A. Henrion Caude et dessiné par Meta Core®. Ce réseau a présenté plusieurs difficultés, vu l'existence de flèches non orientées (couleur grise, voir la Figure 2.3). Nous l'avons complété, en nous référant à la littérature, par l'introduction d'autres interactions entre les gènes. Afin de choisir le bon sens d'interaction, nous avons étudié toutes les combinaisons possibles, afin de déterminer toutes les

architectures possibles du réseau. Par la suite, nous avons comparé les architectures de chaque réseau par rapport à un réseau initial sans aucune flèche grise, afin de déterminer les interactions critiques, c'est-à-dire susceptibles, lorsqu'elles sont absentes ou qu'elles changent de valeurs (activation ou inhibition), de modifier considérablement le paysage des attracteurs du réseau (en nombre, nature et/ou taille de bassins). Nous avons testé à la fois la modification d'une seule flèche (ou arc), puis de deux,. . . , jusqu'au réseau complet avec toutes les flèches sélectionnées. L'hypothèse de départ était d'étudier le comportement des trois gènes clés : Bcl-w, TGF-β and elf-2α kinase, dont l'expression a été suivie chez des patients souffrant d'atrésie biliaire.

2.3.2.3 Outils utilisés

Nous avons utilisé une interface graphique développée au sein de notre laboratoire (thèse d'H. Ben Amor), nommée "Network-Designer", qui propose 4 types de mises à jour des gènes :

1. mise à jour parallèle, où les gènes sont mis à jour au même temps.
2. mise à jour séquentielle, où les gènes sont mis à jour successivement dans un ordre déterminé.
3. mise à jour des blocs en parallèle, où les gènes d'un bloc sont mis à jour séquentiellement, les blocs étant, eux, mis à jour parallèlement
4. mise à jour des blocs en séquentiel, où les gènes d'un bloc sont mis à jour parallèlement, les blocs étant, eux, mis à jour séquentiellement.

Le travail consiste ici à traduire le réseau avec cette interface sous la forme suivante :

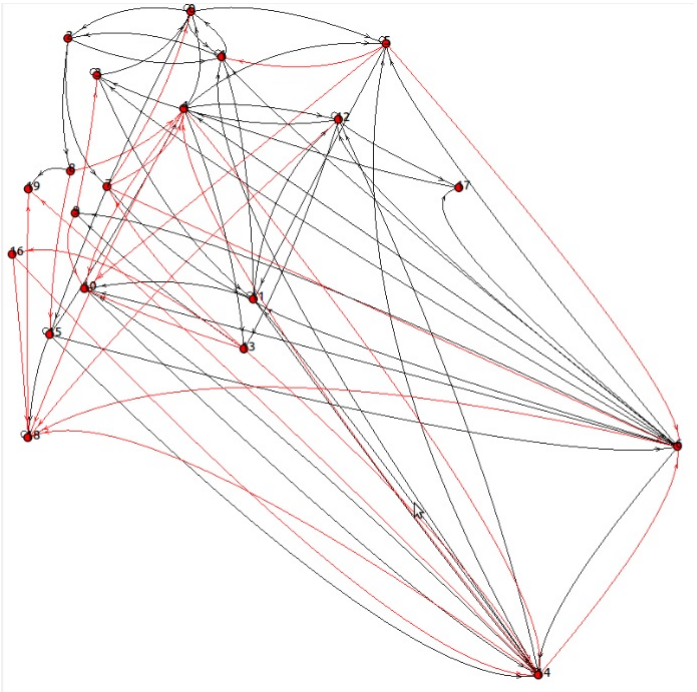


FIGURE 2.4 – Représentation du réseau de régulation pour l'atrésie biliaire avec "NetworkDesigner".

2.3.2.4 Résultats obtenus

Les résultats relatifs à ce réseau sont présentés dans l'article "Genetic regulatory networks : focus on attractors of their dynamics". Rappelons juste que nous avons obtenus 7 attracteurs répartis en 5 points fixes et 2 cycles limites.

Acta Biotheor (2013) 61:367–383
DOI 10.1007/s10441-013-9190-8

ORIGINAL ARTICLE

MitomiRs, ChloromiRs and Modelling of the microRNA Inhibition

J. Demongeot · H. Hazgui · S. Bandiera · O. Cohen · A. Henrion-Caude

Received: 24 December 2012 / Accepted: 20 July 2013 / Published online: 28 August 2013

Abstract MicroRNAs are non-coding parts of nuclear and mitochondrial genomes, preventing the weakest part of the genetic regulatory networks from being expressed and preventing the appearance of a too many attractors in these networks. They have also a great influence on the chromatin clock, which ensures the updating of the genetic regulatory networks. The post-transcriptional inhibitory activity by the microRNAs, which is partly unspecific, is due firstly to their possibly direct negative action during translation by hybridizing tRNAs, especially those inside the mitochondrion, hence slowing mitochondrial respiration, and secondly to their action on a large number of putative m-RNA targets like those involved in immunetworks; We show that the circuits in the core of the interaction graphs are responsible for the small number of dedicated attractors that correspond to genetically controlled functions, partly due to a general filtering by the microRNAs. We analyze this influence as well as their impact on important functions like the control

J. Demongeot (✉) · H. Hazgui · O. Cohen
AGIM, FRE CNRS 3405, Faculty of Medicine of Grenoble,
University J. Fourier, 38 700 La Tronche, France
e-mail: Jacques.Demongeot@agim.eu

H. Hazgui
e-mail: Hana.Hazgui@agim.eu

O. Cohen
e-mail: olivier.cohen@ujf-grenoble.fr

J. Demongeot
Escuela de Ingeniería Civil Informática y Departamento de Ingeniería Biomedica,
Universitad de Valparaiso, Valparaiso, Chile

S. Bandiera · A. Henrion-Caude
INSERM U 781, Hôpital Necker, Université Paris Descartes, 149 rue de Sèvres, 75015 Paris, France
e-mail: simonetta.bandiera@inserm.fr

A. Henrion-Caude
e-mail: alexandra.caude@inserm.fr

by the p53 network over the apoptosis/proliferation system and the homeostasis of the energy metabolism. In this last case, we show the role of two kinds of microRNAs, both involved in the control of the mitochondrial genome: (1) nuclear microRNAs, called mitoMirs, inhibiting mitochondrial genes and (2) putative mitochondrial microRNAs inhibiting the tRNAs functioning.

Keywords MicroRNAs · Chromatin clock control · Genetic networks · Energetic regulation · Mitochondrial respiration · Attractors

1 Introduction

The microRNAS (often denoted miRs) play an important role of non-specific inhibition in many circumstances of the cell life, like chromatin clock control and have a big influence on many metabolic controls of functions like energy systems, cell cycle and defence systems against pathogens. We present successively in Sects. 2 and 3 their action in controlling chromatin clock and "immunetworks", which are genetic regulatory networks devoted to the control and maintenance of the immunologic system, like the innate system of defence represented by the Toll Like Receptors (TLR), control network already present in insects. In Sects. 4 and 5, we will study more generally the global unspecific inhibitory regulation of the translation by microRNAs in different important physiological frameworks involving several functions like p53 control, mitochondrial and chloroplast respiration. In Sect. 6, we propose some perspectives about the general architecture of genetic regulatory networks taking into account the presence (recently discovered) of circular RNAs inhibiting the microRNAs.

2 MicroRNas, Chromatine Clock and Genetic Network Modelling

The genes coding for enzymes involved in the chromatin clock as histone acetyltransferases, endonucleases, exonucleases, helicase, replicase, polymerase,…, called clock genes (inducing for example the progressive unpacking of the Hox genes from the chromatin during development), are inhibited by many microRNAS preventing some blocks of genes to be co-expressed (cf. Fig. 1), e.g., RNA-dependent Helicase P68 and Endonuclease CCR4 are inhibited by miR-20, Helicase-DNA-binding protein KIAA1416 by miR181b, Exoribonuclease 2 and DNA Polymerase θ by miR93 and RNA Polymerase II transcription factor by miR-206 (Fig. 1). We use the circular Hamming distance between microRNAs string and a reference RNA circle of equal length, i.e., the minimal number of positions at which the corresponding symbols are different, when the circle origin is shifted (Demongeot and Moreira 2007a). The reference circle AL of length 22 (Demongeot and Moreira 2007b), close to Lewin's invariant sub-sequences of the tRNA secondary structures (Lewin et al. 2011), contains a triplet of any amino-acids synonymy class of the genetic code (including punctuation classes) and can be considered as a primitive "matrimonial agency" (or an ancestral ribosome) for

amino-acids favouring their peptidic bonds (Hobish et al. 1995; Demongeot et al. 2009a, b).

The proximity of microRNAs (Griffiths-Jones et al. 2005; http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/) to AL is measured by the significance level equal to 95 % (resp. 99.5 %) if the circular Hamming distance d is less or equal to $7 = 22 - 15$ (resp. 5), indicated by ** (resp. ***) as calculated from an unilateral test based on the distribution of the supremum of 22 independent binomial variables B(22, 0.375), greater than the variable (d-22) (Demongeot and Moreira 2007a). Figure 2 shows that (1) several levels of enzymes, like helicase are controlled via the inhibition of the translation of their mRNA by several microRNAs like mi-R20 and miR-181b, and (2) several levels of enzymes like helicase and endonuclease (resp. exoribonuclease and polymerase) are controlled by the same microRNA miR-20 (resp. miR-93). A microRNA exerts its influence in the framework of a genetic regulatory network including these enzyme genes, we call in the following threshold Boolean random regulatory network (getBren), which is just a set N of n random automata defined as follows (Hopfield 1982; Hartwell et al. 1999; Weaver et al. 1999; Kauffman 1969; Thomas 1973; Demongeot et al. 2003; Demongeot and Sené 2008; Demongeot and Waku 2012a, b; Demongeot et al. 2012a):

1. any random automaton i of the getBren N owns at time t a state $x_i(t)$ valued in {0,1}, 0 (resp. 1) meaning that gene i is inactivated (resp. activated). The global state of the getBren at time t, called configuration in the sequel, is then defined by: $x(t) = (x_i(t))_{i \in \{1,n\}} \in \Omega = \{0,1\}^n$
2. the getBren N of size n is a triplet (W,Θ,P), where:
   - W is a matrix of order n, where the coefficient $w_{ij} \in$ IR represents the interaction weight equal to the influence gene j has on gene i. $A = \text{Sign}(W) = (\alpha_{ij} = \text{sign}(w_{ij}))_{i,j}$ is the adjacency matrix of the interaction graph G.

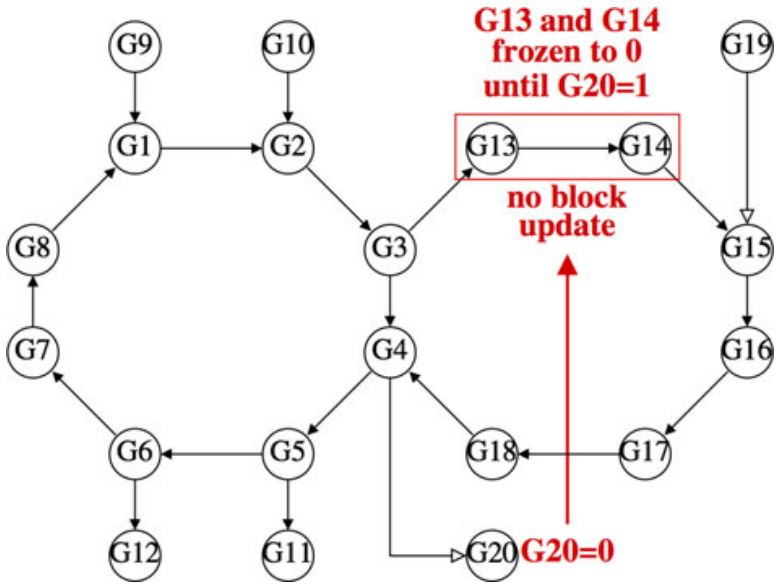


Fig. 1 Influence of the chromatin clock on the genetic network dynamics: toy theoretical example in which G20 (a "leaf node" of the interaction graph) expresses a chromatin clock's enzyme, which prevents the expression of the updating block G13–G14

Sequences	Identification	Circular Hamming distance to AL
5'_UACURRY UUCRARYURARYGG_3'	Lewin's tRNA (R=A or G; Y=U or C)	5***
5'_UACUGCC AUUCAAGAUGAAUGG_3'	AL	
3'_AUGGACG –UGAUAUUCGUGAAAU_5'	hsa-miR-20	8
5'_CAGCAGTAATTATGGTGCACTTTT_3'	Helicase P68 ENST00000225792 112-136	
3'_AUGGACGU--------GAUAUUCG UGAAAU_5'	hsa-miR-20	8
5'_TCCCAGCACCATTCTATAGGCAACTTTT_3'	Endonuclease CCR4 ENST00000264903 3312-3340	
5'_GUACUGCCAUUCAAGAUGAAUG_3'	AL	
3'_UUGGGUGGCUGUCGUUACUUACAA_5'	hsa-miR-181b	7**
5'_ AACCTTACAAGTGT AATGAATGTG_3'	Helicase-DNA-binding protein KIAA1416 ENST00000333363 0-24	
5'_CAUUCAAGAUGAAUGGUACUGC_3'	AL	
3'_GAUGGACGUGCUUGUCGUGAAA_5'	hsa-miR-93	7**
5'_TTAATAAAAC – TACAGTACTTT_3'	Exoribonuclease 2 ENST00000262489 85-106	
3'_GAUGGAC---- GUGCUUG--UCGUGAAA_5'	hsa-miR-93	7**
5'_ATGCTTGAAATATGCACTTAGCACTTT_3'	DNA Polymerase θ ENST00000264233 293-320	

Fig. 2 Sequence alignments and circular Hamming distances to AL for miR-20 (resp. miR-181b and miR-93) from (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/) and its inhibited targets, RNA-dependent Helicase P68 and Endonuclease CCR4 (resp. Helicase-DNA-binding protein KIAA1416, Exoribonuclease 2, and DNA Polymerase). The number of *stars* indicates the significance level of the distance to AL

- Θ is an activation threshold vector of dimension n, its component θ_i being the activation threshold attributed to automaton i
- M: $P(\Omega) \rightarrow [0,1]^{m\times m}$ (where $P(\Omega)$ is the set of subsets of Ω and $m = 2^n$) is a Markov transition matrix, built from local probability transitions P_i giving the new state of gene i at time $t+1$ according to W, Θ, and configuration $x(t)$ of N at time t such that:

$$\forall i = 1,\ldots,n,\ \forall g \in \{0,1\}, \quad \beta \in \Omega, P_{i,g}^{\beta}(\{x_i(t+1) = g | x(t) = \beta\}) = \exp[g(\Sigma_{j\in Ni} w_{ij}\beta_j - \theta_i)/T]/[1 + \exp[(\Sigma_{j\in Ni} w_{ij}\beta_j - \theta_i)/T],$$

where N_i is the neighbourhood of gene i in the getBren N, i.e., the set of genes j (including possibly i) such that $w_{ij} \neq 0$, and $P_{i,g}^{\beta}$ is the probability for gene i of passing to state g knowing the configuration state β on N_i.

An important feature of the getBren dynamics is the existence of genes influencing directly the opening of the DNA inside the chromatine, hence allowing or not the gene expression. When these genes are controlled by microRNAs, it is necessary to generalize getBren structure by considering that the possibility to update the expression of a block of genes $i_1,\ldots,i_q \leq n$ at iteration t is depending on the state of r "clock" genes (i.e., involved in the chromatine updating clock) $k_1,\ldots,k_r \geq n$, depending on s microRNAs $l_1,\ldots,l_s$ (Fig. 1). Then the transition for a gene i, can be written as:

$\forall g\in\{0,1\}$, $\beta\in\{0,1\}^n$, if $\forall\ j = 1,\ldots,r$, $x_{kj}(t) = 1$, involving $\forall\ j = 1,\ldots,\ s$, $x_{lj}(t-1) = 0$ (microRNAs not inhibit), then:

$$\forall i = 1,\ldots,n,\ \forall g \in \{0,1\}, \quad \beta \in \Omega, P_{i,g}^{\beta}(\{x_i(t+1) = g | x(t) = \beta\}) = \exp[g(\Sigma_{j\in Ni} w_{ij}\beta_j - \theta_i)/T]/[1 + \exp[(\Sigma_{j\in Ni} w_{ij}\beta_j - \theta_i)/T]$$

and, if not, this rule is available only for genes $i \neq i_1,\ldots,i_q$ and: $\forall\ j = 1,\ldots,q$, $P_{ij,0}^{\beta}(\{x_{ij}(t+1) = 0 | x(t) = \beta\}) = 1$.

3 The Immunetworks

The genetic regulatory networks involved in the control of both innate and adaptive immunologic systems (ruling respectively the fixed and antigen-dependent pools of antibody receptors), called "immunetworks", have only few asymptotic dynamical behaviours, called attractors (Demongeot et al. 2011a). This small number of attractors is directly linked with the possibilities of differentiation of immuno-competent cells and is controlled inside the interaction graphs of immunetworks by their circuits (Demongeot et al. 2012b), giving to the network the possibility to have more than one attractor (due to positive circuits, having inside an even number of inhibitions) and the possibility for these attractors to be stable (due to negative circuits, having inside an odd number of inhibitions), i.e., to have a sufficient number of initial configurations of gene states going after a transient evolution to the same attractor.

The activation of Natural Killer (NK) cells, involved in the innate immune response, is controlled by the ligands of the Toll Like Receptors (TLR) and RP105 is a TLR-related protein acting as regulator of the proliferation of B cells, responsible of the humoral adaptive immune response (Elkon et al. 2007; Miyake et al. 2000). B-cells lacking RP105 were shown to be severely impaired in antibody production. Protein ICAM1 is an intercellular adhesion molecule present in low concentrations in membranes of leucocytes involved in humoral adaptive immune response. The network controlling the Toll Like Receptor TLR2 and ICAM1 expression (Fig. 3) contains 2 positive 4-circuits (i.e., with 4 nodes) passing through genes MAP3K and BCL10 having the common node NFkB, and a negative 4-circuit passing through the gene TNIP1, giving birth to only one attractor (Fig. 5) leading to the activation of TLR2. ICAM1 is inhibited by the human miR 297 and the activator GATA-3 of BCL10 by the human miR 200c (Fig. 4).

The pUNO sub-sequence of RP105, activator of TRL2 (Liu et al. 2013), contains the sequence AGATGAA belonging both to Tψ-loop common to Lewin's tRNA and AL, which signifies its affiliation to an ancestral functional genome, because this

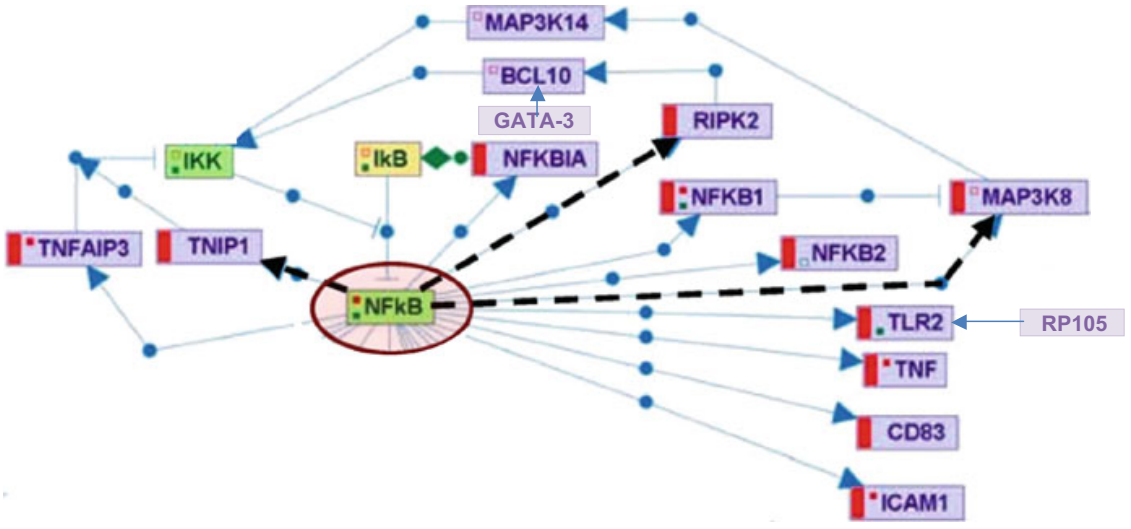


Fig. 3 Network controlling the level of Toll Like Receptor TLR2 and of ICAM1 [after (Elkon et al. 2007), links with bar indicating inhibitions]

Sequences	Identification	Circular Hamming distance to the miR anti-sense
3'-AGGUAGUAAUGGGC-CGUCAUAA-5'	miR 200c	
5'- TGCATT TTTGCAGGAGCAGTATC-3'	GATA-3 sequence starting in position 57	5***
3'-GUACGUGUACGUGUGUAUGUA-5'	hsa-miR 297	
5'-CCTCCCCA------CCCACATACAT-3'	ICAM1 sequence starting in position 832	7**

Fig. 4 Sequence alignments and circular Hamming distances of GATA-3 and ICAM1 to their miR anti-sense

conserved region is recognised by the ribosome to allow a tRNA-ribosome complex to form (Demongeot and Moreira 2007a, b).

4 MicroRNas and p53 Control

The p53 protein is involved in many genetic regulatory networks, especially those controlling the cell cycle and apoptosis. The network centred on p53 given in Fig. 5 shows the importance of this key protein, both controlling as activator (resp. inhibitor) the transcription of microRNAs like miR 143 (resp. miR 20a) and being negatively controlled by microRNAs like miR 122. The presence of a positive 3-circuit (p53, miR 34, SIRT 1) gives two fixed state configurations and two state limit cycles of period 3 as attractors for parallel updating (Fig. 6) in absence of microRNAs inhibiting p53 and if not, only one fixed state configuration: p53 = 0, miR-34 = 0, SIRT 1 = 0.

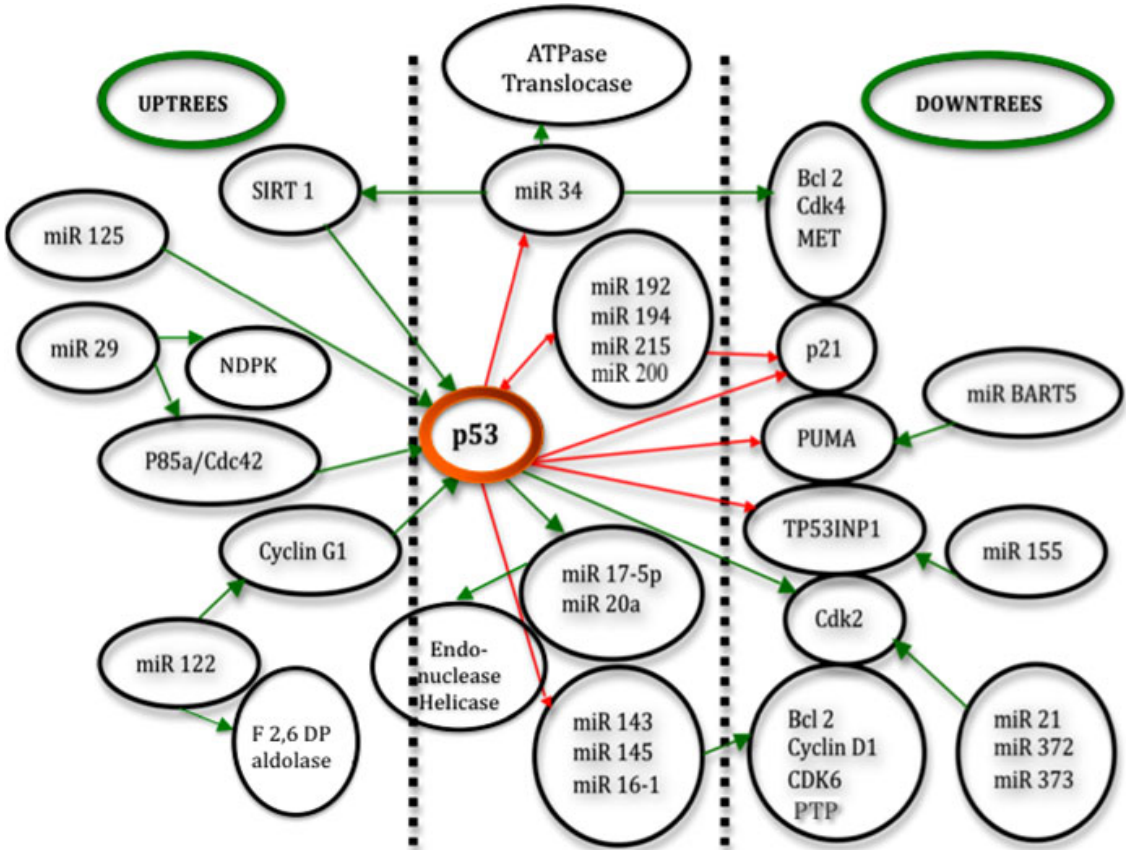


Fig. 5 Regulation of protein p53 showing relationships with microRNAs as inhibitors (*green arrows*) or activators (*red arrows*)

ℓ \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	2	2	3	3	5	5	8	10	15	19	31	41	64
2	1	1	2	3	3	4	5	8	10	14	19	31	41	63
3	1	2	1	3	3	6	5	8	8	15	19	33	41	64
4	1	1	2	1	3	4	5	11	10	14	19	24	41	63
5	1	2	2	3	1	5	5	8	10	26	19	31	41	64
6	1	1	1	3	3	1	5	8	8	14	19	63	41	63
7	1	2	2	3	3	5	1	8	10	15	19	31	41	158
8	1	1	2	1	3	4	5	1	10	14	19	24	41	63
9	1	2	1	3	3	6	5	8	1	15	19	33	41	64
10	1	1	2	3	1	4	5	8	10	1	19	31	41	63
11	1	2	2	3	3	5	5	8	10	15	1	31	41	64
12	1	1	1	1	3	1	5	11	8	14	19	1	41	63
13	1	2	2	3	3	5	5	8	10	15	19	31	1	64
14	1	1	2	3	3	4	1	8	10	14	19	31	41	1

p \ n	1	2	3	4	5	6	7	8	9	10	11	12	21	22
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	–	1	–	1	–	1	–	1	–	1	–	1	–	1
3	–	–	2	–	–	2	–	–	2	–	–	2	2	–
4	–	–	–	3	–	–	–	3	–	–	–	3	–	–
5	–	–	–	–	6	–	–	–	–	6	–	–	–	–
6	–	–	–	–	–	9	–	–	–	–	–	9	–	–
7	–	–	–	–	–	–	18	–	–	–	–	–	18	–
8	–	–	–	–	–	–	–	30	–	–	–	–	–	–
9	–	–	–	–	–	–	–	–	56	–	–	–	–	–
10	–	–	–	–	–	–	–	–	–	99	–	–	–	–
11	–	–	–	–	–	–	–	–	–	–	186	–	–	186
12	–	–	–	–	–	–	–	–	–	–	–	335	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	99858	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	190557
T_n^+	2	3	4	6	8	14	20	36	60	108	188	352	99880	190746

Fig. 6 *Top* number of attractors for a couple of positive r-circuit and negative l-circuit sharing one node. *Bottom* number of attractors of period p for a n-positive circuit in parallel updating [after (Demongeot et al. 2012b)]

5 Energetic Control

5.1 MicroRNAs and Oxidative Phosphorylation

The cellular energy systems of the cell are essentially composed of glycolysis and aerobic oxidation. In eukaryotes, later stages of the second system occur in mitochondria, with enzymatic steps like ATPase and translocase. For each corresponding gene, it is possible to find at least one nuclear microRNA inhibiting its activity (Fig. 7 Top). More, mitochondrial microRNAs (Barrey et al. 2011) have been described, as well as 2 nuclear miRs susceptible to hybridize with a perfect anti-match

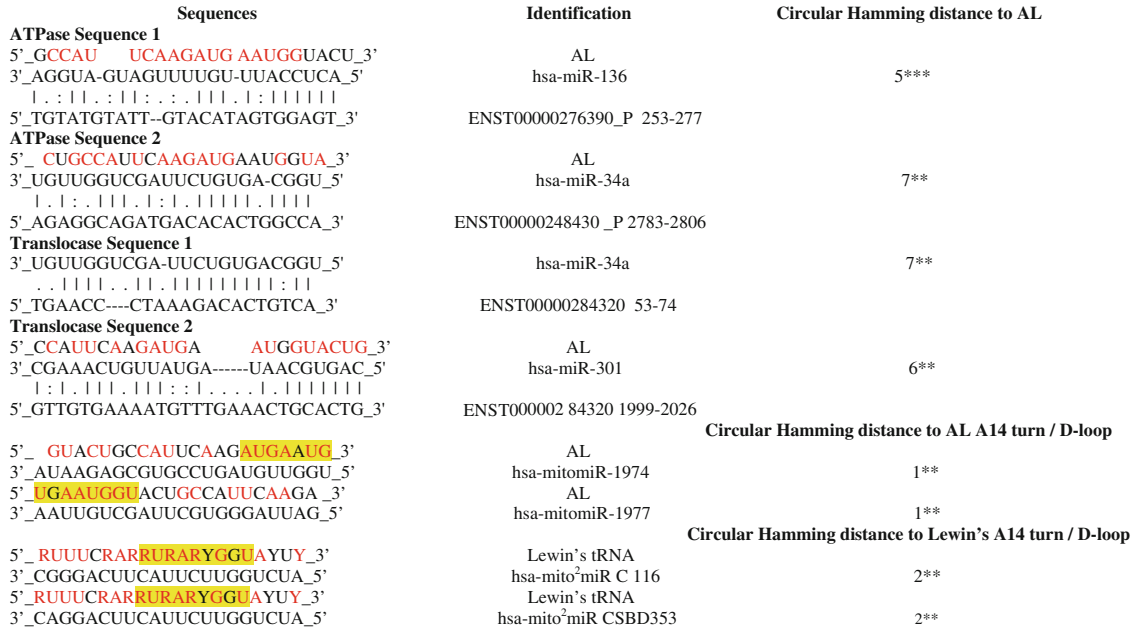


Fig. 7 *Top* sequence alignments and distances to AL for human miRs136, 34a and 301 from (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/) and their inhibited targets, ATPase and Translocase. *Bottom* sequences of nuclear mitomiRs and mito²miRs from mitochondrial non coding genome (Bandiera et al. 2011; Demongeot et al. 2011b; Sbisa et al. 1997; Cui et al. 2007) and alignment with the AL and Lewin's tRNA A14 turn and D-loop sub-sequences in yellow (R = A or G; Y = U or C)

mitochondrial genes, miR-1974 and miR-1977 (Bandiera et al. 2011; Demongeot et al. 2011b), targeting two mitochondrial tRNA genes, TRNE and TRNN, which code respectively for ATP8 (the gene of ATPsynthase protein 8, a subunit of the mitochondrial ATPase) and for ND4L, a protein providing instructions for making NADH. On Fig. 7 Bottom, we see the possibility to hybridize tRNA loops (especially the D-loop and the A14 turn of the tRNA, responsible of its tertiary structure) by nuclear mitomiRs 1974 and 1977 and on Fig. 7 Bottom, we see two small mitochondrial RNAs called mito²miRs C116 and CSBD 353 (Sbisa et al. 1997; Cui et al. 2007), coming from the non coding part of mitochondrial genome, which causes an unspecific inhibitory 'noise' due to a direct regulation (inside mitochondrion matrix, without membrane transfer like for nuclear miRs), leading possibly to a decrease in levels of all mitochondrial proteins, due to the hybridization of common parts of mitochondrial tRNAs and Lewin's tRNA: D-loop, A14 turn and Tψ-loop.

5.2 MicroRNas and Glycolysis

Many MiRs from (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/) are proposed as candidates for inhibiting all glycolytic steps until aldolase (Fig. 8 Top), a possible alternative (or complement) to the direct interaction between metabolites and RNA aptamers, their absence favouring Krebs cycle functioning, if translocase and ATPase have sufficient level inside the mitochondrial inner membrane (Demongeot et al. 2007) and LDH is inhibited, or conversely their presence explains the role of the lactate shuttle from brain astrocytes to neurons (Aubert et al. 2005, 2007;

Fig. 8 *Top* sequence alignments of nuclear miRNAs inhibiting high glycolysis genes (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/). Bottom: Unspecific hybridization of the tRNA D-loop (in *yellow*) by nuclear miRNAs inhibiting low glycolysis and adenylate/guanylate pool genes (http://mirdb.org/miRDB/)

Genc et al. 2011). Glycolysis contains a key step, phosphofructokinase (PFK), a highly non-linear allosteric enzyme (Demongeot and Laurent 1983) with effectors ATP and ADP, whose pools are regulated by miR-302 and 29a, which inhibit respectively ADK and DNPK (Figs. 8 Bottom, 9) in a negative regulatory circuit causing, for critical values of the fructose entry flux J, oscillations for all glycolytic metabolites, with a period of several minutes (Boiteux et al. 1975; Demongeot and Seydoux 1979; Hervagault et al. 1983; Demongeot and Kellershohn 1983; Demongeot and Doncescu 2009).

Concerning the energy balance in neurons, recently proposed Astrocyte-Neuron Lactate Shuttle Hypothesis (ANLSH) suggests that the glial glucose metabolism is almost completely anaerobic, hence the lactate generated by astrocytes is transferred to neurons, which consume this lactate coming from the extracellular space,

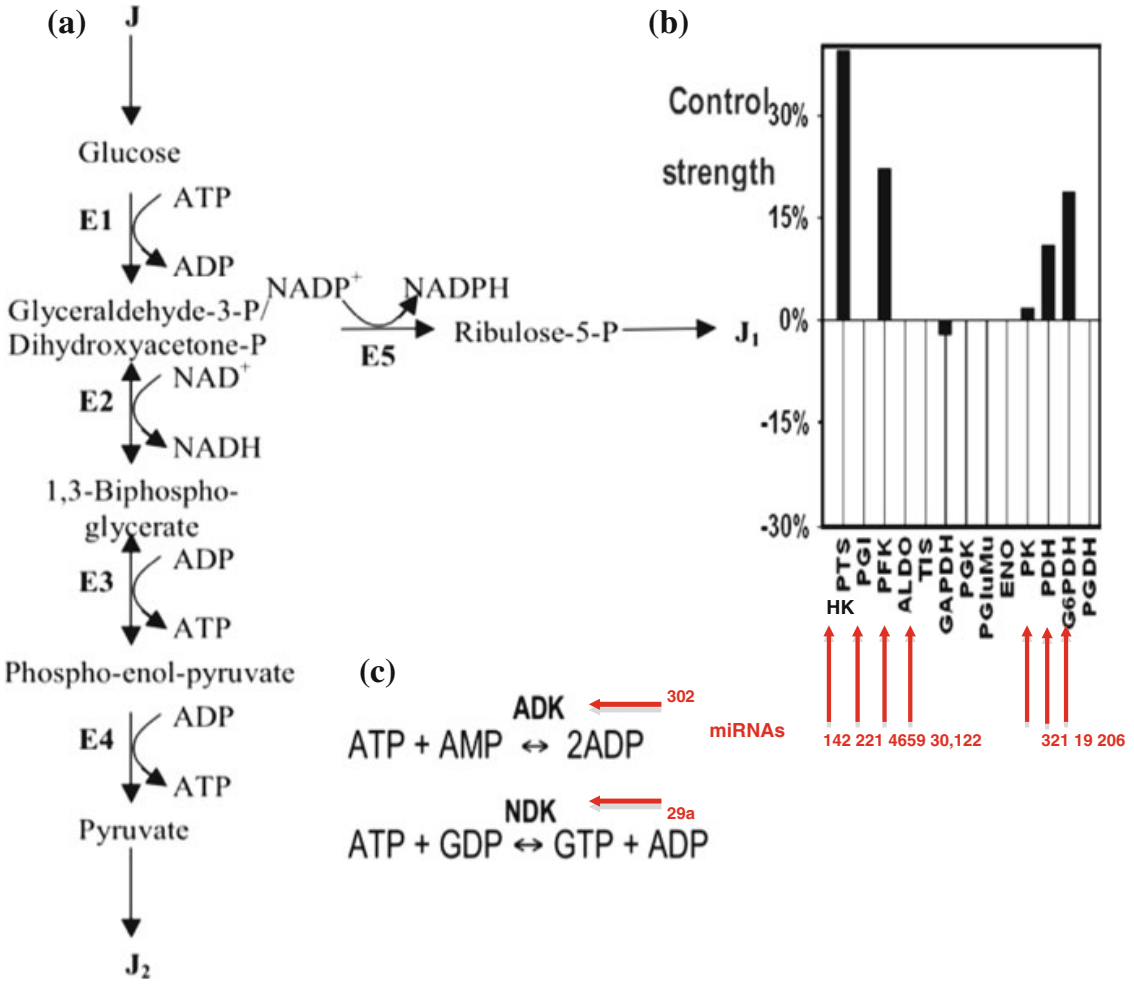


Fig. 9 **a** Glycolysis main steps with indication **b** of their enzyme control strengths and of their inhibitory microRNAs (*red arrows*). E1 denotes the 4 enzymes of the high glycolysis (hexokinase HK, phosphogluco-isomerase PGI, phosphofructo-kinase PFK and aldolase ALDO), E2 denotes the glyceraldehyde-3P-dehydrogenase, E3 denotes the 3 enzymes of the low glycolysis (phosphoglycerate-kinase, phosphoglycerate-mutase and enolase ENO), E4 denotes the pyruvate-kinase PK and E5 the 3 enzymes of the pentose pathway (glucose-6P-dehydrogenase G6PDH, 6P-glucono-lactonase and phosphogluconate-dehydrogenase PGDH, alternative to phospho-transferase system). **c** Adenylate (guanylate) pool ruled by ADK (NDK)

partially replenished by the astrocytes production, giving to neurons an extra-pyruvate production from lactate, and an extra-oxygen and glucose consumption and a high ATP level, theoretically predicted and experimentally observed (Aubert et al. 2005, 2007; Genc et al. 2011). This ATP level depends on ATPase, Translocase, Adenylate Kinase and Nucleoside DiPhosphate Kinase concentrations, which are in human under the negative control of 5 miRNAs, the hsa-miR-136, 34, 301, 302 and 29a (Fig. 8) acting as boundary control nodes (i.e., nodes at the edge of a subnetwork feeding into another subnetwork) regulating the neuronal oxidative system efficacy (Fig. 9), in proportion of the control strength, i.e., the percentage of glycolytic flux ruled by the enzymes they inhibit (Kaczer and Burns 1973; Ovadi 1988; Reder 1988; Baconnier et al. 1993; Bier et al. 1996; Wolf and Heinrich 2000; Ruoff et al. 2003; Ritter et al. 2008).

5.3 Physiological Role of microRNAs in Cell Energetics

The possible inhibitory unspecific noise by miRs due to their direct interaction with mitochondrial tRNAs has a global resultant on the intermediary metabolism of the superior organisms like mammals. Their role is essential in the inhibition/activation balance of network dynamics preventing too much attractors to be expressed, which could lead either to pathologic rhythms, or to steady states with too heavy or weak concentrations of certain metabolites.

Following the comparative physiology study about cell consumption of oxygen by mammals done in Weibel et al. (1991), let consider V, the total mitochondrial volume of muscle and V_{O_2max}, the O_2 consumption rate per unit mass of whole body mass M of these mammals. Table 1 shows that non-ruminant animals like horse and dog have an efficacy in O_2 consumption about 2.5 times more than ruminants like steer and goat. We have then studied in the non-coding mitochondrial genome (called mitochondrial D-loop, not to be confused with tRNA D-loop…) of the non-ruminants if there were less mitomiRs candidates than in that of ruminants. Table 2 shows that in ruminants such mitomiRs do exist and contain sub-sequences (between 6 and 7 bases long) that are identical to mitochondrial tRNA loops. In the non-ruminants, however, these sub-sequences are totally absent in the non-coding mitochondrial genome of the non-ruminants. These genomes have been extracted from the classical genetic data bases dedicated to the whole mitochondrial (http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html) and chloroplast (http://www.ncbi.nlm.nih.gov/nuccore/EF1155; http://www.ncbi.nlm.nih.gov/nuccore/AB042240) genomes: Megasun from the University of Montreal (http://megasun.bch.umontreal.

Table 1 Comparison between cellular oxygen consumption of mammals (Weibel et al. 1991)

	V_{O_2max}/M (ml s^{-1} kg^{-1})	V/M (ml/kg)
Dog	2.29	40.6
Goat	0.95	13.8
Horse	2.23	30.0
Steer	0.85	11.6

Table 2 Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the non-coding mitochondrial genomes of bos (total length 8493 bases) and equus (total length 4535 bases) (http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html) (bos is the genus to which steer belong and equus is horse)

Sub-sequences	Observed bos	Expected bos	Observed equus	Expected equus
taccac	2	2	0	1
taccat	1	2	0	1
acc gtt	2	2	0	1
accatt	1	2	0	1
cttgaa	1	2	0	1
atttgaa	1	0.5	0	0.25
Total	8	10.5 ± 6.5 (m ± 2σ)	0	5.25 ± 4.6 (m ± 2σ)

ca/ogmp/projects/other/cp_list.html), and Nuccore from the NCBI (http://www.ncbi.nlm.nih.gov/nuccore/EF1155; http://www.ncbi.nlm.nih.gov/nuccore/AB042240).

From the Tables 2, 3, and 4, we see that the absence of mitochondrial small sub-sequences, we have called mito^{2}miRs, in the non-coding mitochondrial genomes of non ruminants (for a total length of 10024 bases) susceptible to hybridize the mitochondrial tRNAs inside the mitochondrion matrix, is highly significant (observed number equal to 0 and mean expected number 10 with a standard deviation equal to 3.2).

A classical normal test of difference between empiric means shows that the number of ruminant mito^{2}miRs is significantly (with a threshold of 10^{-3}) superior to that of non ruminant. The absence of possibility of inhibition of the tRNA transcription inside the non ruminant mitochondria can participate to the better efficacy of the cell respiration in these non ruminants with respect to the ruminants, in other terms the inhibition of the translation by interaction of mito^{2}miRs with tRNAs in the mitochondria of ruminants may be responsible for their less efficient respiration, even if the difference in aerobic capacity (Weibel et al. 1991) can be also related to building more same mitochondria.

Table 3 Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the non-coding mitochondrial genomes of capra (total length 1,531 bases) and canis (total length 3,239 bases) (http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html) (capra is goat and canis dog)

Sub-sequences	Observed capra	Expected capra	Observed canis	Expected canis
taccac	1	0.4	0	0.8
taccat	0	0.4	0	0.8
accgtt	1	0.4	0	0.8
accatt	1	0.4	0	0.8
cttgaa	0	0.4	0	0.8
atttgaa	0	0.1	0	0.2
Total	3	2.1 ± 2.9	0	4.2 ± 4.1

Table 4 Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the non-coding mitochondrial genomes of ruminants (total length 10,024 bases) and non ruminants (total length 7,774 bases) (http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)

Sub-sequences	Observed ruminants	Expected ruminants	Observed non ruminants	Expected non ruminants
taccac	3	2.4	0	1.9
taccat	1	2.4	0	1.9
accgtt	3	2.4	0	1.9
accatt	2	2.4	0	1.9
cttgaa	1	2.4	0	1.9
atttgaa	1	0.6	0	0.5
Total	11**	12.7 ± 7.1	0**	10 ± 6.3

**Corresponds to a significance level of 1 ‰ of error in rejecting the hypothesis that the difference is null between ruminants and non ruminants

In same spirit, we analyzed data coming from non-coding chloroplast genome (http://www.ncbi.nlm.nih.gov/nuccore/EF1155; http://www.ncbi.nlm.nih.gov/nuccore/AB042240) of cereals, one belonging to C_4 plants, the sorghum (Sorghum bicolor), and the other from C_3 plants, the wheat (Triticum æstivum) which represent the most important kind grown in temperate countries. C_4 plants are known to have a better efficacy of their chloroplast diurnal respiration than C_3 plants, as we can see on Table 5 (Byrd et al. 1992; Farquhar et al. 1980), where sorghum is shown to be 1.5 times more efficient than wheat, by comparing CO_2 consumption (i.e., taking up carbon during photosynthesis and losing it during respiration) rate per unit surface of whole surface S of the leaves of these cereals.

In these cereals, the CO_2 assimilation is usually limited by the capacity of photosynthetic electron transport to supply ATP and NADPH to regenerate RuBP (Byrd et al. 1992; Farquhar et al. 1980). Then we can compare as for the animal mitochondria the presence of sub-sequences common to the non-coding chloroplast genome and to the chloroplast tRNA loops. From the Table 6, we see that the number of chloroplast small sub-sequences, we are calling chloromiRs, in the non-coding chloroplast genome of a C_4 plant, the sorghum (Sorghum bicolor, for a total

Table 5 Comparison between cellular CO_2 diurnal consumption between 2 cereals, sorghum and wheat (Byrd et al. 1992)

	V_{CO_2max}/S mmol s^{-1} m^{-2}
Sorghum bicolour	41.1 ± 2.8
Triticum æstivum	27.9 ± 1

Table 6 Number of common sub-sequences of length 6 and 7 between the chloroplast tRNA loops and the non-coding chloroplast genomes of sorghum (total length 15,852 bases) and triticum (total length 16,014 bases)

Sub-sequences	Observed sorghum	Expected sorghum	Observed triticum	Expected triticum
taccaCT	0	1	0	1
taccaTT	4	1	4	1
TaccgCt	0	1	0	1
TaccGtt	1	1	1	1
atttgaa	3	1	4	1
GTTTGAA	1	1	5	1
ATTCGAA	3	1	5	1
GTTCGAA	3	1	3	1
ACTTGAA	0	1	2	1
GCTTGAA	0	1	1	1
ACTCGAA	3	1	3	1
GCTCGAA	1	1	1	1
Total	19**	11.7 ± 6.8 (m ± 2σ)	29**	11.6 ± 6.8 (m ± 2σ)

**Corresponds to a significance level of 1 ‰ of error in rejecting the hypothesis that the difference is null between sorghum and triticum

length of 15,852 bases) susceptible to hybridize the chloroplast tRNAs inside the chloroplast matrix is highly significantly less than the same number observed in a C_3 plant, the wheat (Triticum æstivum, for a total length of 16,014 bases). A classical normal test of difference between empiric means shows that the number of the C_4 plant chloromiRs is significantly (with a threshold of 10^{-3}) less than that of the C_3 plant. The diminution of the possibility of inhibition of the tRNA function inside the chloroplast of the sorghum can participate to the better efficacy of the cell diurnal respiration in this C_4 plant with respect to the less efficient C_3 plant, the wheat.

6 Conclusion and Perspectives

The sensitivity of real biological networks with respect to the inhibitory action exerted by microRNAs has to be studied in further more systematic studies, to be performed in order to confirm the influence of boundary negative interactions. A theoretical tool to do it could be the Hopfield-like regulatory interaction networks (Demongeot and Sené 2008; Demongeot and Waku 2012a, b), in order to make more precise the influence of the microRNAs on the number of attractors, which is conjectured to diminish, when they are multiple on the boundary of the interaction graph of the network, reinforcing the robustness of the networks. More, we would have to take into account that the microRNAs can also been inhibited as proved by the recent discovery of circular RNAs, which are called "RNA sponges"; these circular RNA sequences (ciRS) have been found firstly in plants, then in human (Surono et al. 1999) and now are unambiguously validated in animals: in Memczak et al. (2013) and Hansen et al. (2013), the authors exhibit a human ciRS, ciRS-7, antisense gene of cerebellum degeneration-related protein CDR1, which contains binding sites for miR-7, perfectly conserved from annelids to humans, with practically the same role in protein translation regulation, despite of the fact that it regulates in *Drosophila* some proteins not regulated in mammals (Li et al. 2009); miR-7 functions in several interlocking feedback and feedforward loops, e.g., in a negative circuit involving the couple of genes Egfr/RAS, increasing the stability of the MAPK network. Conserved miRNAs like miR-7 seem to diminish the number of attractors, increasing the robustness of the genetic regulatory networks. Then, the second layer of regulation represented by the ciRs could serve to prevent an over-inhibition by critical miRNAs, leading to pathologic down-regulations. On Fig. 10, we see that a good match exist between a sequence called AGAT, homologous from 13 species to a human CDR1 sequence (Memczak et al. 2013) and between the half part of the ABEL sequence, proposed in (Weil et al. 2004) as an archetypal sequence for the ancient genetic code. ABEL contains, like the reference sequence AL, from start codon AUG until end codon UGG, one and only one codon per

```
ABEL sequence      5'-AUGGCACUGAAUGUUCCAAGAU-3'
AGAT sequence                  5'-GTCTTCCAAGAT-3'
hsa-miR-7       5'-AACAAAAUCACUAGUCUUCCA-3'
```

Fig. 10 Matches between the ABEL, AGAT and human miR-7 sequences

synonymy class of amino-acids (Demongeot and Moreira 2007a, b; Demongeot et al. 2009a, b). ABEL, close to AL, shares with any miRNA subsequences of length 5 (like UUCCA) more than expected by chance, and is close to sequences coming from many UTR viral genomes (Demongeot et al. 2009a).

The antisense sequence 3′–5′ of AGAT matches the corresponding sequence 5′–3′ of miR-7, which explains the inhibitory power of ciRS-7. From miRNAs bases (http://mirdb.org/miRDB/), we found the sequence UCUUCCA in about 25 of 500 miRNAs, significantly more than expected (about 2 ± 4). In (http://mamsap.it.deakin.edu.au/~amitkuma/mirna_targetsnew/sequence.html), 16S and 23S rRNA data sets and trees spanning all sequenced type strains of classified species of Archaea and Bacteria are reconstructed, with 249 times the sequence AUCUUCCA, significantly more than the expected frequency (about 160 ± 40). Hence, ciRs containing AGAT sequence inhibit other miRS than miR-7, a major argument for introducing circular RNAs in the future models of genetic regulation by microRNAs (Demongeot et al. 2013a, b; Antonopoulos et al. 2013; Bandiera et al. 2013) to be further investigated.

References

Antonopoulos C, Basios V, Demongeot J, Nardone P, Thomas R (2013) Linear and nonlinear arabesques: a study of closed chains of negative 2-element circuits. Int J Bifurcat Chaos (in press); arXiv:1302.1071

Aubert A, Costalat R, Magistretti PJ (2005) Brain lactate kinetics: modeling evidence for neuronal lactate uptake upon activation. Proc Natl Acad Sci USA 102:16448–16453

Aubert A, Pellerin L, Magistretti PJ, Costalat R (2007) A coherent neurobiological framework for functional neuroimaging provided by a model integrating compartmentalized energy metabolism. Proc Natl Acad Sci USA 104:4188–4193

Baconnier P, Pachot P, Demongeot J (1993) An attempt to generalize the control coefficient concept. J Biol Syst 1:335–347

Bandiera S, Rüberg S, Girard M, Cagnard N, Hanein S, Chrétien D, Munnich A, Lyonnet S, Henrion-Caude A (2011) Nuclear outsourcing of RNA interference components to human mitochondria. PLoS ONE 6:e20746

Bandiera S, Matégot R, Demongeot J, Henrion-Caude A (2013) MitomiRs: delineating the intracellular localization of microRNAs at mitochondria. Free Radic Biol Med. doi:10.1016/j.freeradbiomed.2013.06.013

Barrey E, Saint-Auret G, Bonnamy B, Damas D, Boyer O, Gidrol X (2011) Pre-microRNA and mature microRNA in human mitochondria. PLoS ONE 6:e20220

Bier M, Teusink B, Kholodenko BN, Westerhoff HV (1996) Control analysis of glycolytic oscillations. Biophys Chem 62:15–24

Boiteux A, Goldbeter A, Hess B (1975) Control of oscillating glycolysis of yeast by stochastic, periodic, and steady source of substrate: a model and experimental study. Proc Natl Acad Sci USA 72:3829–3833

Byrd GT, Sage RF, Brown RH (1992) A comparison of dark respiration between C3 and C4 plants. Plant Physiol 100:191–198

Cui P, Ji R, Ding F, Qi D, Gao H, Meng H, Yu J, Hu S, Zhang H (2007) A complete mitochondrial genome sequence of the wild two-humped camel (Camelus bactrianus ferus): an evolutionary history of camelidae. BMC Genomics 8:241–244

Demongeot J, Doncescu A (2009) Modelling the glycolysis. In: IEEE AINA'09. IEEE proceedings, Piscataway, pp 930–935

Demongeot J, Kellershohn N (1983) Glycolytic oscillations: an attempt to an "in vitro" reconstitution of the higher part of glycolysis. Lect Notes Biomaths 49:17–31

Demongeot J, Laurent M (1983) Sigmoidicity in allosteric models. Math Biosci 67:1–17
Demongeot J, Moreira A (2007a) A circular Hamming distance, circular Gumbel distribution, RNA relics and primitive genome. In: Barolli L et al (eds) IEEE AINA' 07. IEEE Press, Piscataway, NJ, pp 719–726
Demongeot J, Moreira A (2007b) A circular RNA at the origin of life. J Theor Biol 249:314–324
Demongeot J, Sené S (2008) Asymptotic behavior and phase transition in regulatory networks. II simulations. Neural Netw 21:971–979
Demongeot J, Seydoux F (1979) Oscillations glycolytiques: modélisation d'un système minimum à partir des données physiologiques et moléculaires. In: Delattre P, Thellier M (eds) Elaboration et justification de modèles. Maloine, Paris, pp 519–536
Demongeot J, Waku J (2012a) Robustness in biological regulatory networks. I mathematical approach. Comptes Rendus Mathématique 350:221–224
Demongeot J, Waku J (2012) Robustness in biological regulatory networks. II application to genetic threshold Boolean random regulatory networks (getBren). Comptes Rendus Mathématique 350:225–228
Demongeot J, Aracena J, Thuderoz F, Baum TP, Cohen O (2003) Genetic regulation networks: circuits, regulons and attractors. CR Biol 326:171–188
Demongeot J, Glade N, Hansen O, Moreira A (2007) An open issue: the inner mitochondrial membrane (IMM) as a free boundary problem. Biochimie 89:1049–1057
Demongeot J, Drouet E, Moreira A, Rechoum Y, Sené S (2009a) MicroRNAs: viral genome and robustness of the genes expression in host. Phil Trans R Soc A 367:4941–4965
Demongeot J, Glade N, Moreira A, Vial L (2009b) RNA relics and origin of life. Int J Mol Sci 10:3420–3441
Demongeot J, Elena A, Noual M, Sené S, Thuderoz F (2011a) Immunetworks. J Theor Biol 280:19–33
Demongeot J, Henrion-Caude A, Lontos A, Promayon E (2011b) General architecture of a genetic regulation network. Applications to embryologic and immunologic control. In: Lenaerts T et al (eds) ECAL'11, advances in artificial life, proceedings of the 11th European conference on the synthesis and simulation of living systems. MIT Press, Cambridge, pp 1–8
Demongeot J, Noual M, Sené S (2012a) Combinatorics of Boolean automata circuits dynamics. Discret Appl Math 160:398–415
Demongeot J, Noual M, Sené S (2012b) Combinatorics of Boolean automata circuits dynamics. Discret Appl Maths 160:398–415
Demongeot J, Hazgui H, Vuillerme N (2013) MicroRNAs: unspecific inhibitory regulation in immunologic control and in mitochondrial respiration. In: Barolli L et al (eds) IEEE AINA' 13. IEEE proceedings, Piscataway, pp 1509–1516
Demongeot J, Cohen O, Doncescu A, Henrion-Caude A (2013) MitomiRrs and energetic regulation. In: Barolli L et al (eds) IEEE AINA' 13. IEEE proceedings, Piscataway, pp 1501–1508
Elkon R, Linhart C, Halperin Y, Shiloh Y, Shamir R (2007) Functional genomic delineation of TLR-induced transcriptional networks. BMC Genomics 8:394
Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO_2 assimilation. Planta 149:78–90
Genc S, Kurnaz IA, Ozilgen M (2011) Astrocyte—neuron lactate shuttle may boost more ATP supply to the neuron under hypoxic conditions—in silico study supported by in vitro expression data. BMC Syst Biol 5:162
Griffiths-Jones S, Marshall M, Khanna A, Eddy SR, Bateman A (2005) Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 33:121–124
Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495:384–388
Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402:47–52
Hervagault JF, Duban MC, Kernevez JP, Thomas D (1983) Multiple steady states and oscillatory behavior of a compartmentalized phosphofructokinase system. Proc Natl Acad Sci USA 80:5455–5459
Hobish MK, Wickramasinghe NSMD, Ponnamperuma C (1995) Direct interaction between amino-acids and nucleotides as a possible physico-chemical basis for the origin of the genetic code. Adv Space Res 15:365–375
Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. PNAS 79:2554–2558

Kaczer H, Burns JA (1973) The control of flux. Symp Soc Exp Bot 28:65–104
Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 22:437–467
Lewin B, Krebs JE, Kilpatrick ST, Goldstein ES (2011) Genes X. Jones & Bartlett, Sudbury
Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW (2009) A microRNA imparts robustness against environmental fluctuation during development. Cell 137:273–282
Liu B, Zhang N, Liu Z, Fu Y, Feng S, Wang S, Cao Y, Li D, Liang D, Li F, Song X, Yang Z (2013) RP105 involved in activation of mouse macrophages via TLR2 and TLR4 signaling. Mol Cell Biochem 378:183–193
Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495:333–338
Miyake K, Ogata H, Nagai Y, Akashi S, Kimoto M (2000) Innate recognition of lipopolysaccharide by Toll-like receptor 4/MD-2 and RP105/MD-1. J Endotoxin Res 6:389–391
Ovadi J (1988) Old pathway-new concept: control of glycolysis. Trends Biochem Sci 13:486–490
Reder C (1988) Metabolic control theory: a structural approach. J Theor Biol 135:175–201
Ritter JB, Genzela Y, Reichl U (2008) Simultaneous extraction of several metabolites of energy metabolism and related substances in mammalian cells: optimization using experimental design. Anal Biochem 373:349–369
Ruoff P, Christensen MK, Wolf J, Heinrich R (2003) Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. Biophys Chem 106:179–192
Sbisa E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140
Surono A, Takeshima Y, Wibawa T, Ikezawa M, Nonaka I, Matsuo M (1999) Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. Hum Mol Genet 8:493–500
Thomas R (1973) Boolean formalisation of genetic control circuits. J Theor Biol 42:563–585
Weaver DC, Workman CT, Stormo GD (1999) Modeling regulatory networks with weight matrices. Pac Symp Biocomp 4:112–123
Weibel ER, Taylor CR, Hoppeler H (1991) The concept of symmorphosis: a testable hypothesis of structure-function relationship. Proc Natl Acad Sci USA 88:10357–10361
Weil G, Heus K, Faraut T, Demongeot J (2004) An archetypal basic code for the primitive genome. Theor Comp Sci 322:313–334
Wolf J, Heinrich R (2000) Effect of cellular interaction on glycolytic oscillations in yeast. Biochem J 345:321–334

2013 27th International Conference on Advanced Information Networking and Applications Workshops

MicroRNAs: unspecific inhibitory regulation in immunologic control and in mitochondrial respiration

J. Demongeot[1,2]*, H. Hazgui[1], N. Vuillerme[1]

Abstract — **The microRNAs are responsible of a post-transcriptional inhibitory activity partly unspecific due firstly to their possible direct negative action during translation by hybridizing the tRNAs especially inside the mitochondrion, hence slowing the mitochondrial respiration, and secondly to the large number of their putative targets like in immunetworks or in cancer control. We will show that in each case the circuits in the core of the interaction graphs are responsible of a few number of dedicated attractors, responsible of the genetically controlled functions, this small amount of possible behaviours being partly due to the microRNA general braking.**

Keywords: MicroRNAs; Genetic regulatory networks; Immunetworks; Mitochondrial respiration; Attractors

I. INTRODUCTION

"Immunetworks" are genetic regulatory networks devoted to the control and maintenance of the immunologic system. Since the innate system of defence represented by the Toll Like Receptors (TLR), already present in insects, mammals have developed an adaptive immune system during the embryonic maturation of their T Cells Receptors α, β and γ (TCRα, TCRβ and TCRγ) from strategies of DNA rearrangements essentially under the control of the RAG gene. We will describe the immunologic networks (called immunetworks) in charge of controlling the concentration of both TLR's and TCR's. The genetic regulatory networks involved in the control of both innate and adaptive immunologic systems (called "immunetworks") have only a few number of asymptotic dynamical behaviours, called attractors. This small number of attractors is directly linked with the possibilities of differentiation of immuno-competent cells and it is controlled inside the interaction graphs of immunetworks by inner circuits [1], giving to the network the possibility to have more than one attractors (due to positive circuits, having inside an even number of inhibitions) and the possibility for these attractors to be stable (due to negative circuits, having inside an odd number of inhibitions), that is to have a sufficient number of initial configurations of gene states giving birth, after a certain dynamical evolution, to the attractor behaviour. We will describe successively these circuits involved in adaptive and innate immunologic systems and give a way to calculate the reduction of the attractor numbers due to the interaction between intersecting circuits and search for their unspecific inhibitory regulation by microRNAs [2-9].

Manuscript received 15th November 2012.
[1]University J. Fourier Grenoble, AGIM CNRS FRE 3405, Faculty of Medicine, 38700 La Tronche, France
*corresponding author: Jacques.Demongeot@agim.eu,
e-mail: Hana.Hazgui@agim.eu

II. THE ADAPTIVE IMMUNETWORKS

The adaptive immune networks are essentially made of three couple of tangential circuits (Fig. 1) concerning essentially the key genes GATA3 (Transcriptional activator binding to DNA sites with the consensus sequence [AT]GATA[AG]) which controls (negatively) T Cells Receptors β (TCRβ), PU.1 (*PUrine*-rich box-1 gene) which controls (negatively) the gene RAG (recombination-activating gene) responsible of the V(D)J rearrangements giving birth to the TCRα receptors, and Zap70 (gene of the Zeta-chain-associated protein kinase 70) which controls (negatively) the TCRβ synthesis. These circuits are inserted in a global immunetwork (Fig. 2), whose attractors are those of the 3 couple of circuits, the rest of the network being essentially reducible to up- and downtrees connected to these circuits.

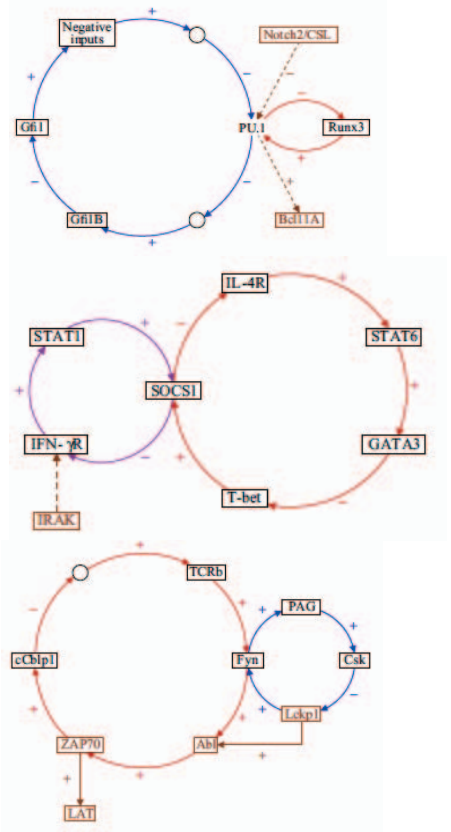


Figure 1. Left: negative 3-circuit tangent to positive 5-circuit controlling GATA3. Middle: negative 6-circuit tangent to negative 2-circuit controlling PU.1. Right: negative 6-circuit tangent to negative 4-circuit controlling ZAP70.

978-0-7695-4952-1/13 $26.00 © 2013 IEEE
DOI 10.1109/WAINA.2013.84

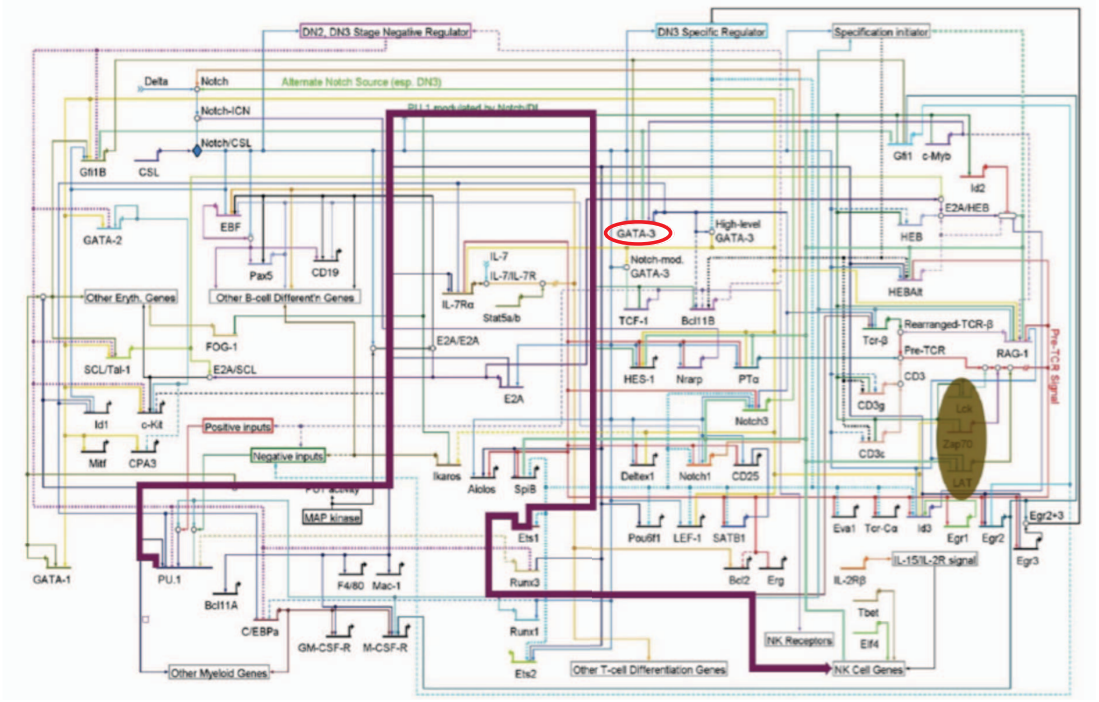


Figure 2. The global immunetwork controlling the genes GATA3 (red circle), PU.1 which activates the Natural Killer cells (violet arrow) and ZAP70 (full brown circle) (after [11]).

ℓ \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	2	2	3	3	5	5	8	10	15	19	31	41	64
2	1	1	2	3	3	4	5	8	10	14	19	31	41	63
3	1	2	1	3	3	6	5	8	8	15	19	33	41	64
4	1	1	2	1	3	4	5	11	10	14	19	24	41	63
5	1	2	2	3	1	5	5	8	10	26	19	31	41	64
6	1	1	1	3	3	1	5	8	8	14	19	63	41	63
7	1	2	2	3	3	5	1	8	10	15	19	31	41	158
8	1	1	2	1	3	4	5	1	10	14	19	24	41	63
9	1	2	1	3	3	6	5	8	1	15	19	33	41	64
10	1	1	2	3	1	4	5	8	10	1	19	31	41	63
11	1	2	2	3	3	5	5	8	10	15	1	31	41	64
12	1	1	1	1	3	1	5	11	8	14	19	1	41	63
13	1	2	2	3	3	5	5	8	10	15	19	31	1	64
14	1	1	2	3	3	4	1	8	10	14	19	31	41	1

p \ n	1	2	3	4	5	6	7	8	9	10	11	12	21	22
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	–	1	–	1	–	1	–	1	–	1	~	1	–	1
3	–	–	2	–	–	2	–	–	2	–	~	2	2	–
4	–	–	–	3	–	–	–	3	–	–	~	3	–	–
5	–	–	–	–	6	–	–	–	–	6	~	–	–	–
6	–	–	–	–	–	9	–	–	–	–	~	9	–	–
7	–	–	–	–	–	–	18	–	–	–	~	–	18	–
8	–	–	–	–	–	–	–	30	–	–	~	–	–	–
9	–	–	–	–	–	–	–	–	56	–	~	–	–	–
10	–	–	–	–	–	–	–	–	–	99	~	–	–	–
11	–	–	–	–	–	–	–	–	–	–	186	–	–	186
12	–	–	–	–	–	–	–	–	–	–	–	335	–	–
21	–	–	–	–	–	–	–	–	–	–	~	–	99858	–
22	–	–	–	–	–	–	–	–	–	–	~	–	–	190557
T_n^+	2	3	4	6	8	14	20	36	60	108	188	352	99880	190746

Figure 3. Top: Calculation of the attractor number due to such intersecting circuits in case of negative of size l and positive of size r circuits intersection (after [9]). Bottom: same calculation for a unique isolated positive circuit.

The Table of Fig. 3 Top shows that a positive 5-circuit tangential to a negative 3-circuit has only 3 attractors, less than an isolated positive circuit which has 8 attractors (Fig. 3 Bottom), showing that for controlling RAG and NK cells, i.e. both adaptive and innate mechanisms, PU.1 needs the attractor in which it is silencing both RAG and NK cells, and two other attractors for which it is inhibiting RAG and co-activating NK cells, and conversely.

2.1. The immunetwork responsible of the Toll Like Receptor (TLR) expression

The natural activation of Natural Killer (*NK) cells, involved in innate immune response,* is controlled by PU.1 gene (Fig. 2) and by *the* ligands of the Toll Like Receptors (TLR). *The* RP105 (CD180) is a TLR-related protein identified in B cells (responsible of the humoral adaptive immune response) [12-14] and acting as regulator of B cell proliferation. B-cells lacking RP105 were shown to be severely impaired in antibody production. The protein ICAM1 is a type o intercellular adhesion molecule continuously present in low concentrations in the membranes of leucocytes involved in humoral adaptive immune response. The network controlling the Toll Like Receptors (TLR) and of ICAM1 expression contains a couple of circuits, one positive 4-circuits tangential to a negative 4-circuits, giving only one attractor (Fig. 4), which corresponds to the activation of TLR2.

2.2. The links with the microRNAS

All the genes introduced in the previous Sections have links with microRNAs exerting a negative control on them and susceptible to decide what attractor will occur, by cancelling their target gene activity.

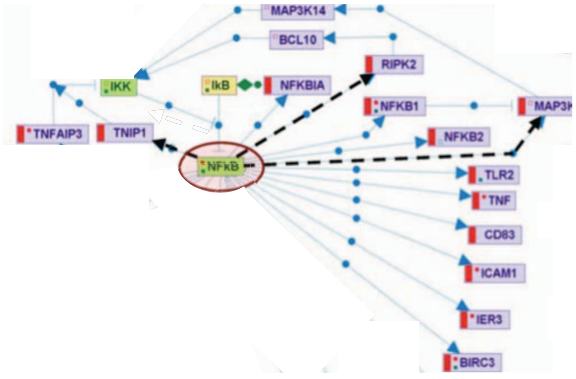


Figure 4. The network controlling the production of the Toll Like Receptors of ICAM1 (after [12]).

For example, we have:
1) for pUNO-hRP105 sub-sequence (4937 bp) of the TLR protein, the hybridization is made by the human microRNA miR 200a [6]:

Sequences Identification Anti-matches
3'-UGUAGCAAUGGUCUGUCACAAU-5' miR 200a
5'-TTGTGC TCATTGAGATGAATGG-3' 12
TRL pUNO-hRP 105 sequence starting in position 531
5'-TACTGCCATTCAAGATGAATGG-3'AL15 matches

2) for the GATA-3 gene, by the human microRNA miR 200c:

Sequences Identification Anti-matches
3'-AGGUAGUAAUGGGC-CGUCAUAA-5' miR 200c
5'- TGCATT TTTGCAGGAGCAGTATC-3' 17
GATA-3 sequence starting in position 57

3) for the ICAM1 sub-sequence CD54 cDNA (1615 bp), by the human microRNA miR 297:

Sequences Identification Anti-matches
3'-GUACGUGUACGUGUGUAUGUA-5' miR 297
5'-CCTCCCCA------CCCACATACAT-3' 15
ICAM1 sequence starting in position 832

We can remark in the above matching that gene expressing the TLR pUNO-hRP105 sub-sequence contains the sequence AGATGAA frequently observed in the not coding genome as part of the CAAGATGAA sequence, belonging both to the Tψ-loop of the tRNAs and to the reference sequence AL [2-4], which signifies in general the affiliation to an ancestral genome.

III. Global Influence of MicroRNAs

3.1. Physiological role in cell energetics

The inhibitory unspecific noise from the microRNAs has a global resultant on the intermediary metabolism of the superior organisms like mammals [15]. Their role is essential in the brake/acceleration balance of the network dynamics preventing artificial attractors to be expressed, which could lead either to pathologic rhythms, or to steady states with too heavy concentrations of certain metabolites.

	V_{O2max}/M (ml.sec^{-1}.kg^{-1})	V/M (ml/kg)
Dog	**2.29**	**40.6**
Goat	**0.95**	**13.8**
Horse	**2.23**	**30.0**
Steer	**0.85**	**11.6**

Figure 5. Comparison between cellular oxygen consumption of mammals [16].

Following the comparative physiology study about cell consumption of oxygen by mammals done in [16], let consider V, the total mitochondrial volume of muscle and V_{O2max}, the O_2 consumption rate per unit mass of whole body mass M of the mitochondria of these mammals. The Fig. 5 shows that non-ruminant animals like horse and dog have an efficacy in O_2 consumption about 2.5 times more than ruminants like steer and dog. We have then studied in the UTR mitochondrial genome of the non-ruminants if there was less mitomiRs candidates than in those of ruminants.

The Fig. 6 shows that such mitomiRs exist and contain long sub-sequences (whose length is between 6 and 7 bases) common with the mitochondrial tRNA loops. On the contrary, these sub-sequences are totally absent in the UTR mitochondrial genome of the non-ruminants. These genomes have been extracted for these different animals from the classical genetic data bases dedicated to the whole mitochondrial [17] and chloroplast [18, 19] genomes: Megasun, from the University of Montreal [17], and Nuccore, from the NCBI [18, 19].

Sub-sequences	Observed bos	Expected bos	Observed equus	Expected equus
TACCAC	2	2	0	1
TACCAT	1	2	0	1
ACC GTT	2	2	0	1
ACCATT	1	2	0	1
CTTGAA	1	2	0	1
ATTTGA A	1	0.5	0	0.25
Total	8	10.5 ± 6.5 (m ± 2σ)	0	5.25 ± 4.6 (m ± 2σ)

Figure 6. Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the UTR mitochondrial genomes of bos (total length 8493 bases) and equus (total length 4535 bases) [17].

From the Figs. 6 to 8, we see that the absence of mitochondrial small sub-sequences, we have called mito^{2}miRs, in the UTR mitochondrial genomes of non ruminants (for a total length of 10024 bases) susceptible to hybridize the mitochondrial tRNAs inside the mitochondrion matrix is highly

significant (observed number equal to 0 and mean expected number 10 with a standard deviation equal to 3.2).

A classical normal test of difference between empiric means shows that the number of ruminant mito[2]miRs is significantly (with a threshold of 10^{-3}) superior to that of non ruminant. The absence of possibility of inhibition of the tRNA transcription inside the non ruminant mitochondria can participate to the better efficacy of the cell respiration in these non ruminants with respect to the ruminants.

Sub-sequences	Observed capra	Expected capra	Observed canis	Expected canis
TACCAC	1	0.4	0	0.8
TACCAT	0	0.4	0	0.8
ACCGTT	1	0.4	0	0.8
ACCATT	1	0.4	0	0.8
CTTGAA	0	0.4	0	0.8
ATTTGAA	0	0.1	0	0.2
Total	3	2.1 ± 2.9	0	4.2 ± 4.1

Figure 7. Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the UTR mitochondrial genomes of capra (total length 1531 bases) and canis (total length 3239 bases) [17].

Sub-sequences	Observed ruminants	Expected ruminants	Observed non ruminants	Expected non ruminants
TACCAC	3	2.4	0	1.9
TACCAT	1	2.4	0	1.9
ACCGTT	3	2.4	0	1.9
ACCATT	2	2.4	0	1.9
CTTGAA	1	2.4	0	1.9
ATTTGAA	1	0.6	0	0.5
Total	11**	12.7 ± 7.1	0**	10 ± 6.3

Figure 8. Number of common sub-sequences of length 6 and 7 between the mitochondrial tRNA loops and the UTR mitochondrial genomes of ruminants (total length 10024 bases) and non ruminants (total length 7774 bases) [17].

In the same spirit, we have analyzed data coming from UTR chloroplast genome [18, 19] of cereals, one belonging to the C_4 plants, the sorghum (Sorghum bicolor), and the other from the C_3 plants, the wheat (Triticum æstivum) which represent the most important kind grown in temperate countries. The C_4 plants are known to have a better efficacy of their chloroplast diurnal respiration than the C_3 plants, as we can see on Fig. 9 [20, 21], where the sorghum is shown to be 1.5 times more efficient that the wheat, by comparing the CO_2 consumption rate per unit surface of whole surface S of the leaves of these cereals.

	V_{CO2max}/S mmol.sec^{-1}.m^{-2}
Sorghum bicolour	**41.1 ± 2.8**
Triticum æstivum	**27.9 ± 1**

Figure 9. Comparison between cellular CO_2 diurnal consumption between 2 cereals, sorghum and wheat [91].

In these cereals, the CO_2 assimilation is usually limited by the capacity of photosynthetic electron transport to supply ATP and NADPH to regenerate RuBP [20, 21]. Then we can compare as for the animal mitochondria the presence of sub-sequences common to the UTR chloroplast genome and to the chloroplast tRNA loops.

Sub-sequences	Observed sorghum	Expected sorghum	Observed triticum	Expected triticum
TACCACT	0	1	0	1
TACCATT	4	1	4	1
TACCGCT	0	1	0	1
TACCGTT	1	1	1	1
ATTTGAA	3	1	4	1
GTTTGAA	1	1	5	1
ATTCGAA	3	1	5	1
GTTCGAA	3	1	3	1
ACTTGAA	0	1	2	1
GCTTGAA	0	1	1	1
ACTCGAA	3	1	3	1
GCTCGAA	1	1	1	1
Total	19**	11.7 ± 6.8 (m ± 2σ)	29**	11.6 ± 6.8 (m ± 2σ)

Figure 10. Number of common sub-sequences of length 6 and 7 between the chloroplast tRNA loops and the UTR chloroplast genomes of sorghum (total length 15852 bases) and triticum (total length 16014 bases).

From the Fig. 10, we see that the number of chloroplast small sub-sequences, we are calling chloromiRs, in the UTR chloroplast genome of a C_4 plant, the sorghum (Sorghum bicolor, for a total length of 15852 bases) susceptible to hybridize the chloroplast tRNAs inside the chloroplast matrix is highly significantly less than the same number observed in a C_3 plant, the wheat (Triticum æstivum, for a total length of 16014 bases).

A classical normal test of difference between empiric means shows that the number of the C_4 plant chloromiRs is significantly (with a threshold of 10^{-3}) less than that of the C_3 plant. The diminution of possibility of inhibition of the tRNA transcription inside the chloroplast of the sorghum can participate to the better efficacy of the cell diurnal respiration in this C_4 plant with respect to the less efficient C_3 plant, the wheat.

3.2. Pathological role of microRNAs in cancer

The role of the metabolic pathways like glycolysis or oxidative phosphorylation in cancerogenesis have been emphasized these last years [22-32], like the revisit of a the Warburg hypothesis, which postulates that cancerogenesis could be favoured by an insufficient cellular respiration. This hypothesis came from the observation that most cancer cells produce energy produced from a rate of the high part glycolysis followed by lactic fermentation greater than the rate of the low part of glycolysis followed by oxidative phosphorylation.

For example, we can notice in patients with melanoma the presence of microRNAS like miR-221 reducing the isomerase activity in melanoma cells [30], hence increasing the high glycolysis rate (hence favouring the pentose phosphate pathway and the Warburg effect [24, 25]) as well as the proliferative growth rate of the melanoma cells by targeting of c-kit, p27 and p57, and in the patients with prostate cancer the presence of microRNAs like miR-34a, repressing the oxidative phosphorylation through the enzymes Translocase and ATPase (hence favouring the Warburg effect) and the cell cycle through E2F and CD44, inhibits the prostate cancer stem cells and prevents metastasis [31, 32].

In the numerous papers devoted to microRNAs and cancer, we can notice the possibility of an under-expression of the microRNA has-miR-320 which reduces the PFK activity [22], favouring the an high glycolysis rate. Associated like in the melanoma data with a blockage of the oxidative phosphorylation or with an over-expression of hsa-miR-19a, which inhibits the pyruvate dehydrogenase, we could have another example reinforcing the Warburg hypothesis.

Another example of miR-dependent cancer is the Hepatocellular Carcinoma (HCC), the major primary liver cancer. The glypican-3 (GPC3) is one of the most abnormally expressed genes in this cancer, which could play a role in liver carcinogenesis. In [26, 27] using a a functional screening, the authors found that miR-96, miR-129-1-3p, miR-219-5p miR-1271, miR-1291 and miR-1303 differentially control GPC3 expression in HCC cells. More precisely, miR-219-5p exerts tumor-suppressive effects in hepatic carcinogenesis through its negative regulation of GPC3 expression. Eventually, we can notice the role of the tumor suppressors like p53, p63, and p73, which are highly controlled by and controlling numerous micro-RNAs (cf. Fig. 11 and [23, 28, 29]). The role we have already commented of p53 in the dual regulation of the apoptosis versus proliferation processes is probably the main origin of its major influence in cancerogenesis and tumoro-suppression, depending on its up- or down-regulation by the microRNAs.

IV. General architecture of a genetic regulatory network and problem of its robustness

The general architecture of a genetic regulatory network is given like in Figs. 2, 4 or 11 by a digraph framework in which the nodes sources are often micro-RNAs exerting their partly unspecific basic inhibitory influence, the nodes sink are are important genes controlling vital functions like the RAG responsible for the TCR building in the immune system (cf. Fig. 2 and [10]).

The network robustness (or resilience) [33] can be defined as the capacity to return at its ordinary asymptotic dynamical behavior (called attractor) after endogenous or exogenous perturbations affecting:

- the state of certain of its genes, *e.g.*, their specific silencing by micro-RNAs
- the state of its boundary, notably the appearance of new regulations from a mutation in the UTR genome giving birth to new micro-RNAs
- its architecture, by creating new links between proteins needing to taken into account non-linear interactions [34].

A study about the influence of the microRNAs on the robustness of a network needs the exact counting of the dynamical attractors of this network, then to know for example the reduction or on the contrary the amplification factor caused for example respectively by the circuit opening due to the effective total inhibition of a gene (until its possible functional knock-out) by a microRNA, or by the appearance of new genes, which corresponds to important architectural perturbations. A parameter devoted to represent the robustness of a biological network is its evolutionary entropy (cf. [34-38]), defined as the Kolmogorov-Sinaï entropy of the Markov process underlying the gene states updating. The microRNAs can have a double opposite influence on this parameter, causing its increase (and hence that of the network robustness) when they create an unspecific inhibitory “noise”, dispatching into the gene states space the probability to have more possible phenotypes, and they can also provoke a decrease of the robustness, by opening circuits of the strong connected components of the network, hence by diminishing the number of circuits, responsible of the attractor number of the whole network.

The example of the Fig. 11 shows that we can count the attractor number of the main sub-network of a functional network like that organized around the gene Engrailed. These attractor numbers are given in blue on the Fig. 11, as well as their nature (steady state or limit-cycle). The global attractor of the whole network is more difficult to get, in particular when two sub-networks create a new strong connected component, like in the case of the MPK and Cytoskeleton sub-networks in Fig. 11, which takes a new level of complexity needing further theoretical researches to be understood and solved.

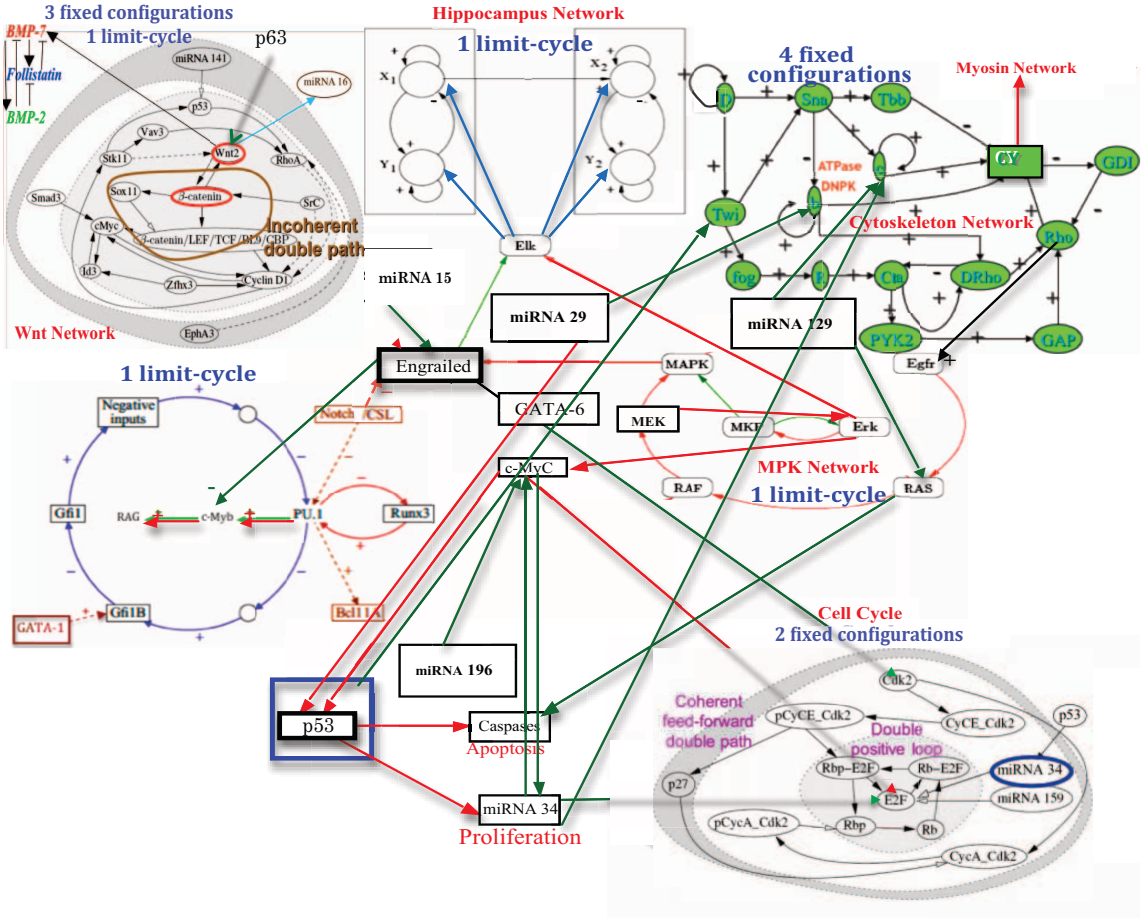


Figure 11. Engrailed network with its regulatory environment of up or down-sub-networks. The number and nature of attractors of sub-networks are indicated in blue.

The dynamics of the Engrailed network can be explained in terms of non-linear interactions [34] and of state dependent updating rule. Indeed, when multiple interactions are occurring on the same gene, like ERF, in the strong connected component at the centre of the cell cycle sub-network, it is possible to model in a non-linear way the interactions between the effectors (notably miR 34 and c-MyC) in order to take into account their mutual influence (here a mutual inhibition) which causes a global effect which is not the algebraic addition of their marginal actions, but either a potentialising effect (if effectors are in synergy or agonist) or conversely a counteracting effect (if they are antagonist, like in the ERF case). Another important feature of the dynamics is the existence of genes influencing directly the opening of the DNA inside of chromatine, hence allowing or not the gene expression. If these genes are controlled by microRNAs [49], it is necessary to generalize the getBren structure by considering that the possibility to update a block of genes at iteration t is depending on the state of r "clock" genes (*i.e.*, involved in the chromatine updating clock) $k_1,\ldots k_r$ (like histone acetyltransferase, endonucleases, exonucleases, helicase, replicase, polymerases) depending on s microRNAs, $l_1,\ldots,l_s$. Then the transition for a gene i, such as i does not belong to $\{k_1,\ldots k_r\}$, could be written as:

$\forall g \in \{0,1\}$, $\beta \in \{0,1\}^n$, if $\forall\, j=1,\ldots,r$, $x_{kj}(t) = 1$, then

$P_{i,g^\beta}(\{x_i(t+1)=g \,|\, x(t)=\beta\}) =$

$\exp[g(\Sigma_{j\in Ni} w_{ij}\beta_j-\theta_i)/T]/[1+\exp[(\Sigma_{j\in Ni} w_{ij}\beta_j-\theta_i)/T]$

(involving $\forall\, j=1,\ldots,\ s,\ x_{lj}(t\text{-}1) = 0$, if the microRNAs $l_1,\ldots,l_s$ are dominant),

and:

$P_{i,g^\beta}(\{x_i(t+1)=\beta_i \,|\, x(t)=\beta\})=1$, if not.

To make the rule more precise, we could for the sake of simplicity, decide that the indices $k_1,\ldots k_r$ of the r "clock" genes are $1,\ldots, r$ and that we can have the 3 possible following behaviours:

i) if $y(t) = \Pi_{i=1,\ldots,r} x_i(t) = 1$, then the above rule is available
ii) if $y(t)=0$ and $\Sigma_{s=t,\ldots,t-c} y(s) > 0$, then $x_i(t+1)=x_i(t-s^*)$, where s^* is the last time before t, where $y(s^*) = 1$
iii) if $y(t)=0$ and $\Sigma_{s=t,\ldots,t-c} y(s) = 0$, then $x_i(t+1)=0$.

The dynamical system remains autonomous (with respect to the time t, *i.e.*, depends on t only through the set of state variables $\{x_i(t-r),\ldots, x_i(t-1)\}$), but a theoretical study of its attractors (like in [50]), with this state dependent updating schedule is very difficult to perform and will be investigated further.

V. Conclusion

The immunetworks are showing interesting properties of attractor numbers reduction, necessary to control in general two main attractors, one inhibiting the function (the brake) and the other activating it (the accelerator). The role of the microRNAs is to provide an unspecific inhibitory noise leaving only the circuits having sufficiently strong interactions to be able to express their attractors. In perspective, we will study systematically all the networks involved in the immunologic defence, in order to check the possible generalization of these observations done only on two reference networks controlling both innate and adaptive immunologic systems in mammals.

References

[1] Demongeot, J.; Henrion-Caude, A.; Lontos, A.; Promayon, E. General architecture of a genetic regulation network. Applications to embryologic and immunologic control. In: Lenaerts, T.; Giacobini, M.; Bersini, H.; Bourgine, P.; Dorigo, M.; Doursat, R. (Eds). ECAL'11, *Advances in Artificial Life, Proceedings of the 11th European Conference on the Synthesis and Simulation of Living Systems*. MIT Press, Cambridge MA, pp. 1-8; 2011.
[2] Demongeot, J.; Moreira, A. A circular RNA at the origin of life. *J. Theor. Biol.* **249**:314-324; 2007.
[3] Demongeot, J.; Drouet, E.; Moreira, A.; Rechoum, Y.; Sené, S. MicroRNAs: viral genome and robustness of the genes expression in host. *Phil. Trans. Royal Soc. A* **367**:4941-4965; 2009.
[4] Demongeot, J.; Glade, N.; Moreira, A.; Vial, L. RNA relics and origin of life. *Int. J. Molecular Sciences* **10**:3420-3441; 2009.
[5] Griffiths-Jones, S.; Marshall, M.; Khanna, A.; Eddy, S.R.; Bateman, A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**:121–124; 2005.
[6] http://mirdb.org/miRDB/
[7] http://mirnamap.mbc.nctu.edu.tw/
[8] Demongeot, J.; Aracena, J.; Thuderoz, F.; Baum, T.P.; Cohen, O. Genetic regulation networks: circuits, regulons and attractors. *C. R. Biologies* **326**:171-188; 2003.
[9] Demongeot, J.; Noual, M.; Sené, S. Combinatorics of Boolean automata circuits dynamics. *Discrete Applied Mathematics* **160**:398-415; 2012.
[10] Demongeot, J.; Elena, A.; Noual, M.; Sené, S., Thuderoz, F. "Immunetworks", attractors and intersecting circuits. *J. Theor. Biol.* **280**:19-33; 2011.
[11] Georgescu, C.; Longabaugh, W.J.R.; Scripture-Adams, D.D.; David-Fung, E.S.; Yui, M.A.; Zarnegar, M.A.; Bolouri, H.; Rothenberg, E.V. A gene regulatory network armature for T lymphocyte specification. *Proceedings of the National Academy of Sciences USA* **105**:20100–20105; 2008.
[12] Elkon, R.; Linhart, C.; Halperin, Y. ; Shiloh, Y.; Shamir, R. Functional genomic delineation of TLR-induced transcriptional networks. *BMC Genomics* **8**:394; 2007.
[13] Bulet, P.; Hetru, C.; Dimarcq, J.L.; Hoffmann, D. Antimicrobial peptides in insects; structure and function. *Dev. and Comparative Immunology* **23**:329-344; 1999.
[14] Miyake, K.; Ogata, H.; Nagai, Y.; Akashi, S.; Kimoto, M. Innate recognition of lipopolysaccharide by Toll-like receptor 4/MD-2 and RP105/MD-1. *J Endotoxin Res.* **6**:389-391; 2000.
[15] Bandiera, S.; Rüberg, S.; Girard, M; Cagnard, N.; Hanein, D.; Chrétien, D.; Munnich, A.; Lyonnet, S.; A. Henrion-Caude, A. Nuclear Outsourcing of RNA Interference Components to Human Mitochondria. *PLoS ONE* **6**:e20746; 2011.
[16] Weibel, E.R.; Taylor, C.R.; Hoppeler, H. The concept of symmorphosis: A testable hypothesis of structure-function Relationship. *Proc. Natl. Acad. Sci. USA* **88**:10357-10361; 1991.
[17] http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list .html
[18] http://www.ncbi.nlm.nih.gov/nuccore/EF115542
[19] http://www.ncbi.nlm.nih.gov/nuccore/AB042240
[20] Byrd, G.T.; Sage, R.F.; Brown, R.H. A Comparison of Dark Respiration between C3 and C4 Plants. *Plant Physiol.* **100**:191-198; 1992.
[21] Farquhar G.D.; von Caemmerer S.; Berry J.A. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta* **149**:78-90; 1980.
[22] Tang, H.; Lee, M.; Sharpe, O.; Salamone, L.; Noonan, E.J.; Hoang, C.D.; Levine, S.; Robinson, W.H.; Shrager, J.B. Oxidative stress-responsive microRNA-320 regulates glycolysis in diverse biological systems. *FASEB J.* **26**: 4710-4721; 2012.
[23] Hainaut, P.; Hollstein, M. p53 and human cancer: the first ten thousand mutations. *Adv. Cancer Res.* **77**:81–137; 2000.
[24] Demetrius, L.A.; DK Simon, D.K. An inverse-Warburg effect and the origin of Alzheimer's disease. *Biogerontology* doi: 10.1007/s10522-012-9403-6; 2012.
[25] Davies, P.C.; Demetrius, L.; Tuszynski, J.A. Cancer as a dynamical phase transition. *Theor. Biol. Med. Model.* **8**:30; 2011.
[26] Maurel, M.; Jalvy, S.; Lareiro, Y.; Combe, C.; Vachet, L.; Sagliocco, F.; Bioulac-Sage, P.; Pitard, V.; Jacquemin-Sablon, H.; Zucman-Rossi, J.; Laloo, B.; Grosset, C.F. A functional screening identifies five miRNAs controlling glypican-3: Role of miR-1271 down-regulation in hepatocellular carcinoma. *Hepatology* doi: 10.1002/hep.25994; 2012.
[27] Huang, N; Lin, J.; Ruan, J.; Su, N; Qing, R.; Liu, F.; He, B.; Ly, C.; Zheng, D.; Luo, R. MiR-219-5p inhibits hepatocellular carcinoma cell proliferation by targeting glypican-3. *FEBS Letters* **586**:884-891; 2012.
[28] Boominathan, L. The Tumor Suppressors p53, p63, and p73 Are Regulators of MicroRNA Processing Complex. *PLoS ONE* **5**:e10615; 2010.
[29] Boominathan, L. The guardians of the genome (p53, TA-p73, and TA-p63) are regulators of tumor suppressor miRNAs network. *Cancer Metastasis Rev.* **29**:613-639; 2010.
[30] Segura, M.F.; Greenwald, H.S.; Hanniford, D.; Osman, I.; Hernando, E. MicroRNA and cutaneous melanoma: from discovery to prognosis and therapy. *Carcinogenesis* **33**:1823-1832; 2012.
[31] Tennant, D.A.; Dur, R.V.; Gottlieb, E. Targeting metabolic transformation for cancer therapy. *Nature Reviews Cancer* **10**:267–277; 2010.
[32] Liu, C.; Kelnar, K.; Liu, B.; Chen, X.; Calhoun-Davis, T.; Li, H.; Patrawala, L.; Yan, H.; Jeter, C.; Honorio, S.; Wiggins, J.F.; Bader, A.G.; Fagin, R.; Brown, D; Tang, D.G. The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nature Medicine* **17**:211-215; 2011.

[33] Blanchini, F.; Franco, E. Structurally robust biological networks. *BMC Systems Biology* **5**:74; 2011.
[34] Demongeot, J.; Sené, S. The singular power of the environment on nonlinear Hopfield networks. In: *CMSB'11*. ACM Proceedings, New York; pp. 55-64; 2011.
[35] Demongeot, J.; Jezequel, C.; Sené, S. Asymptotic behavior and phase transition in regulatory networks. I Theoretical results. *Neural Networks* **21**:962-970; 2008.
[36] Demongeot, J.; Sené, S. Asymptotic behavior and phase transition in regulatory networks. II Simulations. *Neural Networks* **21**:971-979; 2008.
[37] Demetrius, L. Statistical mechanics and population biology. *J. Stat. Physics* **30**:709–753; 1983.
[38] Demetrius, L. Directionality principles in thermodynamics and evolution. *Proc. Natl. Acad. Sci. USA* **9**:3491-3498; 1997.
[39] Kühn, R. *Equilibrium Analysis of Complex Systems*. Lecture Notes 7CCMCS03, King's College, London; 2010
[40] Fogelman Soulié F.; Goles E.; Martinez S.; Mejia C. Energy function in neural networks with continuous local functions. *Complex Systems* **3**:269-293; 1989.
[41] Cosnard M.; Goles E. Discrete states neural networks and energies. *Neural Networks* **10**:327-334; 1977.
[42] Lesne, A. Robustness: confronting lessons from physics and biology. *Biol. Rev. Cambridge Philos. Soc.* **83**:509-532; 2008.
[43] Gunawardena, S. The robustness of a biochemical network can be inferred mathematically from its architecture. *Biological Systems Theory* **328**:581-582; 2010.
[44] Waddington, C.H. *Organizers & Genes*. Cambridge University Press, Cambridge UK; 1940.
[45] Thom, R. *Structural stability and fluctuations and morphogenesis*. Benjamin, Reading Mass; 1972.
[46] Cinquin, O.; Demongeot, J. Positive and negative feedback : striking a balance between necessary antagonists. *J. Theor. Biol.* **216**:229-241; 2002.
[47] Demongeot, J.; Noual, M.; Sené, S. Combinatorics of Boolean automata circuits dynamics. *Discrete Applied Mathematics* **160**:398-415; 2012.
[48] Demongeot, J.; Noual, M.; Sené, S. On the number of attractors of positive and negative Boolean automata circuits. In: Chang, E., Barolli, L. (Eds). *IEEE AINA' 10*. IEEE Proceedings, Piscataway; pp. 782-789; 2010.
[49] Demongeot, J.; Cohen, O.; Doncescu, A.; Henrion-Caude, A. MitomiRs and energetic regulation. In: Barolli, L. et al. (Eds). *IEEE AINA' 13*. IEEE Proceedings, Piscataway; in the present volume.
[50] Demongeot, J.; Goles, E.; Morvan, M.; Noual M.; Sené, S. Attraction Basins as Gauges of Environmental Robustness in Biological Complex Systems. *PloS ONE* **5**:e11793; 2010.

Inhibitory Regulation by microRNAs and Circular RNAs

J. Demongeot, H. Hazgui, J. Escoffier, and C. Arnoult

I AGIM CNRS FRE 3405, Faculty of Medicine, University J. Fourier of Grenoble, 38B700 La Tronche, France

Abstract—**The microRNAs are responsible of a general post-transcriptional inhibitory activity partly unspecific due firstly to a possible direct negative action during translation by hybridizing tRNAs, and secondly to the large number of their putative targets. Recently a second regulation layer has been discovered involving both proteins as microRNA transcription factors and circular RNAs as microRNA inhibitors ("miRs sponges"). We will show the existence of circuits inside the role of miR-7 in the regulation of the spermatozoa morphogenesis.**

Keywords—**MicroRNAs, Genetic regulatory networks, Cyclic RNAs inhibition, spermatozoa morphogenesis regulation.**

I. Introduction

The genetic regulatory networks involved in the control of morphogenesis have only a few number of asymptotic dynamical behaviors, called attractors. This small number of attractors is directly linked with the possibilities of differentiation of the cells in the tissue in construction, *e.g.* the spermatozoa, and this number is controlled inside the interaction graph of the regulatory network by the circuits of its strongly connected components [1-4], giving to the network the possibility to have if necessary, when the tissue has more than one function, more than one attractors (due to positive circuits, having inside an even number of inhibitions) and the ability to be stable (due to negative circuits, having inside an odd number of inhibitions). We will describe successively a recently discovered regulation due to circular RNAs, inhibiting microRNAs (miRNAs) responsible of a partly unspecific inhibition in genetic networks, then the network regulating the spermatozoa morphogenesis in its controlling environment, and eventually, a method to calculate the reduction of the attractor numbers due to interactions between intersecting circuits and apply it to the control of the gene Dpy9l2, responsible for the morphology of the acrosome in spermatids, involved in fecundation.

II. Circular RNAs

Mature messenger and transport RNAs are inhibited by miRNAs in a partly unspecific way due to the great number of their putative targets [5,6] and other RNAs play also a regulatory role by hybridizing miRNAs, the circular RNA sequences (ciRS) discovered firstly in plants and now unambiguously validated in animals: in [7,8] authors found that a human ciRS, ciRS-7, antisense to the cerebellum degeneration-related protein CDR1, was binding sites of the miRNA, miR-7, perfectly conserved from annelids to humans, with practically the same role in protein translation regulation, despite of the fact that for example it regulates in *Drosophila* some proteins not regulated in mammals [9]. miR-7 functions in several interlocking feedback and feed-forward loops, and we have chosen its role in a negative circuit involving the couple of genes Egfr/RAS, increasing the stability of the MAPK network, *i.e.* buffering them against external perturbations. Conserved miRNAs like miR-7 may increase robustness of genetic regulatory networks. Hence a second layer of regulation could serve to prevent an over-inhibition by such critical miRNAs, leading to pathologic down-regulations. On Fig. 1, we can see good matches between a sequence called AGAT homologous from 13 species to a human CDR1 sequence using BLAT and the UCSC Genome browser [7]. This sequence AGAT matches the half part of the ABEL sequence, proposed as an archetypal sequence for the ancient genetic code [10-12]. The main characteristic of ABEL is that it contains from start codon AUG until end codon UGG one and only one codon per synonymy class of amino-acids [10]. ABEL is close to another sequence AL, which can be considered as a the barycentre of the miRNA set for Hamming distance, and AL shares with any miRNA subsequence of lengh 5 (like UUCCA) more than expected by chance [11], and is close to sequences coming from the UTR viral genomes [12].

ABEL sequence 5'-AUGGCACUGAAUGUUCCAAGAU-3'
AGAT sequence 5'-GTCTTCCAAGAT-3'
hsa-miR-7 5'-AACAAAAUCACUAGUCUUCCA-3'

Fig. 1 Matches between the ABEL, AGAT and human miR-7 sequences

The antisense sequence 3'-5' of AGAT anti-matches the corresponding sequence 5'-3' of miR-7, which explains the inhibitory power of ciRS-7. By searching in miRNAs bases [6], we found the sequence UCUUCCA in about 25 of among 500 miRNAs, significantly more than expected (about 2±4). In [5], are reconstructed 16S and 23S rRNA data sets and trees spanning all sequenced type strains of classified species of Archaea and Bacteria, and we found inside these RNAs 249 times the sequence AUCUUCCA, significantly more than expected frequency (about 160±40).

L.M. Roa Romero (ed.), *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*,
IFMBE Proceedings 41,
DOI: 10.1007/978-3-319-00846-2_309,

III. THE SPERMATOZOA MORPHOGENESIS REGULATION

The Dpy19l2 protein is expressed in spermatids with a specific localization on the inner nuclear membrane facing the acrosomal vesicle [15,16]. The absence of Dpy19l2 leads to the destabilization of both the the junction between the acroplaxome and the nuclear envelope. Consequently, the acrosome fails to be linked to the nucleus leading to the failure of sperm nuclear shaping and eventually to the elimination of the unbound acrosomal vesicle rendering impossible the fecundation. Hence, the control of Dpy19l2 has to be very reliable and Figs. 2 and 3 show the corresponding regulatory network in its genomic environment (after [17]) and its possible dynamics (after [2]).

IV. NUMBER OF ATTRACTORS

The different components of the sub-networks of the Fig. 3 have been extracted from the recent literature, in particular those concerning the inhibitory action of the miRNAs [19-26]. Let us remark that the gene Dpy19la is inhibited by miR-146a, whose dynamics is dependent on two negative circuits, one of size 12 coming from the miR-146a and passing through the couple of genes Egfr/RAS (counted only for one node) and the gene p21, and the other of size 4 coming from the miR-7 passing through the node Egfr/RAS, the gene RAF and the couple of genes MEK/ERK (counted only for one node). These two negative circuits are tangent on the node Egfr/RAS.

The number of attractors of the sub-networks of the Fig. 3 can be calculated by using an algorithm presented in [2], by supposing that each gene can have only two states, 1 if the gene is expressed and 0 if not. The central elements of the network architecture fixing the number of possible asymptotic dynamical behavior of the gene states (called attractor), when the updating schedule is supposed to be synchronous, are the circuits of the strongly connected components of the network. Using the results of [2], it is for example possible to calculate the number of attractors of any sub-network of the Fig. 3. For example, at the intersection of the line 12 and of the column 4, we meet the integer 2, which is the number of the attractors of the couple of two tangential negative circuits of size respectively 12 and 4. One of the two attractors corresponds to both miR-146a and ciRS-7 expressed. Then miR-7 is not expressed, but thegenes Egfr/RAS, RAF and MEK/ERK of the negative circuit of length 4 as well as the gene Dpy19la are not expressed. This attractor corresponds to a pathologic spermatozoa morphogenesis. On the contrary, when ciRS-7 is expressed, but miR-146a is inhibited by c-Myc, then the genes of the negative circuit of length 4 are expressed as well as the gene Dpy19la, and the corresponding attractor leads to a physiologic spermatozoa morphogenesis.

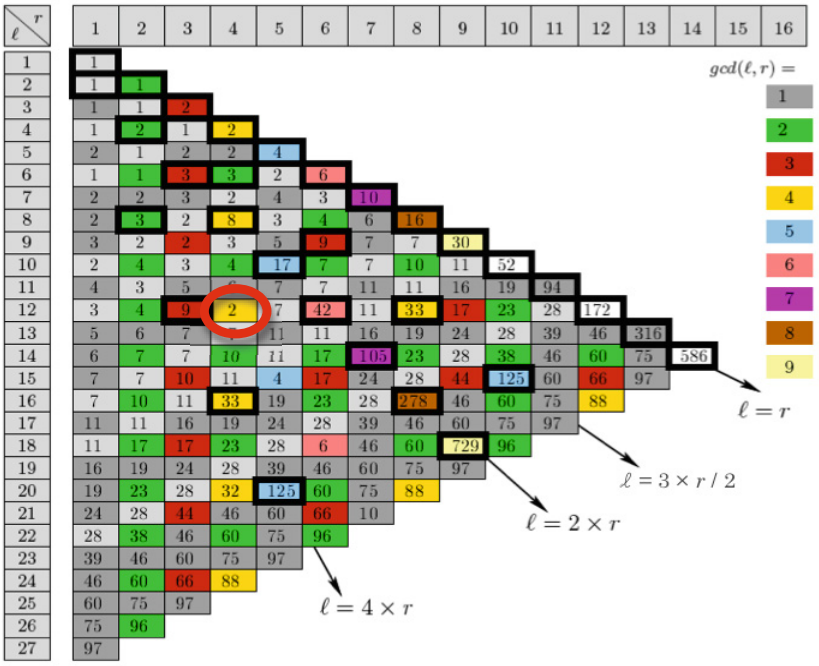


Fig. 2 Attractors number of tangential negative circuits of size l and r. Colors indicate the different values of the greatest common divisor (gcd).

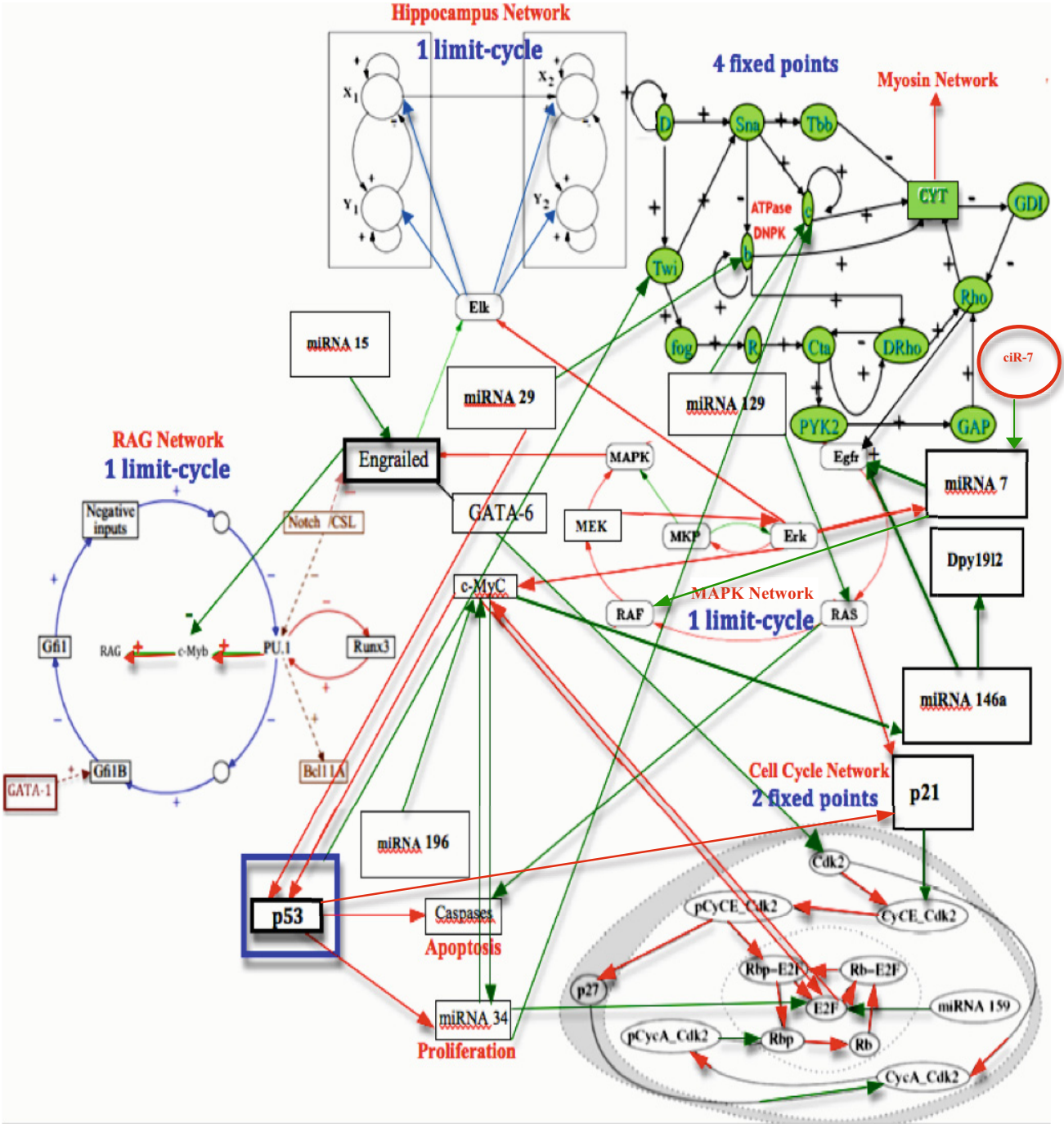


Fig. 3 Genetic network controlling the gene Dpy19la with its environment centred on the gene Engrailed with its up and down-sub-networks. The number and nature of attractors of these sub-networks are indicated in blue. Bottom left: interaction signed digraph modelling the genetic regulation network controlling the cell cycle in mammals [18]. Black (resp. white) arrows represent activations (resp. inhibitions).

The size of the genetic regulatory networks is increasing and the complexity of their architecture leads to use recent mathematical approaches on the dynamics of the Boolean networks, in order to identify their attractors (their number and their nature). The recent discovery of two layers of gene inhibition, due respectively to the circular RNAs and to the micro-RNAs, partly unspecific, renders this mathematical framework still more necessary. Surprisingly, the addition of a great number of such inhibitory RNAs is in general reducing the number of possible asymptotic configurations of the genetic regulatory networks and increases their stability and robustness [2-4], leading to biological functions with a few number of physiologic and pathologic phenotypes, hence focusing a tissue on a specific well controlled role inside the whole organism, resisting to environmental external perturbations, like here the spermatozoa morphogenesis.

ACKNOWLEDGMENT

The present work has been supported by the ANR Project REGENER.

REFERENCES

1. Cinquin O, Demongeot J (2002) Positive and negative feedback : striking a balance between necessary antagonists. J Theor Biol 216:229-241
2. Demongeot J, Noual M, Sené S (2012) Combinatorics of Boolean automata circuits dynamics. Discrete Appl Math 160:398-415
3. Demongeot J, Cohen O, Doncescu A, Henrion-Caude A (2013) MitomiRs and energetic regulation, IEEE Proc., AINA' 13, pp 1501-1508
4. Demongeot J, Goles E, Morvan M, Noual M, Sené S (2010) Attraction Basins as Gauges of Environmental Robustness in Biological Complex Systems. PloS ONE 5:e11793
5. http://mamsap.it.deakin.edu.au/~amitkuma/mirna_targetsnew/sequence.html
6. http://mirdb.org/miRDB/
7. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak S D, Gregersen L H, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495:333-338
8. Hansen T B, Jensen T I, Clausen B H, Bramsen J B, Finsen B, Damgaard C K, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495:384-388
9. Li X., Cassidy J.J., Reinke C.A., Fischboeck S., Carthew R.W. (2009) A microRNA Imparts Robustness Against Environmental Fluctuation During Development. Cell 137: 273–282
10. Weil G, Heus K, Faraut T, Demongeot J (2004) An archetypal basic code for the primitive genome. Theoret Comp Sc 322:313-334
11. Demongeot J, Moreira A (2007) A circular RNA at the origin of life. J Theor Biol 249:314-324
12. Demongeot J, Drouet E, Moreira A, Rechoum Y, Sené S (2009) MicroRNAs: viral genome and robustness of the genes expression in host. Phil Trans Royal Soc A 367:4941-4965
13. Kumar A, Wong A K, Tizard M L, Moore R J, Lefèvre C (2012) miRNA_Targets: a database for miRNA target predictions in coding and non-coding regions of mRNAs. Genomics 100:352-356
14. http://www.arb-silva.de/fileadmin/silva_databases/living_tree/ LTP_release_93/type_strains_LTP_s93_unaligned.fasta
15. Pierre V, Martinez G, Coutton C, Delaroche J, Yassine S, Novella C, Pernet-Gallay K, Hennebicq S, Ray PF, Arnoult C (2012) Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. Development 139:2955-2965
16. Escoffier J, Boisseau S, Serres C, Cheng C C, Kim D, Stamboulian S, Shin H S, Campbell K P, De Waard M, Arnoult C (2007) Expression, localization and functions in acrosome reaction and sperm motility of $Ca_V3.1$ and $Ca_V3.2$ channels in sperm cells: An evaluation from $Ca_V3.1$ and $Ca_V3.2$ deficient mice. J Cell Physiol 212: 753-763
17. Demongeot J, Pempelfort H, Martinez J M, Vallejos R, M. Barria M, Taramasco C (2013) Information design of biological networks: application to genetic, immunologic, metabolic and social networks., IEEE Proc., AINA' 13, pp 1533-1540
18. Kohn K W (1999). Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. Molecular Biology of the Cell 10:2703–2734.
19. Maurel M, Jalvy S, Lareiro Y, Combe C, Vachet L, Sagliocco F, Bioulac-Sage P, Pitard V, Jacquemin-Sablon H, Zucman-Rossi J, Laloo B, Grosset C F (2013) A functional screening identifies five miRNAs controlling glypican-3: Role of miR-1271 down-regulation in hepatocellular carcinoma. Hepatology 57:195-204
20. Huang N, Lin J, Ruan J, Su N, Qing R, Liu F, He B, Lv C, Zheng D, Luo R (2012) MiR-219-5p inhibits hepatocellular carcinoma cell proliferation by targeting glypican-3. FEBS Letters 586:884-891
21. Boominathan L (2010) The Tumor Suppressors p53, p63, and p73 Are Regulators of MicroRNA Processing Complex. PLoS ONE 5:e10615
22. Boominathan L (2010) The guardians of the genome (p53, TA-p73, and TA-p63) are regulators of tumor suppressor miRNAs network. Cancer Metastasis Rev 29:613-639
23. Segura M F, Greenwald H S, Hanniford D, Osman I, Hernando E (2012) MicroRNA and cutaneous melanoma: from discovery to prognosis and therapy. Carcinogenesis 33:1823-1832
24. Tennant D A, Dur R V, Gottlieb E (2010) Targeting metabolic transformation for cancer therapy. Nature Reviews Cancer 10:267-277
25. Georgescu C, Longabaugh W J R, Scripture-Adams D D, David-Fung E S, Yui M A, Zarnegar M A, Bolouri H, Rothenberg E V (2008) A gene regulatory network armature for T lymphocyte specification. Proceedings of the National Academy of Sciences USA 105:20100–20105
26. Elkon R, Linhart C, Halperin Y, Shiloh Y, Shamir R (2007) Functional genomic delineation of TLR-induced transcriptional networks. BMC Genomics 8:394
27. Jiang L, Liu X, Chen Z, Jin Y, Heidbreder C E, Kolokythas A, Wang A, Dai Y, Zhou X (2010) MicroRNA-7 targets IGF1R (insulin-like growth factor 1 receptor) in tongue squamous cell carcinoma cells. Biochem J 432:199–205

CHAPTER

Genetic Regulatory Networks: Focus on Attractors of Their Dynamics

7

J. Demongeot[1], H. Hazgui[1], and A. Henrion Caude[2]
Université J. Fourier Grenoble, Faculté de Médecine, AGIM CNRS/UJF FRE 3405, La Tronche, France[1]
Université Paris Descartes, INSERM U 781, Hôpital Necker-Enfants Malades, Paris, France[2]

1 INTRODUCTION

The present genomes are the result of a long evolution from the start of the life on the earth until the appearance of mammals and human. We will try in this chapter to show that the control of important genetic networks involved both on defense and energy processes of cells in numerous living systems is under the dependence of different regulators, among them microRNAs and circular RNAs. We show that these genetic networks have only a small number of asymptotic dynamical behaviors, called *attractors*. This small number is directly linked with the possibilities of differentiation of the concerned cells and is controlled inside the interaction graphs of the genetic networks, whose nodes are genes and signs of arrows between genes indicate the presence of interactions between these genes, + (resp. −) in the case of activation (resp. inhibition), by the circuits (closed paths between genes) of the strong connected components (*i.e.*, subgraphs containing a path between any couple of their genes), giving the network the possibility to have more than one attractor (and due to positive circuits, it is made of an even number of inhibitions) and the possibility for an attractor to be stable and possibly oscillating (due to negative circuits, made of an odd number of inhibitions); *i.e.*, having a large number of initial configurations of gene states giving birth, after a dynamical evolution, to its asymptotic behavior. Section 2 describes the circuits involved in adaptive and innate immunologic systems, giving a way to calculate the reduction of the attractor numbers due to the presence of intersecting circuits and to the inhibitory regulation by microRNAs often responsible of periodic protein signals (Bandiera et al., 2011, 2013; Bulet et al., 1999; Demongeot and Besson, 1983, 1996; Demongeot et al., 2003, 2009a, 2009b, 2010, 2011, 2012, 2013a, 2013b, 2014a, b; Weil et al., 2004). Section 3

presents the Ferritin control network, regulating the iron metabolism in mammals, section 4 is devoted to the study of the engrailed morphogenetic network, and section 5 concerns the network controlling a disease called biliary atresia.

2 IMMUNETWORKS

2.1 THE IMMUNETWORK RESPONSIBLE OF THE TOLL-LIKE RECEPTOR (TLR) EXPRESSION

The activation of natural killer (NK) cells, involved in innate immune response, is controlled by the ligands of the Toll-like receptors (TLRs) (see Figure 7.1 and Bulet et al., 1999; Elkon et al., 2007; Miyake et al., 2000). The gene *GATA-3* is activating the gene *BCL10,* which is crucial for NFκB activation by T- and B-cell receptors (Zhou et al., 2004), and the protein ICAM1 is a type of intercellular adhesion molecule continuously present in low concentrations in the membranes of leucocytes involved in the blood adaptive immune response. The network controlling TLR and ICAM1 expression contains a couple of circuits, one positive five-circuit tangent to a negative five circuits, giving only one attractor (see Figure 7.1 and Table 7.1, red circle), which corresponds to the activation of the gene TLR.

2.2 THE LINKS WITH THE microRNAs

Most of the genes introduced here have links with microRNAs exerting a negative control on them and then, susceptible to deciding if the unique physiologic attractor will occur, by cancelling their target gene activity. Here are two examples of such microRNAs, negatively regulated by the circular RNA ciRs7 (Hansen et al., 2011, 2013):

1. for the subsequence pUNO-hRP105 of the TLR 2 gene (4937 bp) (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/; Miyake et al., 2000), close to the reference sequence AL (*cf.* Annex A6 and Demongeot, 1978; Demongeot and Besson, 1983, 1996; Demongeot et al., 2003, 2009a, 2009b; Demongeot and Moreira, 2007), the hybridization is made by the microRNA miR 200a:
   5'-CCAUUCAAGAUGA**AUGGU**ACUG-3' AL 14 anti-matches
   5'-UCAUUGUUAUGCUACAGGUAUU-3' ciRs7 14 anti-matches
   3'-UGUAGCAAUGGUCUGUCACAAU-5' hsa miR 200a 12 anti-matches
   5'-UUGUGCUCAUUGAGAUGAAUGG-3' pUNO-hRP105 mRNA starting in position 531
   5'-UACUGCCAUUCAAGAUGAAUGG-3' AL 15 matches
   5'-CUGCCAUUCCUGAAGAAUAGCA-3' ciRs7 17 matches
   5'-AGGGAGCUACAAUUCAAGAUGA-3' ciRs7 17 matches (significance of 2x17 matches: 2.5‰)
2. For the GATA 3 gene, hybridization is made by miR 200c (http://mirdb.org/miRDB/; http://mirnamap.mbc.nctu.edu.tw/):
   5'-GCCAUUCAAGAUGA–A**UGGUA**CU-3' AL 13 anti-matches

5'-ACCAUCAUUAUCCCUAUUUUACA-3' ciRs7 15 anti-matches
3'-AGGUAGUAAUGGGC–CGUCAUAA-5' has miR 200c 15 anti-matches
5'-UCUGCAUUUUUGCAGGAGCAGUA-3' GATA 3 mRNA starting in position 57

The gene expressing TLR contains the AGAUGAAUGG subsequence, belonging both to the D-loop of many tRNAs, to the reference sequence AL (Demongeot, 1978; Demongeot and Besson, 1983, 1996; Demongeot et al., 2003, 2009a, 2009b; Demongeot and Moreira, 2007) and to the circular RNA ciRs7, which signifies its affiliation to an ancestral genome, confirming the old origin of the innate immunologic system (Bulet et al., 1999; Elkon et al., 2007; Miyake et al., 2000) (see Annex A5 for the significance of the matches). In case of parallel updating (with $T=0$), the network controlling the TLR production has 4 (resp. 1) attractor, if miR200c is (resp. not) expressed (see Tables 7.1 and 7.6 left bottom, red circles).

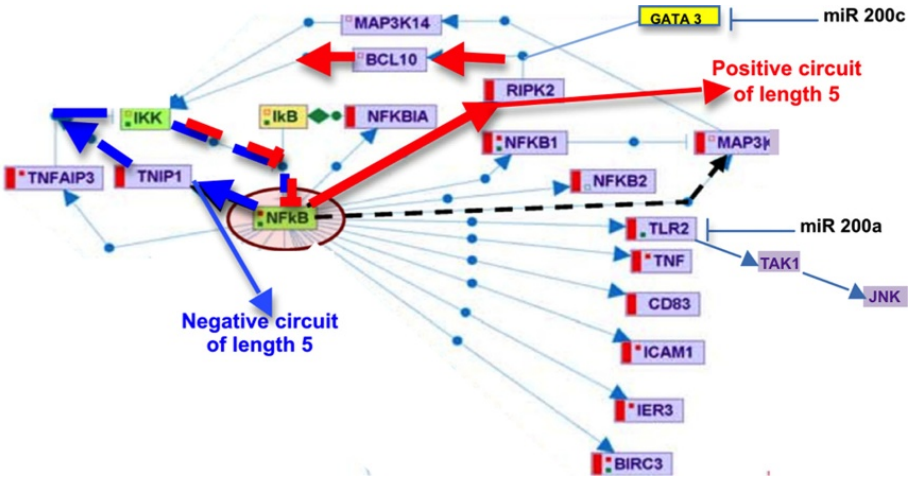


FIGURE 7.1

The network controlling the production of the TLR and the gene ICAM1 (adapted after Elkon et al., 2007).

2.3 THE ADAPTIVE IMMUNETWORKS

The adaptive immunetworks are essentially made of three couples of tangent circuits (Figure 7.2) concerning the key genes GATA 3, transcriptional activator binding to DNA sites with the consensus sequence [AT]GATA[AG]), which controls negatively T cell receptors β (TCRβ), PU.1 (PUrine-rich box-1 gene), controlling negatively the recombination-activating gene (RAG) responsible of the V(D)J rearrangements giving birth to the TCRα receptors, and Zap70 (Zeta-chain-associated protein kinase 70 gene), controlling negatively TCRβ synthesis (Demongeot et al., 2012; Georgescu et al., 2008). These circuits are inserted into a global immunetwork (Figure 7.3), whose

Table 7.1 Total number of attractors in parallel dynamics (with $T=0$), with 2 tangent circuits, left-circuit being negative of length l and right-circuit positive of length r (*cf.* Demongeot et al., 2012, and Annex A6). .

l \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	2	2	3	3	5	5	8	10	15	19	31	41	64
2	1	1	2	3	3	4	5	8	10	14	19	31	41	63
3	1	2	1	3	3	6	5	8	8	15	19	33	41	64
4	1	1	2	1	3	4	5	11	10	14	19	24	41	63
5	1	2	2	3	1	5	5	8	10	26	19	31	41	64
6	1	1	1	3	3	1	5	8	8	14	19	63	41	63
7	1	2	2	3	3	5	1	8	10	15	19	31	41	158
8	1	1	2	1	3	4	5	1	10	14	19	24	41	63
9	1	2	1	3	3	6	5	8	1	15	19	33	41	64

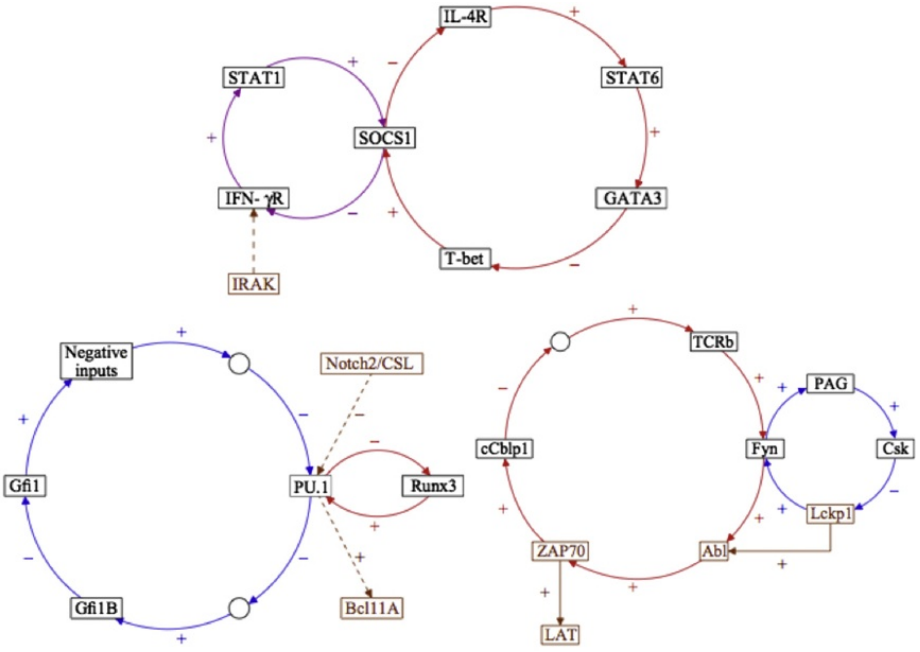


FIGURE 7.2

Left: Negative six-circuit tangent to a negative two-circuit controlling PU.1. Middle: negative three-circuit tangent to a positive five-circuit controlling GATA 3. *Right:* Tangent negative six-circuit and four-circuit controlling ZAP70 (adapted after Demongeot et al., 2012).

attractors are those of the three couples of circuits, the rest of the network being reducible to up- and down-trees connected to the circuits.

In the case of parallel updating (with $T=0$; *cf.* Annex A1), Table 7.7 shows that a negative six-circuit tangent to a negative two-circuit has only one attractor (Table 7.7, blue circle at the bottom right), less than an isolated negative six-circuit

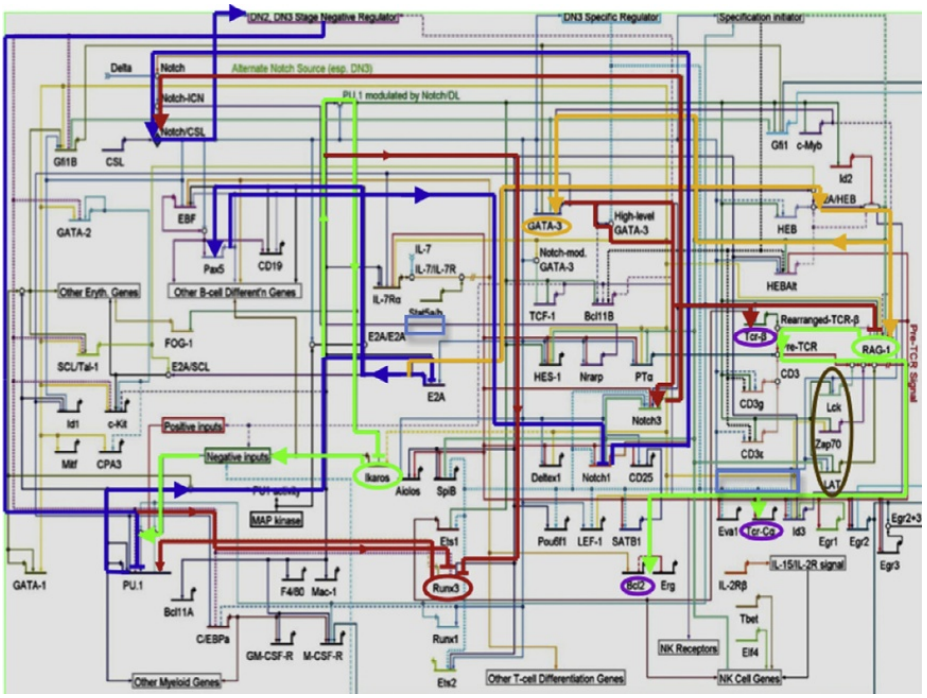


FIGURE 7.3

The global immunetwork controlling genes GATA 3 (orange), RAG (green), and ZAP70 (brown), with a negative circuit of length 6 (dark blue) controlling PU.1, which activates the NK cells through Ets1 (pink), and with circuits from the strong connected component controlling genes TCR-#, TCR-" and Bcl2 (light blue) (adapted after Georgescu et al., 2008). .

that has six attractors (Table 7.2 blue circle), showing that for controlling RAG, Bcl1, and NK cells (*i.e.*, in the case of both adaptive and innate mechanisms), if Notch2/CSL is silent, PU.1 is on and hence can activate both RAG and Bcl1, as well as promote NK cells. On the contrary, if Notch2/CSL inhibits PU.1, the immunologic system is paralyzed. In the same way, we can show that GATA 3 and ZAP70 networks each have three attractors (Table 7.7, orange circles on the right).

3 THE IRON CONTROL NETWORK

The regulatory network controlling iron metabolism contains 10 elements, with one positive circuit of length 6 and one negative of length 4 (Figure 7.4, bottom): the number of attractors is 4 (Table 7.3), following the rules of Table 7.2 (green circle). Depending on the inhibition by miR-485 or miRNA sponge ciRs7 (Hansen et al., 2011, 2013; Hentze et al., 1987; Sangokoya et al., 2013), we get either of two fixed

Table 7.2 Total Number of Attractors of Period p in Parallel Updating for a Unique Isolated Negative Circuit of Size n (After Demongeot et al., 2012). .

p \ n	1	2	3	4	5	6	7	8
2	1	-	1	-	1	-	1	-
4	-	1	-	-	-	1	-	-
6	-	-	1	-	-	-	-	-
8	-	-	-	2	-	-	-	-
10	-	-	-	-	3	-	-	-
12	-	-	-	-	-	5	-	-
14	-	-	-	-	-	-	9	-
16	-	-	-	-	-	-	-	16
Total	1	1	2	2	4	6	10	16

configurations or one limit-cycle of configurations (the second having a negligible attraction basin size, equal to 4% of the possible initial conditions). Same attractors are observed for the first eight nodes, when miR-485 or anticiRs7 are expressed. The presence of an ancestral subsequence in ciRs7 sequences (*cf.* section 2.2 and Figure 7.4, top) is in favor of the seniority of the iron control system:

5'-AUGGGGCAACAUAUUGUAUGAA-3' FPN1a 14 anti-matches
3'-UCUCUCCUCUCGGCACAUACUG-5' miR-485 15 anti-matches
5'-UCUUUAUGUCCUCUACUGGCAGAGAGGAUGGGGGAGU
UGUGUAUUCUUCCAGGUUC-3' ciRs7
5'-UCAAGAUGAAUGGUACUGCCAU-3' AL 12 matches
14 matches 5'-CCUGUUGGUCUCUUCCAGGUAC-3' IRP
10/17 anti-matches 3'-CUGGAUCAGUGGAUCUA-5' IRE-FPN1a

We can calculate a robustness parameter for the iron control network based on evolutionary entropy, defined by

$$E = \log 2^{10} - \mathrm{E}_{\mathrm{attractor}} = 10\log 2 - \Sigma_{k=1,m}\,\mu(C_k)\log\mu(C_k),$$

where m is the attractor number and $C_k = B(A_k) \cup A_k$ is the union of the attractor A_k and of its attraction basin $B(A_k)$ (*cf.* Annex A2). Hence, $E_{attractor} = -\Sigma_{k=1,4}\mu(C_k)\log\mu(C_k) = 1/2\log2 + 1/4\log4 + 0.21\log(0.21) + 0.039\log(0.039)$. When $E_{attractor}$ decreases (*e.g.*, if c-Myc is cancelled, provoking the disappearance of one attractor), then the robustness of the network increases. In the stochastic parallel updating case with $T > 0$, we can calculate the derivative of E with respect to the randomness parameter T, the interaction weights being supposedly the same for each interaction (Demongeot et al., 2013a). This derivative gives an indication about the sensitivity to noise of the network.

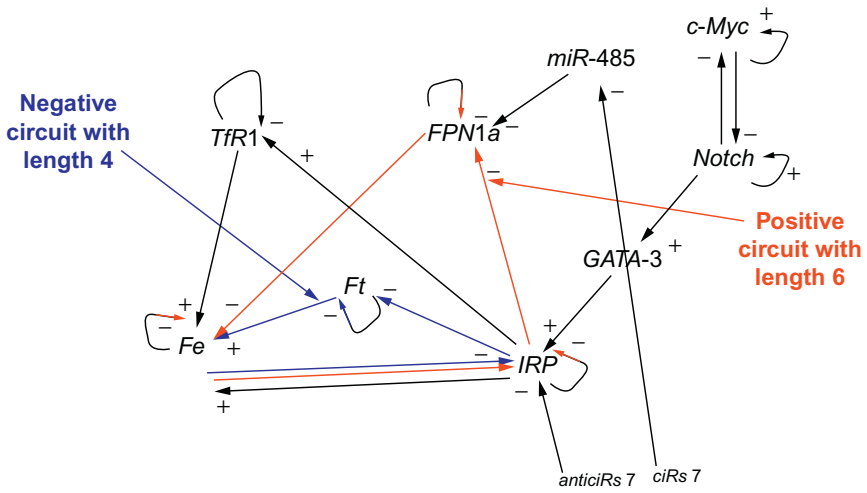


FIGURE 7.4

Top: matches between miR-485-3p and its target FPN1a (FerroPortiNe), and between the microRNAs sponge ciRs7 and its target miR-485-3p in the iron control network. Bottom: Iron control genetic network with coexistence of numerous circuits between FPN1a, Ft (Ferritin), iron regulatory protein (IRP) and transferrin receptor (TfR1), with a positive circuit of length 6 and a negative one of length 4.

Table 7.3 Recapitulation of the four attractors of the iron metabolic system, with miR-485 and anticiRs7 not expressed (state 0) and other genes expressed (state 1) or not (state 0) and (bottom) the attraction basin sizes for parallel updating (with $T=0$).

Position	Gene	Fixed Point 1	Fixed Point 2	Limit Cycle 1			Limit Cycle 2	
1	TfR1	0	0	0	0	1	1	0
2	FPN1a	0	0	0	0	0	0	0
3	c-Myc	0	1	0	0	0	0	0
4	Notch	0	0	1	1	1	1	1
5	GATA-3	0	0	1	1	1	1	1
6	IRP	0	0	0	1	0	0	1
7	Ft	0	0	0	0	0	0	0
8	Fe	0	0	0	0	1	0	1
9	miR-485	0	0	0	0	0	0	0
10	anticiRs7	0	0	0	0	0	0	0
	Attraction basin size	512	256	216			40	

4 MORPHOGENETIC NETWORKS

The morphogenetic network centered on the engrailed gene (*En* on Figure 7.5) controls in vertebrates the segmentation phase, as well as the morphogenesis of feathers and hairs in birds and mammals (Demongeot and Waku, 2012; Michon et al., 2008).

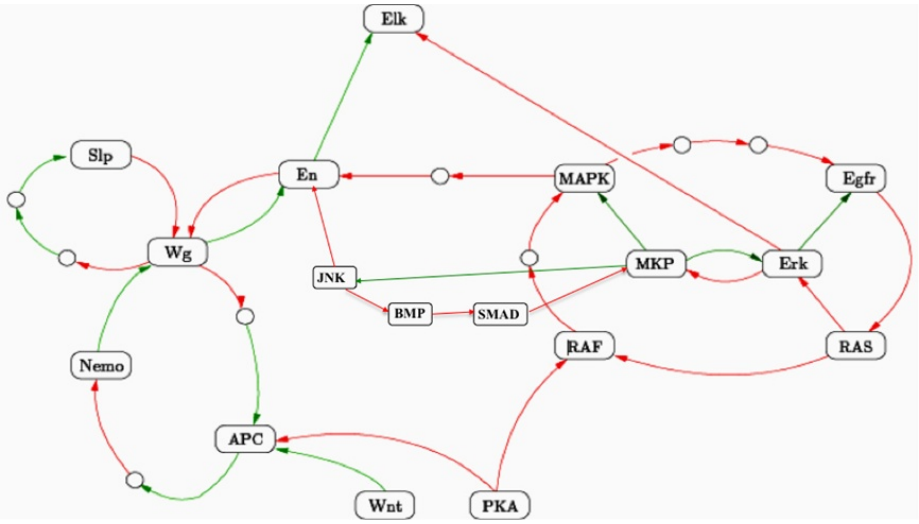


FIGURE 7.5

Engrailed centered network in *Drosophila*, with positive arcs in red and negative arcs in green.

.

Engrailed is not only required for the segmentation phase in vertebrates, but it efficiently represses the activity of numerous transcriptional factors like Elk (Saenz-Robles et al., 1995; Vickers and Sharrocks, 2002). Note that the engrailed centered network shown in Figure 7.5 was built by bringing together information about repressions and inductions from various sources:

- Inhibition of Engrailed by Wg and its activation by JNK is given in Gettings et al. (2010).
- Action of MPK through JNK inhibition is studied by Mantrova and Hsu (1998).
- Egrf positive feedback circuit is explained by McEwen and Peifer (2005).
- The whole Wnt pathway can be found in http://www.kegg.jp/kegg/pathway/dme/dme04310.
- Inhibition of ERK by MKP is described in Weil et al. (2004).
- Activation of Nemo by ARM, inhibited by APC, inhibited itself by Wg through Dsh (http://dev.biologists.org/content/131/12/2911/ F8.large.jpg).
- Inhibition of Wg by Nemo is described by Zeng and Verheyen (2004).

The genetic regulatory network of Figure 7.4 possesses six negative circuits of respective sizes 7, 5, 4, 3, and 2, and two positive circuits of respective sizes 7 and 4. Following Demongeot et al. (2012), we know that each of these circuits bring the following total number of attractors when isolated with parallel deterministic updating ($T = 0$):

- Negative circuit of size 7: 10 attractors (*cf.* Mathematical Annex Table 7.7, left bottom blue circle)
- Negative circuit of size 5: 4 attractors (*cf.* Mathematical Annex Table 7.7, left bottom red circle)
- Negative circuit of size 4: 2 attractors (*cf.* Mathematical Annex Table 7.7, left bottom orange circle)
- Negative circuit of size 3: 2 attractors (*cf.* Mathematical Annex Table 7.7, left bottom orange circle)
- Negative circuit of size 2: 2 attractors (*cf.* Mathematical Annex Table 7.7, left bottom orange circle)
- Positive circuit of size 7: 20 attractors (*cf.* Mathematical Annex Table 7.7, left top blue circle)
- Positive circuit of size 4: 6 attractors (*cf.* Mathematical Annex Table 7.7, left top voilet circle).

Table 7.4 Attractors in Parallel Mode (with $T=0$) for the MPK/ERK Centered Subnetwork, with Negative 7- and 3-Circuit Tangent

Gene	PF	Cycle Limit 1				Cycle Limit 2			
En	0	1	0	0	0	0	1	0	1
Elk	0	0	0	0	1	0	0	0	0
MAPK	0	0	0	0	1	1	0	1	0
Egfr	0	1	0	0	0	0	1	0	1
RAS	0	0	1	0	0	1	0	1	0
Erk	0	0	0	1	0	0	1	0	0
MKP	0	0	0	0	1	0	0	1	0
RAF	0	0	0	1	0	0	1	0	1
SMAD	0	0	0	0	0	0	0	0	0
BMP	0	0	0	0	0	0	0	0	0
JNK	0	0	0	0	0	0	0	0	0
TBA=2048	1320				648				80
ABRS = 1	0,6445				0,31640625				0,0390625

The results obtained recently on tangent and intersecting circuits [*cf.* Melliti et al. (2014); Richard, 2011) and Table 7.7 in Annex A3) show a drastic reduction of the attractor number to 3 or 1 (depending on source nodes PKA and Wnt), because the engrailed centered network has the following:

- One positive circuit of size 7 intersecting a negative circuit of size 7: 1 attractor, instead of 200 if the circuits are disjoint (*cf.* Mathematical Annex Table 7.7, right top, red circle)
- One positive circuit of size 7 intersecting a negative circuit of size 3: 5 attractors, instead of 20 if the circuits are disjoint (*cf.* Mathematical Annex Table 7.7, right top, blue circle)

- One negative circuit of size 7 tangent to a negative circuit of size 4 : 2 attractors (*cf.* Mathematical Annex Table 7.7, right bottom red circle)
- One negative circuit of size 7 intersecting a negative circuit of size 3: 3 attractors, instead of 20 if the circuits are disjoint (*cf.* Mathematical Annex Table 7.7, right bottom, green circle)
- One negative circuit of size 4 tangent to a negative circuit of size 2: 2 attractors (*cf.* Mathematical Annex Table 7.7, right bottom red circle)
- One negative circuit of size 3 tangent to a negative circuit of size 2: 2 attractors (*cf.* Mathematical Annex Table 7.7, right bottom brown circle)
- Three co-tangent circuits, one positive circuit of size 4, one negative circuit of size 5, and one negative circuit of size 2: 1 attractor, instead of 24 if the circuits are disjoint (*cf.* Mathematical Annex Table 7.7, right voilet circles)

The corresponding attractors are given in Table 7.4, depending on the expression of the gene sources of the up-trees controlling the circuits; *i.e.*, *Wnt* and *PKA*. If both genes *Wnt* and *PKA* are expressed, there are three attractors with only one limit-cycle of period 6 as asymptotic dynamical behavior of the engrailed centered network, for which the gene *En* is not expressed and *Elk* is expressed one-third of the time, inhibiting both the dorsal closure and allowing the excitable cells differentiation. If *PKA* is silent and *Wnt* is either expressed or silent, there are also three attractors: two fixed points, where neither *En* nor *Elk* are expressed, and a limit cycle of period 4, where *En* is expressed half the time and *Elk* is expressed a quarter of the time. If the gene *Wnt* is silent and *PKA* expressed, there exists only a limit cycle of period 6, where *En* is expressed half the time and *Elk* is expressed 1/6 of the time. Other examples of genetic regulatory networks involving *Wnt* are given in Michon et al., (2008). All these examples show that more generally, the architecture of a genetic regulatory network consists of the strong connected components of its interaction graph, to which are attached three kinds of substructures:

- a set of up-trees, issued from the sources of the interaction graph of the network, made either of small RNAs (like microRNAs, translational inhibitors), or of genes repressors or inductors, self-expressed without any other genes controlling them, like the genes *Wnt* and *PKA*
- a set of circuits in the core (in a graphical sense) of the strong connected components of the interaction graph. These circuits are unique or multiple, reduced to one gene (if there is an auto-control loop) or made of several ones, negative or positive, and disjoint or not, like the circuits involving *Erk*
- a set of down-trees going to the sinks of the interaction graph; *i.e.*, to genes controlled by but not controlling any other genes, like the gene *Elk*

In *Drosophila* embryo, using the interaction graph of the engrailed centered network, a simple model based on the knowledge about the asymptotic dynamics of the network (*i.e.*, its attractors) shows that *Wg* is expressed and inhibits *Dsh* during the Mixer cell formation at the para-segment boundaries, during the polarization of epidermal cells during dorsal closure in *Drosophila,* where *Wnt* and *PKA* are expressed (Gettings et al., 2010).

5 BILIARY ATRESIA CONTROL NETWORK

The genetic network controlling the morphogenesis of the biliary canal can be summarized as in Figures 7.6 and 7.7 (Bessho et al., 2013; Choe et al., 2003; Girard et al., 2011, 2012; Kohsaka et al., 2002; Luedde et al., 2008; Matte et al., 2010; Nouws and Shadel, 2014; Ranganathan et al., 2011; Xiao et al., 2014). By using the gene expression data comparing normal individuals and patients suffering from biliary atresia (*cf.* Figure 7.6, left, and Melliti et al., 2014; Meyer and Nelson, 2011), we can locate and study three key genes inside or around the network, *Bcl-w, TGF-ß, and elf-2α* kinase (Figure 7.7 and Choe et al., 2003), and study the attractors of the network (Table 7.5).

Transcript identity	Intensity		Ratio (BA/ Normal)
	Normal	BA	
Bcl-w	16.7	28.8	1.7
Laminin BP (binding protein)	0.7	7.7	11
HRS (HGF-regulated tyrosine kinase substrate)	0.5	3.5	6.9
Thymosin β-4	0.6	4.1	6.5
Thymosin β-10	0.5	2.5	5.4
TGF-ß	0.6	2.1	3.3
TIMP (tissue inhibitor of metalloproteinase)-1	1.8	3.9	2.2
SRP (signal recognition particle) 4	2	7.1	3.6
SRP9	1.7	5.6	3.3
SNAP (soluble NSF attachment proteins) 45	1.2	5.6	4.5
Alu RNA BP	0.8	6.9	8.5
suptSh (human homologue of the yeast transcription factor, SPT5)	1.9	4.5	2.3
eIF-2α kinase	4.5	12.1	2.7

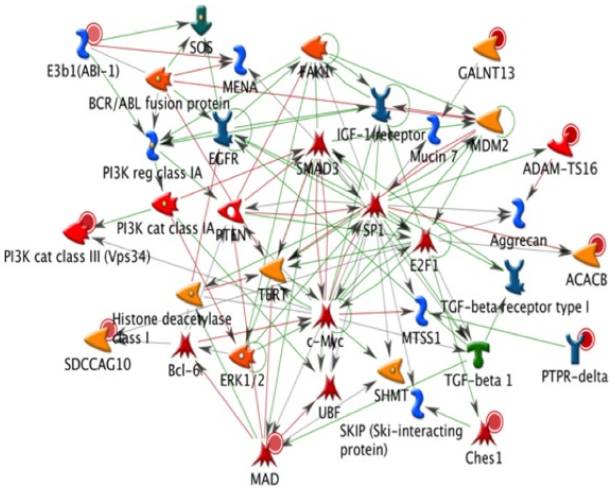


FIGURE 7.6

Left: biliary atresia gene expression (Choe et al., 2003), with three markers: *Bcl-w, TGF-ß,* and *elf-2α* kinase, whose level of expression is measured in a control group (Normal) and in patients suffering from biliary atresia (noted here as BA). *Right:* Genetic network controlling biliary atresia.

We can add to the network shown in Figure 7.6, provided initially by MetaCore™ and checked after in the literature. The following information has been added in Figure 7.7 to the network in Figure 7.6:

- The gene *JAG1* (mutated in case of biliary atresia (Kohsaka et al., 2002; Matte et al., 2010) activates Notch, necessary for the activation of EGFR (Franco et al., 2006; Ranganathan et al., 2011).
- Human microRNAs *miR 29* and *miR 39b* inhibits *IGF 1* and *PI3K*, respectively (Bessho et al., 2013; Gottwein et al., 2011; Hand et al., 2012).
- The gene *Bcl-w*/*Bcl-2* is inhibited by *SMAD 3,* and activated by *ERK ½* and *MAPK* (Kang and Pervaiz, 2013). Its protein is phosphorylated (hence inhibited) by *JNK* (Singh et al., 2009).
- The gene eIF-2α kinase phosphorylates (hence, inhibits) eIF-2α (Gurzov and Eizirik, 2011; Lee et al., 2000) and activates IRP and MAPK.

The strong connected component of the interaction graph of the biliary atresia network (Figure 7.7) shows the existence of a negative circuit of length 11 (by counting the auto-loops) tangent on the gene *FAK1* to a negative circuit of length 5 (Figure 7.8). The theoretical results of Demongeot et al. (2012) shows that we can expect seven attractors, which can be simulated (Table 7.5), whose only two limit-cycles (their attraction basins representing about half of all possible expression patterns) show the presence of the same gene expression than that detected in patients suffering from biliary atresia; *i.e.*, an increased expression of the three genes *Bcl-w, TGF-ß,* and *elf-2α* kinase (Choe et al., 2003), an overexpression of IGF1 and a down-expression of JAG1, mutated in biliary atresia (Kohsaka et al., 2002), provoking a weak expression of EGFR (Matte et al., 2010). That partially validates the control network proposed for biliary atresia syndrome.

We can break the negative circuit of length 5 by using the microRNA *hsa miR 39b,* where one of the targets is the gene *PI3K* (Xiao et al., 2014; Kohsaka et al., 2002):

- 3'-uguaaacauccuacacucagcu-5' *hsa miR 30b*
- 5'-AUGGUACUGCCAUUCAAGAUGA-3' AL 13 anti-matches

The main result of the disappearance of the negative circuit of length 5 is an increase of the attractor number, passing from 7 to 10 (*cf.* Table 7.2, red circle). Two other negative circuits do not increase the attractor number:

- The negative circuit of length 3 *FAK1-MDM2-IGF-1* is tangent at *FAK1* to the two previous circuits of lengths 7 and 4, but this coupling brings only one attractor (Figure 7.9, blue circle). It can be broken by the microRNA *hsa miR 29* (Matte et al., 2010) and can only change the configuration of expressed genes; for example, in the presence of *miR 29,* the *IGF-1* state becomes 0 in the attractor states of Table 7.5:
  - 3'-acugauuucuuuugguguucag-5' *hsa miR 29*
  - 5'-CCAUUCAAGAUGAAUGGUACUG-3' AL 13 anti-matches
- The negative circuit of length 3 FAK1-PI3K-PTEN also brings only one attractor and its break by miR 39b (Figure 7.7) changes only the attractor states by fixing at 0 the state of PI3K (*cf.* Table 7.5).

From the mathematical analysis of the attractors of the biliary atresia network, we can infer the existence of seven possible stationary behaviors; among them, only four have an attraction basin that is sufficiently stable (*i.e.*, containing a sufficient number of initial conditions to resist to large perturbations in the state space of the expression configurations). These four attractors represent the final evolution of 96% of the possible initial configurations; two of them being fixed configurations (called *fixed points* and denoted as PF in Table 7.5), and the other two being periodic configurations (called *limit cycles* and denoted as CL in Table 7.5). Among these attractors, notice that only one corresponds to all the characteristics observed in the pathologic case of the biliary atresia: *E2F1, TGF-ß,* and *Bcl-w* are always expressed, *EGFR* is weakly expressed, as is *ERK ½,* due to its absence of inhibition of *Bcl-w*. The unique attractor satisfying these constraints is the fixed configuration PF 4 on Table 7.5. It has a small attraction basin (ABRS = 2%), but it can represent a nonphysiologic

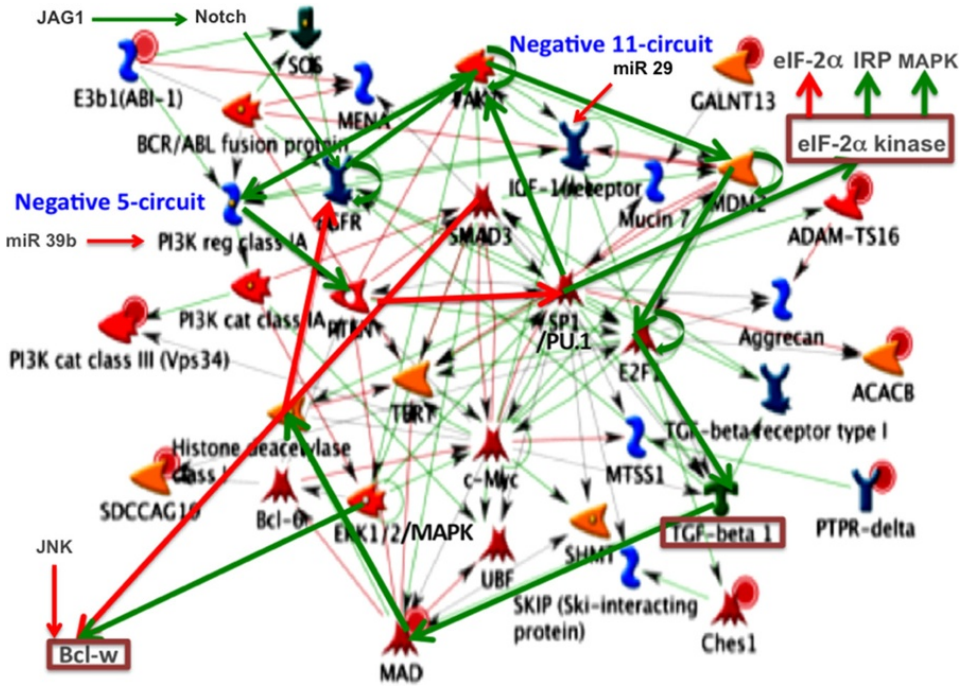


FIGURE 7.7

Biliary atresia control network, with three important genes: two on its frontier (up-tree), *Bcl-w* and *elf-2α* kinase; and one inside the strong connected component, *TGF-ß 1.* The strong connected component (scc) contains two tangent (in FAK1) negative circuits: one of length 11 passing through *TGF-ß 1,* and the other of length 5, passing through *PI3K.* Both can be broken by the microRNA *miR 39b.* The scc is on the control of *JAG1* by Notch, activates (through SP1/ PU.1 and *eIF-2α* kinase) the genes *IRP* and *MAPK,* and inhibits the gene *eIF-2α.*

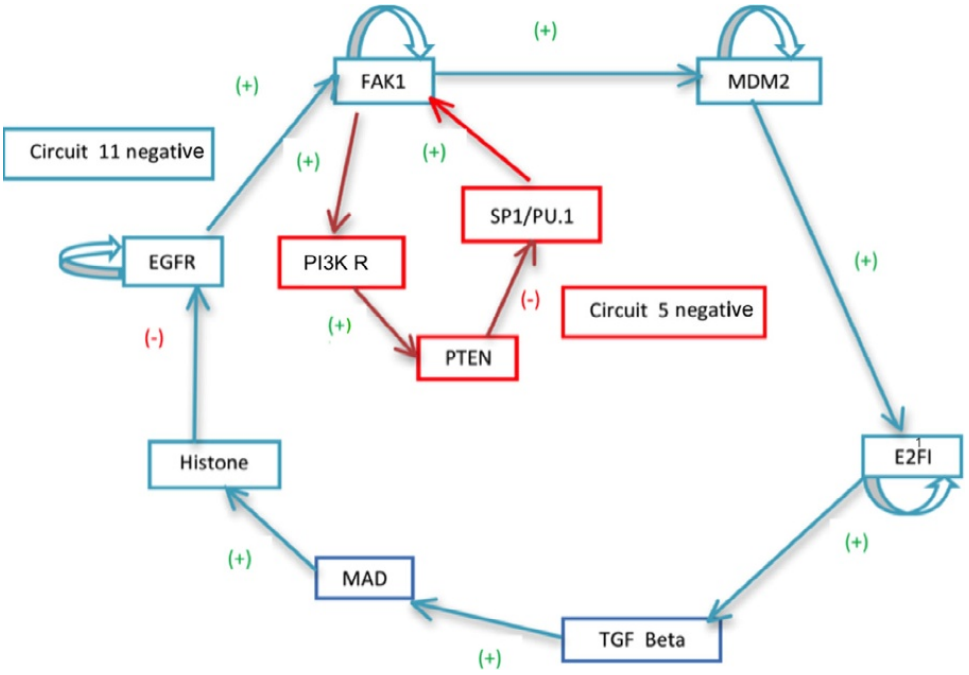


FIGURE 7.8

The two main negative circuits of the biliary atresia control network: one of length 11 and the other of length 5 (taking into account the auto-catalytic loops).

Table 7.5 Description of the Seven Attractors of the Biliary Atresia Control Network (with Parallel Updating and $T=0$)

	PF1	PF2	PF3	PF4	PF5	CL 1		CL 2	
FAK1	0	1	1	1	1	1	1	1	1
IGF-1	0	1	1	1	1	1	1	1	1
PI3K REG	0	1	1	1	1	1	1	1	1
EGFR	0	1	1	0	0	1	1	0	0
SMAD3	0	0	0	0	0	1	0	1	0
MDM2	0	1	1	1	1	1	1	1	1
SP1	0	0	0	0	0	0	0	0	0
PTEN	0	0	0	0	0	0	0	0	0
PI3K CAT CLASS	0	1	1	1	1	1	1	1	1
Histone	0	1	1	1	1	1	1	1	1
TERT	0	0	0	0	0	0	0	0	0
E2F1	0	1	1	1	1	**1**	**1**	**1**	**1**
TGF-ß	0	1	1	1	1	**1**	**1**	**1**	**1**
MAD	0	1	1	1	1	1	1	1	1
C_myc	0	1	0	0	1	1	0	1	0
ERK 1/2	0	1	0	0	1	1	1	1	1
Bcl_6	0	0	0	0	0	0	0	0	0
TGF RECEPTOR	0	1	1	1	1	1	1	1	1
Bcl-w	0	0	0	0	0	**1**	**0**	**1**	**0**
PI3K CLASS III	0	1	1	1	1	1	0	1	0
E3B1	0	0	0	0	0	0	0	0	0
UBF	0	0	0	0	0	0	0	0	0
TBA	336	949200	1020024	82616	37600	1997032		107496	
ABRS	8,01086E-05	0,22630692	0,24319267	0,01969719	0,00896454	0,476129532		0,025629044	
TTBA	4194304								

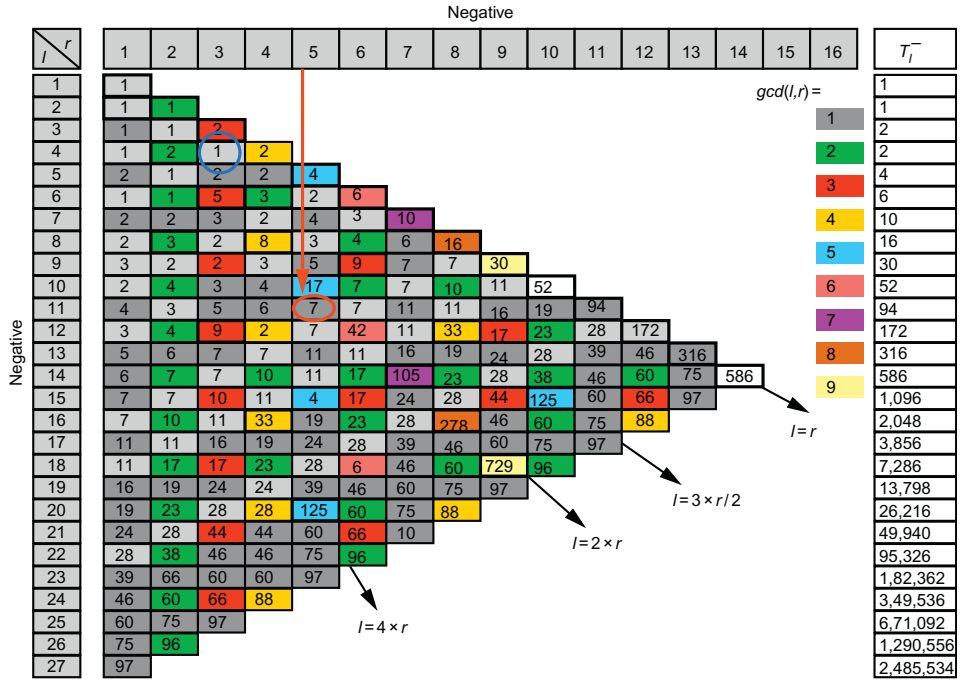


FIGURE 7.9

Calculation of the number of attractors of the biliary atresia control network (Demongeot et al., 2012). At the intersection of the column 5 and the line 11, one can find the attractor number, 7, due to a couple of tangent negative circuits, one of length 5 and the other of length 11..

possibility of differentiation in a few cases depending on initial conditions fixed by the genomic expression It could be interesting to observe in patients the expression activity of the gene *c-Myc,* which is absent in PF4 on Table 7.5, in order to confirm it as a candidate for explaining the pathologic behavior, due to the role of *EGFR* and *ERK ½* in the morphogenesis (see the network described in section 4, earlier in this chapter) that is necessary to activate Nemo.

6 CONCLUSION AND PERSPECTIVES

The genetic networks regulated by small RNAs fitting with ancestral sequences are showing interesting properties with a small number of attractors, allowing to control less than four main attractors in general, some inhibiting the function (the brakes), and the others activating it (the accelerators). The role of the microRNAs is to provide an unspecific inhibitory noise leaving active in the dynamics only the circuits with sufficiently strong interactions to be able to express these attractors. The circular RNAs are inhibiting in an unspecific way the microRNAs in order to have, as in neural networks, the possibility of obtaining a double reciprocal influence (inhibitory

and anti-inhibitory) on mRNAs (*i.e.*, on gene expression). Because the sequences of the small regulatory RNAs offer in general a good fit with ancestral sequences, we can infer that the control of important domains involved in key metabolism or defense processes have been fixed early in the evolution to optimize systems like the immunologic, energetic and morphogenetic ones, which are crucial for survival. This resemblance with ancient genomes can come from two different mechanisms: (i) there are relics, having not mutated, that are still present in the genomes today (*e.g.*, in the most conserved and universal parts of RNA molecules like the tRNA loops); and (ii) there exists still a mimicking of the start of life, such as circular RNA.

In order to survive, it has to solve the following variational problem: (i) to be sufficiently small to remain not denatured in the cell cytoplasm and (ii) hybridize a sufficient number of microRNAs in order to serve as a brake to their inhibitory activity. Because these microRNAs have to be sufficiently nonspecific to target a great number of messenger RNAs, they have to contain in their 22 bases the maximum of codons from different classes of synonymy of the amino-acid. This variational problem is very close to that of the beginning of life, in which primitive RNAs [like the AL sequence (Demongeot, 1978; Demongeot and Besson, 1983, 1996; Demongeot et al., 2003, 2009a, 2009b; Demongeot and Moreira, 2007)] could have occurred and persisted due to the selective advantage to fix a great number of amino acids realizing a protein proto-membrane, an ancestor of the plasmic membrane that exists today. That could suffice to explain the similarity observed between the primitive and present circular RNA. Future studies could perform an exhaustive examination of small RNAs in order to reinforce the hypotheses of both (i) the existence of RNA relics still present in the evolved genomes and (ii) the biosynthesis of RNA molecules similar to those existing at the beginning of life, because the same variational problem is being solved. Further studies on the genetic regulatory networks could show that the role of all their inhibitors is crucial for getting a very limited number of attractors focusing only on the cell functions necessary to survive, eliminating all the nonfunctional attractors.

MATHEMATICAL ANNEX

A1 DEFINITIONS

The mathematical object modeling a real genetic regulatory network is called a *genetic threshold Boolean regulatory network* (denoted in the following as *getBren*). A getBren N can be considered as a set of random automata, defined by the following criteria (Robert, 1980):

1. Any random automaton i of the getBren N owns at time t a state $x_i(t)$ valued in $\{0,1\}$, 0 (resp. 1) meaning that gene i is inactivated or in silence (Respectively activated or in expression). The global state of the getBren at time t, called *configuration* in the sequel, is then defined by $x(t) = (x_i(t))_{i \in \{1,n\}} \in \Omega = \{0,1\}^n$

2. A getBren N of size n is a triplet (W, Θ, P), where
- W is a matrix of order n, where the coefficient $w_{ij} \in$ IR represents the interaction weight gene j has on gene i. Sign$(W) = (\alpha_{ij} = \text{sign}(w_{ij}))$ is the adjacency (or incidence) matrix of a graph G, called the *interaction graph*.
- Θ is a threshold of dimension n, its component θ_i being the activation threshold attributed to automaton i.
- M: $\mathbf{P}(\Omega) \rightarrow [0,1]^{m \times m}$, where $\mathbf{P}(\Omega)$ is the set of all subsets of Ω and $m = 2^n$ is a Markov transition matrix, built from local probability transitions P_i giving the new state of the gene i at time $t+1$ according to W, Θ, and configuration $x(t)$ of N at time t such that

$$\forall g \in \{0,1\}, \beta \in \Omega, P_{i,g}{}^{\beta}(\{x_i(t+1) = g \,|x(t) = \beta\}) = \exp\left[g\left(\Sigma_{j \in Ni} w_{ij}\beta_j - \theta_i\right)/T\right]/Z_i,$$

where $Z_i = [1 + \exp[(\Sigma_{j \in Ni}\ w_{ij}\beta_j$ - $\theta_i)/T]$, N_i is the neighborhood of the gene i in the getBren N; *i.e.*, the set of genes j (including possibly i) such that $w_{ij} \neq 0$, and $P_{i,g}{}^{\beta}$ is the probability for the gene i of passing to the state g at time $t+1$, from the configuration β at time t on N_i. M denotes the transition matrix built from the $P_{i,g}{}^{\beta}$'s. M depends on the update mode chosen for changing the states of the getBren automata. In this chapter, we use the parallel or synchronous mode of updating.

For the extreme values of the randomness parameter T, we have the following:

1. If $T = 0$, getBren becomes a deterministic threshold automata network and the transition can be written as

$$x_i(t+1) = h\left(\Sigma_{j \in Ni} w_{ij} x_j(t) - \theta_i\right),$$

where h is the Heaviside function: $h(y) = 1$, if $y > 0$;

$$\text{h(y)} = 0, \text{ if y} < 0,$$

except for the case $\Sigma_{j \in Ni}\ w_{ij}x_j(t) - \theta_I = 0$, for which, if necessary, 1 and 0 are both chosen with probability ½. In this chapter, we chose $T = 0$.

2. When T tends to infinity, then $P_{i,g}{}^{\beta} = ½$ and each line M_i of M becomes the uniform distribution on the basin of the final class of the Markov matrix M to which i belongs (corresponding to an attraction basin when $T = 0$).

We define (Demongeot et al., 2003) the energy U and frustration F of a getBren N by

$$\forall x \in \Omega, U(x) = \Sigma_{i,\, j \in \{1,\, n\}} \alpha_{ij} x_i x_j = Q_+(N) - F(x),$$

where $Q_+(N)$ is the number of positive edges in the interaction graph G of the network N and $F(x)$ the global frustration of x; *i.e.*, the number of pairs (i,j) where the values of x_i and x_j are contradictory with the sign α_{ij} of the interaction between genes i and j: $F(x) = \Sigma_{i,j \in \{1,n\}} F_{ij}(x)$, where F_{ij} is the local frustration of the pair (i,j) defined by

$F_{ij}(x) = 1$, if $\alpha_{ij} = 1$, $x_j = 1$ and $x_i = 0$, or $x_j = 0$ and $x_i = 1$, and if $\alpha_{ij} = -1$, $x_j = 1$ and $x_i = 1$, or $x_j = 0$ and $x_i = 0$,
$F_{ij}(x) = 0$, elsewhere.

Eventually, we define the random global dynamic frustration D by

$$D(x(t)) = \Sigma_{i,\, j\in\{1,\, n\}} D_{ij}(x(t)),$$

where D_{ij} is the local dynamic frustration of the pair (i,j) defined by

$D_{ij}(x(t)) = 1$, if $\alpha_{ij} = 1, x_i(t) \neq h\left(\Sigma_{j\in Ni} w_{ij} x_j(t) - \theta_i\right)$ or $\alpha_{ij} = -1, x_i(t) = h\left(\Sigma_{j\in Ni} w_{ij} x_j(t) - \theta_i\right)$,
$D_{ij}(x(t)) = 0$, elsewhere.

A2 FIRST PROPOSITIONS

Based on these definitions, we can prove the following propositions [*cf.* Demongeot and Waku (2012) and Demongeot et al. (2013b) for complete results]:

Proposition 1 Let us consider the random energy U and the random frustration F of getBren N having a constant absolute value w for its interaction weights, null threshold Θ, temperature T equal to 1, and being sequentially updated. Then:

1. $U(\mathrm{x}) = \Sigma_{\mathrm{i,j}\in\{1,\mathrm{n}\}} \alpha_{\mathrm{ij}} \mathrm{x_i x_j} = \mathrm{Q_+(N)}$ - F(x), where $\mathrm{Q_+(N)}$ is the number of positive edges in the interaction graph G of the network.
2. $\mathrm{E_\mu}(U) = \partial \log \mathrm{Z}/\partial \mathrm{w}$, where the free energy logZ is equal to the quantity log $(\Sigma_{\mathrm{y}\in\Omega} \exp(\Sigma_{\mathrm{j}\in\mathrm{y},\mathrm{k}\in\mathrm{y}} \mathrm{w_{ij} y_j y_k}))$, and μ is the invariant Gibbs measure defined by $\forall\ \mathrm{x}\in\Omega$, $\mu(\{\mathrm{x}\}) = \exp(\Sigma_{\mathrm{i}\in\mathrm{x},\mathrm{j}\in\mathrm{x}} \mathrm{w_{ij} x_i x_j})/\mathrm{Z}$.
3. $\mathrm{Var_\mu U} = \mathrm{Var_\mu F} = \text{-}\partial \mathrm{E_\mu}/\partial \log \mathrm{w}$, where $\mathrm{E_\mu} = \text{-}\Sigma_{\mathrm{x}\in\Omega}\, \mu(\{\mathrm{x}\}) \log(\mu(\{\mathrm{x}\})) = \log \mathrm{Z} - \mathrm{w E_\mu}(U)$ is the entropy of μ, maximal among entropies corresponding to all probability distributions ν for U having the same given expectation $\mathrm{E_\nu}(U) = \mathrm{E_\mu}(U)$.

Proof: (1) It is easy to check that $U(x) = Q_+(N) - F(x)$ and
(2) the expectation of U, denoted $\mathrm{E}_\mu(U)$, is given by

$$\mathrm{E_\mu(U)} = \Sigma_{\mathrm{x}\in\Omega} \Sigma_{\mathrm{i}\in\mathrm{x},\, \mathrm{j}\in\mathrm{x}} \alpha_{\mathrm{ij}} \mathrm{x_i x_j} \exp\left(\Sigma_{\mathrm{i}\in\mathrm{x},\, \mathrm{j}\in\mathrm{x}} \mathrm{w x_i x_j}\right)/Z = \partial \log Z/\partial w$$

(3) Following Demongeot and Waku (2012), we have $\mathrm{Var}_\mu U = \mathrm{Var}_\mu F = \text{-}\partial E_\mu/\partial \log w$, and E_μ is maximal among the proposed entropies

Proposition 2 Let us consider getBren N with $\mathrm{T} = 0$, sequentially or synchronously updated, defined from a potential P defined by

$$\forall\, \mathrm{x} \in \Omega,\ \mathrm{P(x)} = \Sigma_{\mathrm{k}} ({}^{\mathrm{t}}\mathrm{x A_k x}) \mathrm{x_k} + {}^{\mathrm{t}}\mathrm{x W x} + \Theta \mathrm{x},$$

where A, W, and Θ are integer tensor, matrix, and line vector, respectively. Also suppose that

$$\forall \mathrm{i} = 1, \ldots, \mathrm{n},\ \Delta \mathrm{x_i} \in \{-1, 0, 1\}.$$

If h denotes the Heaviside function, consider now the potential automaton i defined by

$$\mathrm{x_i(t+1)} = \mathrm{h}(-\Delta \mathrm{P}/\Delta \mathrm{x_i} + \mathrm{x_i(t)}),$$

and by the condition $x_i(t+1) \geq 0$, if $x_i(t) = 0$, such that the flow remains in Ω. Then, if the tensor A is symmetrical with vanishing diagonal (*i.e.,* if we have the equalities: ∀ i, j, k = 1,…,n, $a_{ijk} = a_{ikj} = a_{kij} = a_{jki} = a_{jik} = a_{kji}$ and $a_{iik} = 0$), and if each submatrix (on any subset J of indices in {1,…,n}) of A_k and W are nonpositive with vanishing diagonal, P decreases on the trajectories of the potential automaton, for any mode of implementation of the dynamics (sequential, block sequential, and parallel). Hence, the stable fixed configurations of the automaton correspond to the minima of its potential P.

Proof: We have, for a discrete function *P* on *Ω*:

$$\Delta P(x)/\Delta x_i = [P(x_1,\ldots,x_i+\Delta x_i,\ldots,x_n) - P(x_1,\ldots,x_i,\ldots,x_n)]/\Delta x_i$$

and the proof is based on the existence of a Lyapunov function proved in Demongeot et al. (2014).

Proposition 3 Let us consider the Hamiltonian getBren, which is a circuit with constant absolute value w for its interaction weights, null threshold Θ , and temperature T equal to 0, sequentially or synchronously updated, whose Hamiltonian H is defined by

$$H(x(t)) = \sum_{i=1,\ldots,n}(x_i(t) - x_i(t-1))^2/2 = \sum_{i=1,\ldots,n}\left(h\left(w_{i(i-1)}x_{i-1}(t-1) - x_i(t-1)\right)^2/2,\right.$$

which equals the total discrete kinetic energy and the half of the global dynamic frustration D(x(t)). The result remains available if automata network is a circuit in which transition functions are Boolean identity or negation.

Proof: It is easy to check that H(x(t)) = D(x(t))/2.

Proposition 1 is used to estimate the evolution of the robustness of a network because from Demongeot et al. (2014a, 2014b), it results that the quantity $E = E_\mu - E_{attractor}$ (= $\log 2^n - E_{attractor}$, if μ is uniform), called *evolutionary entropy,* serves as a robustness parameter, being related to the capacity that a getBren has to return to μ, the equilibrium measure, after endogenous or exogenous perturbation. $E_{attractor}$ can be evaluated by the quantity

$$E_{attractor} = -\Sigma_{k=1,m\leq 2}\, n\mu(C_k)\log\mu(C_k),$$

where *m* is the number of attractors and $C_k = B(A_k) \cup A_k$ is the union of the attractor A_k and of its attraction basin $B(A_k)$. A systematic calculation of $E_{attractor}$ allows quantifying the complexification of a network ensuring a dedicated regulatory function in different species. For example, the increase of the inhibitory sources in up-trees converging on a conserved subgraph of a genetic network causes a decrease of its attractor number by cutting some inhibited circuits, hence a decrease of $E_{attractor}$ and an increase of the evolutionary entropy *E*, showing that the robustness of a network is positively correlated with its connectivity (*i.e.*, the ratio between the numbers of interactions and genes in the network). Propositions 2 and 3 give examples of extreme cases where the networks are either discrete potential (or gradient) systems, generalizing previous works on continuous dynamics in which authors attempt to explicit Waddington and Thom chreode's energy functions, conserved or dissipated. In Demongeot et al. (2012), a method was proposed for calculating the number of

attractors in the case of circuits with Boolean transitions reduced to identity or negation. These results about attractors counting constitute a partial response to the discrete version of the 16th Hilbert's problem and can be approached by using Hamiltonian energy levels. For example, for a positive circuit of order 8, it is easy to prove that, in case of parallel updating, we have only even values for the global frustration D (they are odd for a negative circuit), corresponding to different values of the period of the attractors (Table 7.6).

Table 7.6 Values of the Global Frustration D, Attractor Numbers and Periods for Positive Circuits of Order 8 with Boolean Transition Identity or Negation

***D* (Frustration)**	**Attractor Number**	**Attractor Period**
0	2	1
2	7	8
4	3	4
4	16	8
6	7	8
8	1	2

A3 TANGENT AND INTERSECTING CIRCUITS

The study of tangent and intersecting circuits in strong connected components of a genetic network is possible when interactions are either identity or negation (Demongeot et al., 2012; http://dev.biologists.org/content/131/12/2911/ F8.large.jpg), with a mixing rule monotonic when different arcs come on the same node. For example, such circuits with one or two genes in common are shown in Figure 7.10.

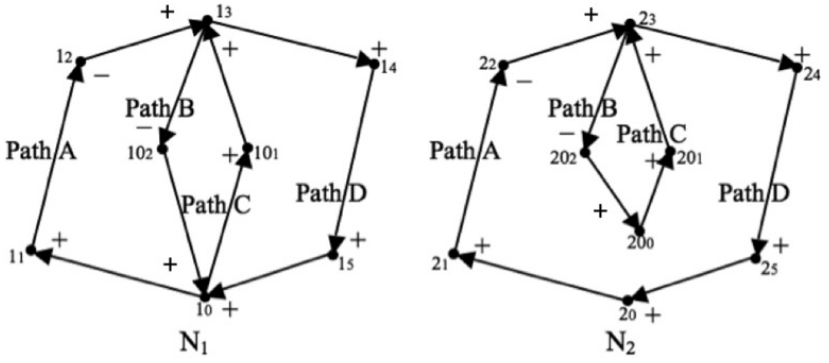


FIGURE 7.10

The two coupled networks N_i (i = 1,2) are each made of the subnetworks N_i and N_{i0}, whose vertices (or nodes) are denoted as ij (j = 0,…,5) and i0k (k = 0,…,2), respectively.

By looking on the networks of Figure 7.10, we see that each is made of four main paths of opposite senses: two are up (A and C), and two down (B and D), with the respective lengths of ℓ_A, ℓ_B, ℓ_C, and ℓ_D, and parities of s_A, s_B, s_C, and s_D, equal to 1 (resp. -1) if they have an even (resp. odd) number of negative arcs, with $s_A = \Pi_{a \in A}$ s_a,, where the sign s_a of the arc a of A is equal to -1 if a is negative (inhibition) and 1 if a is positive (activation).

For example, in Figure 7.10, the path A of N_1 is such as $\ell_A = 3$ and $s_A = -1$. The main paths have in N_1 two common nodes, 10 and 13, and in N_2, only one common node 23. Finally, with N_1, having four paths and two common nodes, there are four combinations giving four possible circuits (A,B), (B,C), (A,D), and (D,C), with a circuit like (A,C) having the parity $s_{A,C} = s_A\, s_C$. In the following discussion, to facilitate the reasoning, we will suppose that the state 0 of a gene is replaced by the state -1.

Let us denote as ((J,L),(L,M)) the general couple of circuits inside the set of the six possible couples created from these four circuits. If N((J,K),(L,M)) denotes the number of possible attractors of ((J,K),(L,M)), then the attractor number of N_1 is the minimum of the values of N((J,K),(L,M)) for the six couples of circuits. We conjectured (Demongeot and Moreira, 2007) that this number was less than the attractor number of N_2, the attractors of N_1 being those that allow the two common nodes to have the same state for each couple of circuits.

The minimum can be made more precise: suppose that two circuits (K,L) and (M,) have two common nodes with the same state. Any attractor of these intersected circuits has the same configurations as one of those of the tangent circuits that we can build by decoupling one on their two common nodes. Indeed, if not, there is at least one node different from the common nodes having an asymptotic sequence of states different for N_1 and N_2. Starting from this node and following the path going from this node to the common node of N_2, we would find for this node a sequence of states different that those observed in N_2 attractors, which is impossible. This reduces considerably the possible attractors for intersecting circuits, since they must always have the same state on their two intersected nodes. Then, counting attractors corresponds to a known combinatorial problem generalizing the necklace problem (*cf.* Demongeot et al., 2012; and Table 7.7). Then two new questions remain open:

Q1: Is a given attractor of the network N_2 respecting the constraint $x_{20_0}(t) = x_{2_0}(t))$, identical to an attractor of N_2?
Q2: If the answer to Q1 is yes, what are the constraints of its period?

The following propositions partially address Q1 and Q2.

Proposition 4 The attractors of the network N_2 respecting the constraint $x_{20_0}(t) = x_{2_0}(t))$, are the precise attractors of N_1.

Proof. From an initial condition identical for N_2 and N_1, where $x_{20_0}(0) = x_{2_0}(0) = x_{1_0}(0)$, the trajectories are the same for all the nodes, if $x_{20_0}(t) = x_{2_0}(t)$, for any $t \geq 1$. For the node 1_0 of the network N_1, we have, if the mixing rule is monotonic (*e.g.*, $\vee$ in N_1 and N_2):

$$x_{1_0}(t) = [s_B \times x_{1_3}(t-\ell_B)] \vee [s_D \times x_{1_3}(t-\ell_D)]$$
$$x_{20_0}(t) = s_B \times x_{2_3}(t-\ell_B) \text{ and } x_{2_0}(t) = s_D \times x_{2_3}(t-\ell_D).$$

Note that the reasoning would be the same with rule $\wedge$ or any composition of $\vee$ and $\wedge$.

By imposing $x_{20_0}(t) = x_{2_0}(t)$, for any $t \geq 1$, then $s_B \times x_{2_3}(t-\ell_B) = s_D \times x_{2_3}(t-\ell_D)$, and we have in the network N_1:

$$x_{1_0}(t) = [s_B \times x_{1_3}(t-\ell_B)] \vee [s_D \times x_{1_3}(t-\ell_D)] = x_{20_0}(t) = x_{2_0}(t).$$

By recurrence on t, this common value for $x_{1_0}(t)$, $x_{20_0}(t)$ and $x_{2_0}(t)$ is equal to

$$s_B \times x_{1_3}(0) = s_B \times x_{2_3}(0), \text{ for any } t = k\ell_B, \text{ with } k \geq 0.$$

The same reasoning can apply to $t \equiv 1,\ldots, \ell_B\text{-}1 \pmod{\ell_B}$. Then, the trajectories being the same, the attractors of N_2 with the constraint $x_{20_0}(t) = x20(t)$, are attractors of N_1.

Let us consider now an attractor of N_1, for which $x_{1_0}(t) = [s_B \times x_{1_3}(t\text{-}\ \ell_B)] \vee [s_D \times x_{1_3}(t\text{-}\ \ell_D)]$. If we identify $x_{1_0}(t)$ and $x_{2_0}(t)$, then, if $x_{1_3}(0) = x_{2_3}(0)$, this attractor is an attractor of N_2, where the two circuits (tangent in 2_3) have the signs $\sup(s_A, s_B)s_C$ and $\sup(s_A, s_B)s_D$, respectively, and where $x_{20_0}(t) = x_{2_0}(t)$

Remark If $\ell_D = \ell_B$, the constraint $s_D = s_B$ is necessary for observing in N_2 the coupling $x_{20_0}(t) = x_{2_0}(t)$. If not, the attractors of N_2 are not necessarily attractors of N_1.

Proposition 5 The attractors of N_2, with the coupling $x_{20_0}(t) = x_{2_0}(t)$, are characterized by the following property on their period p:

- If $s_B \times s_D = 1$ (resp. $s_A \times s_C = 1$), we have
  - p divides $(\sup(\ell_B, \ell_D) - \inf(\ell_B, \ell_D)) \Leftrightarrow p|(\sup(\ell_B, \ell_D) - \inf(\ell_B, \ell_D))$ (resp. p divides $(\sup(\ell_A, \ell_C) - \inf(\ell_A,\ell_C)) \Leftrightarrow p|(\sup(\ell_A,\ell_C) - \inf(\ell_A,\ell_C))$
  - if $s_B \times s_D = -1$ (resp. $s_A \times s_C = -1$), p does not divide $(\sup(\ell_B,\ell_D) - \inf(\ell_B,\ell_D))$ and p divides $2(\sup(\ell_B,\ell_D) - \inf(\ell_B,\ell_D)) \Leftrightarrow \neg[p|(\sup(\ell_B,\ell_D) - \inf(\ell_B,\ell_D))] \wedge p|2(\sup(\ell_B,\ell_D) - \inf(\ell_B,\ell_D))$ (resp. p does not divide $(\sup(\ell_A,\ell_C) - \inf(\ell_A,\ell_C))$ and p divides $2(\sup(\ell_A,\ell_C) - \inf(\ell_A,\ell_C)) \Leftrightarrow \neg[p|(\sup(\ell_A,\ell_C) - \inf(\ell_A,\ell_C))] \wedge p|2(\sup(\ell_A,\ell_C) - \inf(\ell_A,\ell_C))$.

Proof. If p denotes the length (or period) of an attractor of N_2, with the coupling $x_{20_0}(t) = x_{2_0}(t)$, then we have

$$\forall t \geq 1, x_{20_0}(t) = x_{2_0}(t) \Leftrightarrow \forall t \geq 1, s_B \times x_{2_3}(t-\ell_B) = s_D \times x_{2_3}(t-\ell_D) \Leftrightarrow$$
$$\forall t \geq 1, x_{2_3}(t) = s_B \times s_D \times x_{2_3}(t+\ell_B-\ell_D)$$

and

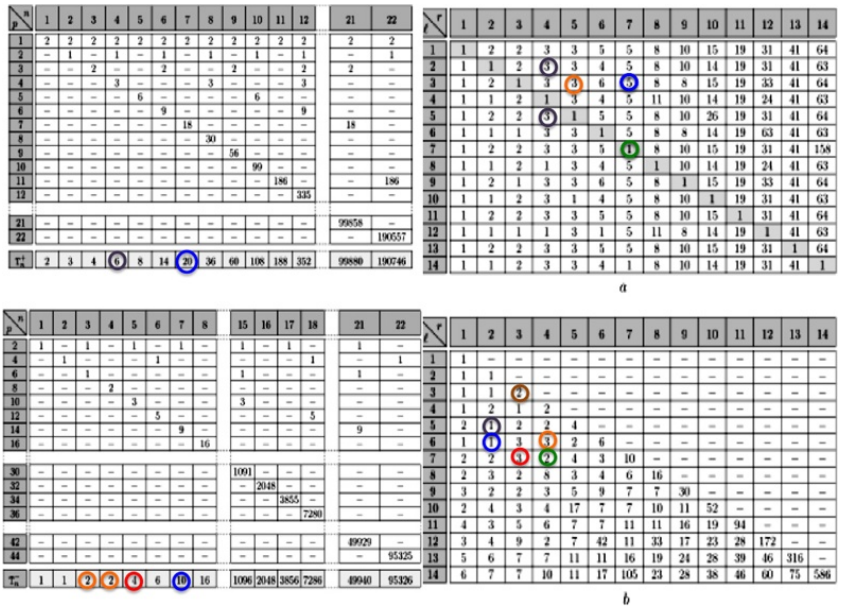
$$\forall t \geq 1, x_{20_0}(t) = x_{2_0}(t) \Leftrightarrow \forall t \geq 1, s_A \times x_{2_3}(t-\ell_A) = s_C \times x_{2_3}(t-\ell_C) \Leftrightarrow$$
$$\forall t \geq 1, x_{2_3}(t) = s_A \times s_C \times x_{2_3}(t+\ell_A-\ell_C).$$

Hence, we have, if $s_B \times s_D = 1$: $x_{20_0}(t) = x_{2_0}(t) \Leftrightarrow \forall\ t \geq 1,\ x_{2_3}(t) = x_{2_3}(t + sup(\ell_B,\ell_D) - inf(\ell_B,\ell_D)) \Leftrightarrow$

$$\forall t \geq 1, p|(\sup(\ell_B, \ell_D) - \inf(\ell_B, \ell_D)).$$

The proof in the case of $s_B \times s_D = -1$ is similar

Table 7.7 *Left:* Total number of attractors of period p for positive (top) and negative (bottom) circuits of order n. *Right:* Total number of attractors in case of tangent circuits, where (a) the left circuit is negative and the right circuit positive and (b) both side circuits are negative with parallel updating and $T=0$ (after Demongeot et al., 2012]). .



A4 STATE-DEPENDENT UPDATING SCHEDULE

A last important feature of the getBren dynamics is the existence of genes influencing directly the opening of the DNA inside the chromatine, hence allowing or disallowing the gene expression. If these genes are controlled by microRNA, it is necessary to generalize the getBren structure by considering that the possibility to update a block of genes at iteration t depends on the state of r clock genes (*i.e.*, involved in the chromatine updating clock) $k_1,\ldots k_r$ (like histone acetyltransferase, endonucleases, exonucleases, helicase, replicase, and polymerases) depending on s microRNAs, $l_1,\ldots,l_s$. Then the transition for a gene i, such as i, does not belong to $\{k_1,\ldots k_r\}$, could be written as

$$\forall g \in \{0,1\}, \beta \in \{0,1\}^n, \text{if } \forall j=1,\ldots,r, x_{kj}(t)=1,$$

then

(i) $P_{i,g}{}^{\beta}(\{x_i(t+1)=g|x(t)=\beta\})=\exp[g(\Sigma_{j\in Ni}\, w_{ij}\beta_j\text{-}\theta_i)/T]/[1+\exp[(\Sigma_{j\in Ni}\, w_{ij}\beta_j\text{-}\theta_i)/T]$, if microRNAs $l_1,\ldots,l_s$ are dominant;

(ii) $P_{i,g}{}^{\beta}(\{x_i(t+1)=\beta_i|x(t)=\beta\})=1$, if not.

The case (i) implies that $\forall\, j=1,\ldots,\ s,\ x_{lj}(t\text{-}1)=0$. To make the transition rule more precise, we can, for the sake of simplicity, decide that the indices $k_1,\ldots k_r$ of the r clock genes are $1,\ldots,\ r$ and then we have the three possible following behaviors:

- If $y(t)=\Pi_{i=1,\ldots,r}\ x_i(t)=1$, then rule (ii) is available.
- If $y(t)=0$ and $\Sigma_{s=t,\ldots,t\text{-}c}\ y(s)>0$, then $x(t+1)=x(t\text{-}s^*)$, where s^* is the last time before t, where $y(s^*)=1$.
- If $y(t)=0$ and $\Sigma_{s=t,\ldots,t\text{-}c}\ y(s)=0$, then $x(t+1)=0$ (by exhaustion of the pool of genes still in expression).

The dynamical system remains autonomous with respect to the time t [*i.e.*, it depends on t only through the set of state variables $\{x(t\text{-}c),\ldots,x(t\text{-}1)\}$], but a theoretical study of its attractors (as in Demongeot et al., 2012), with a state-dependent updating schedule is difficult to perform and will be investigated further in the future.

A5 THE CIRCULAR HAMMING DISTANCE

The most usual way to compare vectors with values in a finite alphabet is through the Hamming distance. Given two vectors $x,\ y \in A^n$, the Hamming distance between them is

$$d_H(x,y)=\#\{i\in\{0,\ldots,n-1\}:x_i\neq y_i\}.$$

In other words, it is the number of positions in which the values of the vectors differ. The function d_H is a metric: it is nonnegative and symmetric, it satisfies the triangle inequality, and a null distance implies identity of the vectors. It is also easy to see that

$$\forall i\in\{0,\ldots,n-1\},\ d_H(x,y)=d_H\left(\sigma^i(x),\sigma^i(y)\right),\ \text{and hence}\ d_H\left(x,\sigma^i(y)\right)=d_H\left(\sigma^{-i}(x),y\right).$$

Using this last property, we define the circular Hamming distance between two rings $[x]$ and $[y]$ as

$$d^c{}_H([x],[y])=\min\ d_H\left(x,\ \sigma^k(y)\right)$$
$$0\leq k\leq n-1$$

In general, the minimum between two metrics is not necessarily a metric, but here it holds.

Lemma 1. d^c_H is a metric on $A^n/\equiv$.

Proof. 1. If $d^c_H([x],[y])=0$, this implies that there exists k such that $d_H(x,\sigma^k(y))=0$; hence:

$$x=\sigma^k(y)\ \text{and}\ [x]=[y].$$

2. Let us now prove the symmetry:

$$d^c{}_H([x],[y])=\min_k d_H\left(x,\sigma^k(y)\right)=\min_k d_H\left(\sigma^{-k}(x),y\right)=\min_k d_H\left(y,\sigma^{-k}(x)\right)=d^c{}_H([y],[x]).$$

3. Let $[x],[y],[z]\in A^n/\equiv$. We must show that the triangular inequality is satisfied; i.e., that:

$$d^c{}_H([x],[y])\leq d^c{}_H([x],[z])+d^c{}_H([z],[y]).$$

Let i, j be such that: $d^c_H([z], [x]) = d_H(z,\sigma^i(x))$, $d^c_H([z], [y]) = d_H(z,\sigma^j(y))$

In addition, we define

$$(y)) \leq d_H\left(\sigma^i(x), z\right) + d_H\left(z, \sigma^j(y)\right) = d^c{}_H([x], [z]) + d^c{}_H([z], [y])$$

The cumulative distribution function $F_{n,k}$ of the circular Hamming proximity $p^c_H = n - d^c_H$, defined by k permutations of a ring of length n can be calculated from the cumulative function $G_{n,p}$ of the binomial law B(n,p) of order n, by the formula:

$$\forall i \in N,\ F_{n,k}(i) = P(\{p^c{}_H \leq i\}) = P\left(\left\{ \sup_{j=1,k} p_{H,j} \leq i \right\}\right) = P\left(\cap_{j=1,k}\left\{ p_{H,j} \leq i \right\}\right) = G_{n,p}(i)^k$$

For example, if n = 22, k = 22, and p = 4/16 = 1/4 (*i.e.,* the case of the circular Hamming distance between a small RNA of length 22 and the sequence AL):

$$F_{22,22}(12) = G_{22,1/4}(12)^{22} \approx (0.9993)^{22} \approx 0.985$$
$$F_{22,22}(13) = G_{22,1/4}(13)^{22} \approx (0.99998)^{22} \approx 0.9996$$

Note the following: the probability that the number of matches of a ring of length 22 with a linear sequence of length 22 is k or more is equal to $P_k = 1 - F_{22,22}(k-1) = 1 - G_{22,1/4}(k\text{-}1)^{22}$; *i.e.*, the probability that the circular Hamming distance is strictly less than k. For k = 14, $P_{13} = 1 - G_{22,1/4}(12)^{22:;}$ is about 1.5%. In the same way, the probability $P_{14} = 1\text{-}G_{22,1/4}(13)^{22}$ is about 0.4‰, and $P_{17} = 1\text{-}G_{22,1/4}(16)^{22}$ is about $8.8\ 10^{-6}$. If the linear sequence is of length 129126 (like ciRs7), then the probability to observe at least 17 matches once is about 5%.

In addition, if n = 22 and p = 6/16 = 3/8 = 0.375 (*i.e.,* the case of Hamming proximity with the constraint to match a substring of length 5 like AUGGU or UGGUA and authorize A–U, C–G and U–G coupling, between a small RNA of length 22 containing the substring and the sequence AL):

$$F_{22}(13) = G_{22,3/8}(13) \approx 0.9885$$

Hence, the probability $P_{14} = 1\text{-}F_{22}(13)$ that Hamming proximity is 14 or more, is about 1.5‰.

A6 THE ARCHETYPAL SEQUENCE AL

In Demongeot and Moreira (2007), a sequence of bases called AL (for ArchetypaL) is described as follows: 5'-UGCCAUUCAAGAUGAAUGGUAC-3' corresponding to a putative circular RNA with a possible hairpin form (*cf.* Figure 7.6, and Demongeot and Moreira, 2007; Meyer and Nelson, 2011; Turk-Mcleod et al., 2012; de Vladar, 2012; Yarus, 1988, 2010, 2013). AL can serve as a primitive ribosome in the sense that its circular form can bind any amino acid [with the weak electromagnetic or van der Waals interactions described in the Direct RNA Templating (DRT) hypothesis on the origin of the genetic code, which is still under debate (Demongeot and Moreira, 2007; Meyer and Nelson, 2011; Turk-Mcleod

et al., 2012; de Vladar, 2012; Yarus, 1988, 2010, 2013) to one of the triplets of its synonymy class in the genetic code, allowing the formation of small peptides (Yarus, 2013).

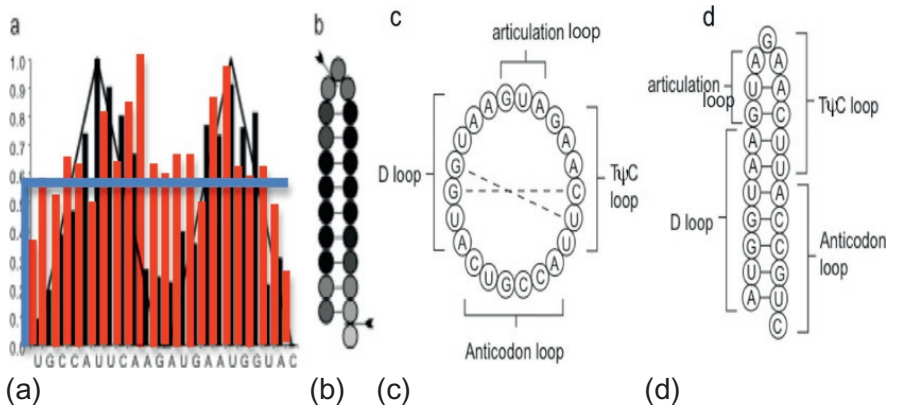


FIGURE 7.11

Relative frequencies of AL quintuplets in word matches, with Rfam sequences (Griffiths-Jones et al., 2005) in gray and ciRs7 sequence in red. For better comparison, the values have been normalized into the [0,1] interval by dividing by the maximal frequency (250/129126 for ciRs7). (a) The line is the (also normalized) distance of each base of AL with respect to the two interbase positions marked in (b); (c) shows the circular sequence of AL, with the correspondence with the tRNA loops; (d) shows the values graphed in (a) as shades of gray on the hairpin form of AL (Demongeot, 1978), with white and black representing the minimum and maximum values, respectively.

The sequence AL share many subsequences, like quintuplets, with small RNAs coming from Rfam (Griffiths-Jones et al., 2005), a database containing information about noncoding RNA families and structured RNA molecules, like transfer RNA (tRNA).

If we compare AL to the sequence of the circular RNA ciRs7, we find qualitative similarities, with 17/22 quintuplets passing the 5% upper threshold of significance (the 5%-threshold number of occurrences of a quintuplet in ciRs being equal to $129126/1024 + 1.6 \times 11 = 144$) of an unrandom frequency of common triplets (Figure 7.11). The occurrence numbers of the 22 successive quintuplets of AL inside the sequence ciRs7 of length 129126 have the following values, with the local maxima in red:

> uucaa (Tψ-loop) 250 ucaag 154 caaga 146 aagau 163 agaug 163 gauga 122 augaa 211 ugaau 238 (articulation loop) gaaug 152 aaugg 145 auggu 156 (D-loop) uggua 120 gguac 62 guacu 90 uacug 143 acugc 129 cugcc 160 (anticodon-loop) ugcca 155 gccau 121 ccauu 198 cauuc 155 auuca 206.

The similarity between the function of circular RNAs, tRNAs and the sequence AL could come from the fact that they are all concerned by the protein synthesis, directly

for the tRNAs and its ancestor AL, and indirectly, by inhibiting translational inhibitors as the microRNAs. This functional proximity could explain the frequent presence as relics of subsequences of AL inside the ciRs and tRNAs.

In the case of the tRNAs, the similitude concerns the conserved (inside and between species) bases of their loops (*cf.* Figure 7.12 and Alexander et al., 2010; Brown et al., 1986; Demongeot and Moreira, 2007; Shigi et al., 2002; Ueda et al., 1992; Yu et al., 2011), as well as of some particular tRNAs, such as where the ordered sequences of the loops is identical to AL except for two bases (Demongeot and Moreira, 2007).

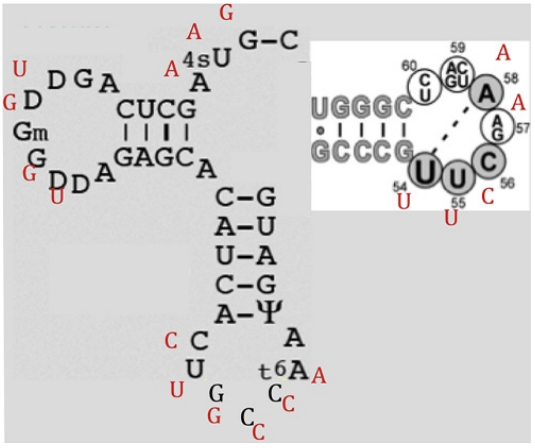


FIGURE 7.12

Matching of AL bases to the conserved bases of the tRNA loops (after Alexander et al., 2010; Brown et al., 1986; Demongeot and Moreira, 2007; Shigi et al., 2002; Yu et al., 2011).

ACKNOWLEDGMENTS

We acknowledge the financial support of the projects ANR-11-BSV5-0021, REGENR, EC Project VPH (Virtual Physiological Human), and Investissements d'Avenir VHP.

REFERENCES

Alexander, R.W., Eargle, J., Luthey-Schulten, Z., 2010. Experimental and computational determination of tRNA dynamics. FEBS Lett. 584, 376–386.

Bandiera, S., Rüberg, S., Girard, M., Cagnard, N., Hanein, S., Chrétien, D., Munnich, A., Lyonnet, S., Henrion-Caude, A., 2011. A nuclear outsourcing of RNA interference components to human mitochondria. PLoS One 6, e20746.

Bandiera, S., Matégot, R., Demongeot, J., Henrion-Caude, A., 2013. MitomiRs: delineating the intracellular localization of microRNAs at mitochondria. Free Radic. Biol. Med. 64, 12–19.

Bessho, K., Shanmukhappa, K., Sheridan, R., Shivakumar, P., Mourya, R., Walters, S., Kaimal, V., Dilbone, E., Jegga, A.G., Bezerra, J.A., 2013. Integrative genomics identifies candidate microRNAs for pathogenesis of experimental biliary atresia. BMC Syst. Biol. 7, 104.

Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C., Sbisà, E., 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. 192, 503–511.

Bulet, P., Hetru, C., Dimarcq, J.L., Hoffmann, D., 1999. Antimicrobial peptides in insects; structure and function. Dev. Comp. Immunol. 23, 329–344.

Choe, B.H., Kim, K.M., Kwon, S., Lee, K.S., Koo, J.H., Lee, H.M., Kim, M.K., Kim, J.C., 2003. The pattern of differentially expressed genes in biliary atresia. J. Korean Med. Sci. 18, 392–396.

de Vladar, H.P., 2012. Amino acid fermentation at the origin of the genetic code. Biol. Direct 7, 6.

Demongeot, J., 1978. Sur la possibilité de considérer le code génétique comme un code à enchaînement. Revue de Biomaths 62, 61–66.

Demongeot, J., Besson, J., 1983. Code génétique et codes à enchaînement I. C.R. Acad. Sc. Série III 296, 807–810.

Demongeot, J., Besson, J., 1996. Genetic code and cyclic codes II. C.R. Acad. Sc. Série III 319, 520–528.

Demongeot, J., Moreira, A., 2007. A circular RNA at the origin of life. J. Theor. Biol. 249, 314–324.

Demongeot, J., Waku, J., 2012. Robustness in biological regulatory networks. III application to genetic networks controlling the morphogenesis. Compt. Rendus Math. 350, 289–292.

Demongeot, J., Aracena, J., Thuderoz, F., Baum, T.P., Cohen, O., 2003. Genetic regulation networks: circuits, regulons and attractors. C. R. Biol. 326, 171–188.

Demongeot, J., Drouet, E., Moreira, A., Rechoum, Y., Sené, S., 2009a. MicroRNAs: viral genome and robustness of the genes expression in host. Phil. Trans. Royal Soc. A 367, 4941–4965.

Demongeot, J., Glade, N., Moreira, A., Vial, L., 2009b. RNA relics and origin of life. Int. J. Mol Sci. 10, 3420–3441.

Demongeot, J., Hazgui, H., Escoffie, J., Arnoult, C., 2010. Inhibitory regulation by micro-RNAs and circular RNAs. In: Medicon'13 IFBME Proceedings. Springer Verlag, New-York, pp. 722–725.

Demongeot, J., Elena, A., Noual, M., Sené, S., Thuderoz, F., 2011. "Immunetworks", attractors and intersecting circuits. J. Theor. Biol. 280, 19–33.

Demongeot, J., Noual, M., Sené, S., 2012. Combinatorics of Boolean automata circuits dynamics. Discr. Appl. Math. 160, 398–415.

Demongeot, J., Hazgui, H., Bandiera, S., Cohen, O., Henrion-Caude, A., 2013a. MitomiRs, Chloromi-Rs and general modelling of the microRNA inhibition. Acta Biotheor. 61, 367–383.

Demongeot, J., Cohen, O., Henrion-Caude, A., 2013b. MicroRNAs and robustness in biological regulatory networks. A generic approach with applications at different levels: physiologic, metabolic, and genetic. In: Aon, M.A., Saks, V., Schlattner, U. (Eds.), Systems Biology of Metabolic and Signaling Networks.In: Springer Series in Biophysics, vol. 16, pp. 63–114.

Demongeot, J., Hamie, A., Hansen, O., Franco, C., Sutton, B., Cohen, E.P., 2014a. Dynalets: a new method of modelling and compressing biological signals. Applications to physiological and molecular signals. C. R. Biol. 337, 609–624.

Demongeot, J., Ben Amor, H., Hazgui, H., Waku, J., 2014b. Robustness in neural and genetic regulatory networks: mathematical approach and biological applications. Acta Biotheor. 62, 243–284.
Elkon, R., Linhart, C., Halperin, Y., Shiloh, Y., Shamir, R., 2007. Functional genomic delineation of TLR-induced transcriptional networks. BMC Genomics 8, 394.
Franco, C.B., Scripture-Adams, D.D., Proekt, I., Taghon, T., Weiss, A.H., Yui, M.A., Adams, S.L., Diamond, R.A., Rothenberg, E.V., 2006. Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. Proc. Natl. Acad. Sci. U. S. A. 103, 11993–11998.
Georgescu, C., Longabaugh, W.J.R., Scripture-Adams, D.D., David-Fung, E.S., Yui, M.A., Zarnegar, M.A., Bolouri, H., Rothenberg, E.V., 2008. A gene regulatory network armature for T lymphocyte specification. Proc. Natl. Acad. Sci. U. S. A. 105, 20100–20105.
Gettings, M., Serman, F., Rousset, F.R., Bagnerini, P., Almeida, L., Noselli, S., 2010. JNK signalling controls remodelling of the segment boundary through cell reprogramming during drosophila morphogenesis. PLoS Biol. 8, e1000390.
Girard, M., Jannot, A.S., Besnard, M., Leutenegger, A.L., Jacquemin, E., Lyonnet, S., Henrion-Caude, A., 2011. Polynesian ecology determines seasonality of biliary atresia. Hepatology 54, 1893–1894.
Girard, M., Besnard, M., Jacquemin, E., Henrion-Caude, A., 2012. Biliary atresia: does ethnicity matter? J. Hepatol. 57, 700–701.
Gottwein, E., Corcoran, D.L., Mukherjee, N., Skalsky, R.L., Hafner, M., Nusbaum, J.D., Shamulailatpam, P., Love, C.L., Dave, S.S., Tushi, T., Ohler, U., Cullen, B.R., 2011. Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. Cell Host Microbe 10, 515–526.
Griffiths-Jones, S., Marshall, M., Khanna, A., Eddy, S.R., Bateman, A., 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33, 121–124.
Gurzov, E.N., Eizirik, D.L., 2011. Bcl-2 proteins in diabetes: mitochondrial pathways of β-cell death and dysfunction. Trends Cell Biol. 21, 424–431.
Hand, N.J., Horner, A.M., Master, Z.R., Boateng, L.A., LeGuen, C., Uvaydova, M., Friedman, J.R., 2012. MicroRNA profiling identifies miR-29 as a regulator of disease-associated pathways in experimental biliary atresia. J. Pediatr. Gastroenterol. Nutr. 54, 186–192.
Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., Kjems, J., 2011. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J. 30, 4414–4422.
Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., Kjems, J., 2013. Natural RNA circles function as efficient microRNA sponges. Nature 495, 384–388.
Hentze, M.W., Rouault, T.A., Caughman, S.W., Dancis, A., Harford, J.B., Klausner, R.D., 1987. A cis-acting element is necessary and sufficient for translational regulation of human ferritin expression in response to iron. Proc. Natl. Acad. Sci. U. S. A. 84, 6730–6734.
http://dev.biologists.org/content/131/12/2911/F8.large.jpg.
http://mirdb.org/miRDB/.
http://mirnamap.mbc.nctu.edu.tw/.
http://www.kegg.jp/kegg/pathway/dme/dme04310.
Kang, J., Pervaiz, S., 2013. Crosstalk between Bcl-2 family and Ras family small GTPases: potential cell fate regulation? Front. Oncol. 2, 206.
Kohsaka, T., Yuan, Z.R., Guo, S.X., Tagawa, M., Nakamura, A., Nakano, M., Kawasasaki, H., Inomata, Y., Tanaka, K., Miyauchi, J., 2002. The significance of human jagged 1 mutations detected in severe cases of extrahepatic biliary atresia. Hepatology 36, 904–912.

Lee, W.J., Kim, S.H., Kim, Y.S., Han, S.J., Park, K.S., Ryu, J.H., Hur, M.W., Choi, K.Y., 2000. Inhibition of Mitogen-Activated Protein Kinase (MAPK) by a Drosophila dual-specific phosphatase. Biochem. J. 349, 821–828.

Luedde, T., Heinrichsdorff, J., de Lorenzi, R., De Vos, R., Roskams, T., Pasparakis, M., 2008. IKK1 and IKK2 cooperate to maintain bile duct integrity in the liver. Proc. Natl. Acad. Sci. U. S. A. 105, 9733–9738.

Mantrova, E.Y., Hsu, T., 1998. Down-regulation of transcription factor CF2 by *Drosophila* Ras/MAPK signaling in oogenesis: cytoplasmic retention and degradation. Genes Dev. 12, 1166–1175.

Matte, U., Mourya, R., Miethke, A., Liu, C., Kauffmann, G., Moyer, K., Zhang, K., Bezerra, J. A., 2010. Analysis of gene mutations in children with cholestasis of undefined etiology. J. Pediatr. Gastroenterol. Nutr. 51, 488–493.

McEwen, D.G., Peifer, M., 2005. Puckered, a *Drosophila* MAPK phosphatase, ensures cell viability by antagonizing JNK-induced apoptosis. Development 132, 3935–3946.

Melliti, T., Noual, M., Regnault, D., Sené, S., Sobieraj, J., 2014. Full characterisation of attractors of two intersected asynchronous Boolean automata cycles. eprint arXiv, 1310.5747v2.

Meyer, S.C., Nelson, P.A., 2011. Can the origin of the genetic code be explained by direct RNA templating? BIO-Complexity 2011, 1–10.

Michon, F., Forest, L., Collomb, E., Demongeot, J., Dhouailly, D., 2008. BMP-2 and BMP-7 play antagonistic roles in feather induction. Development 135, 2797–2805.

Miyake, K., Ogata, H., Nagai, Y., Akashi, S., Kimoto, M., 2000. Innate recognition of lipopolysaccharide by Toll-like receptor 4/MD-2 and RP105/MD-1. J. Endotoxin Res. 6, 389–912.

Nouws, J., Shadel, G.S., 2014. MicroManaging mitochondrial translation. Cell 158, 477–478.

Ranganathan, P., Weaver, K.L., Capobianco, A.J., 2011. Notch signalling in solid tumours: a little bit of everything but not all the time. Nat. Rev. Cancer 11, 338–351.

Richard, A., 2011. Local negative circuits and fixed points in non-expansive Boolean networks. Discr. Appl. Math. 59, 1085–1093.

Robert, F., 1980. Itérations sur des ensembles finis et automates cellulaires contractants. Lin. Algebra Appl. 29, 393–412.

Saenz-Robles, M.T., Maschat, F., Tabata, T., Scott, M.P., Kornberg, T.B., 1995. Selection and characterization of sequences with high affinity for Engrailed proteins of *Drosophila*. Mech. Dev. 53, 185–195.

Sangokoya, C., Doss, J.F., Chi, J.T., 2013. Iron-responsive miR-485-3p regulates cellular iron homeostasis by targeting ferroportin. PLoS Genet. 9, e1003408.

Shigi, N., Suzuki, T., Tamakoshi, M., Oshima, T., Watanabe, K., 2002. Conserved bases in the TΨC loop of tRNA are determinants for thermophile-specific 2-thiouridylation at position 54. J. Biol. Chem. 277, 39128–39135.

Singh, V.B., Pavithra, L., Chattopadhyay, S., Pal, J.K., 2009. Stress-induced overexpression of the heme-regulated eIF-2a kinase is regulated by Elk-1 activated through ERK pathway. Biochem. Biophys. Res. Commun. 379, 710–715.

Turk-Mcleod, R.M., Puthenvedu, D., Majerfeld, I., Yarus, M., 2012. The plausibility of RNA-templated peptides: simultaneous RNA affinity for adjacent peptide side chains. J. Mol. Evol. 74, 217–225.

Ueda, T., Yotsumoto, Y., Ikeda, K., Watanabe, K., 1992. The T-loop region of animal mitochondrial $tRNA^{Ser}$ (AGY) is a main recognition site for homologous seryl-tRNA synthetase. Nucleic Acids Res. 20, 2217–2222.

Vickers, E.R., Sharrocks, A.D., 2002. The use of inducible Engrailed fusion proteins to study the cellular functions of eukaryotic transcription factors. Methods 26, 270–280.

Weil, G., Heus, K., Faraut, T., Demongeot, J., 2004. An archetypal basic code for the primitive genome. Theoret. Comp. Sc. 322, 313–334.
Xiao, J., Xia, S., Xia, Y., Xia, Q., Wang, X., 2014. Transcriptome profiling of biliary atresia from new born infants by deep sequencing. Mol. Biol. Rep. 41, 8063–8069.
Yarus, M., 1988. A specific amino acid binding site composed of RNA. Science 240, 1751–1758.
Yarus, M., 2010. Life from an RNA World: The Ancestor Within. Harvard University Press, Cambridge US.
Yarus, M., 2013. A ribonucleotide origin for life–fluctuation and near-ideal reactions. Orig. Life Evol. Biosph. 43, 19–30.
Yu, F., Tanaka, Y., Yamashita, K., Suzuki, T., Nakamura, A., Hirano, N., Suzuki, T., Yao, M., Tanaka, I., 2011. Molecular basis of dihydrouridine formation on tRNA. Proc. Natl. Acad. Sci. U. S. A. 108, 19593–19598.
Zeng, Y.A., Verheyen, E.M., 2004. Nemo is an inducible antagonist of Wingless signaling during Drosophila wing development. Development 131, 2911–2920.
Zhou, H., Wertz, I., O'Rourke, K., Ultsch, M., Seshagiri, S., Eby, M., Xiao, W., Dixit, V.M., 2004. Bcl10 activates the NF-kB pathway through ubiquitination of NEMO. Nature 427, 167–171.

Collège de France

Complexité-Simplexité | Alain Berthoz, Jean-Luc Petit

La simplexité, ultime avatar de la complexité ?

Jacques Demongeot, Hedi Ben Amor, Hana Hazgui et Athanasios Lontos

Texte intégral

1. Introduction

1 La simplexité peut être considérée comme l'émergence de nouvelles fonctionnalités, issues des interactions entre sous-systèmes (ou modules) d'un système complexe : elle apparaît ainsi comme l'ultime avatar de la complexité, ce qui permet de la définir, sans rupture épistémologique avec le courant scientifique des systèmes complexes, comme la tendance qu'a la nature de sélectionner les fonctions vitales des êtres vivants, en particulier celles les plus utiles à la spécificité et à la

survie d'une espèce, dans la catégorie des mécanismes simples à déclencher, à contrôler et à maintenir.

2 Nous prendrons, comme premier exemple de notre thèse, l'évocation mnésique, qui permet la reconnaissance rapide d'un mot ou d'une image, nécessaire à son identification contextuelle ultérieure (permettant de comprendre le discours d'un locuteur et de lui répondre) ou à un comportement réflexe, sans étape sémantique préalable (permettant la fuite devant l'image d'un prédateur, ou, à tout le moins, l'amorce d'un réflexe primaire, par exemple par adaptation pupillaire ou changement du focus binaural). Cet exemple permettra de mettre en lumière un mode de simplexification fréquent dans le système nerveux, impliquant un triple mécanisme de projection sur une variété lente de l'espace d'état du réseau de neurones impliqué, de moyennisation, et, enfin, de synchronisation des activités des nombreux modules supposés identiques de ce réseau.

3 Le second exemple concerne la réduction du nombre d'attracteurs dans les réseaux de contrôle immunologiques, due à la coalescence de réseaux de contrôle génétiques tangents en un de leurs sommets, représenté par un gène critique pour l'unicité de la fonction contrôlée, à savoir le gène de la recombinase RAG, qui permet la création de néo-gènes immuns dans le chromosome 14 de mammifères tels que la souris et l'homme. Cet exemple met en évidence l'importance des circuits dans le contrôle dynamique de l'expression génétique.

4 Le dernier exemple montre l'existence d'une optimisation de la forme globale d'une cellule ou d'un tissu, qui reçoit ses éléments nutritifs à travers sa surface externe, tendant ainsi à optimiser le rapport S/V de leur surface externe S à leur volume interne V. Ce mécanisme provoque à la fois la division cellulaire et l'arrêt de la croissance du tissu, pour une valeur critique de S/V, fixant ainsi la taille « adulte » de l'organe concerné.

2. Simplexification dans les systèmes dynamiques complexes

5 Le processus de simplexification d'un système dynamique complexe, formé de sous-systèmes (ou modules) en interaction, fait intervenir divers mécanismes, parmi lesquels on peut retenir les suivants :

– la projection sur une variété de l'espace d'état de vitesse lente (« central manifold ») ;

– la moyennisation modulaire et le passage à la limite thermodynamique (« mean field » et loi
des grands nombres), dans le cas de systèmes composés de multiples modules identiques ;
– le codage dynamique et la synchronisation de sous-systèmes ou modules (« phase locking ») ;
– la réduction dynamique par coalescence d'architectures de régulation génétique en interaction (« tangent circuits ») ;
– l'optimisation d'une fonction de coût « universelle » (comme l'index « Cell Surface/Volume Ratio », pour les cellules et tissus).

3. Synchronisation et évocation mnésique

6 Comment coder une information de type chaîne de caractères, image ou son, dans la trajectoire asymptotique d'un système dynamique, et comment retrouver cette information à la suite d'un stimulus provoquant son évocation ? Pour répondre à cette question, nous prendrons pour exemple le réseau hippocampique, dans lequel s'exercent les deux fonctions mnésiques de stockage à long terme et de requête (volontaire consciente ou réflexe post-stimulus) d'informations mémorisées. Nous considérerons deux types d'interaction[1] :
– les interactions lentes, dues aux synapses à récepteurs NMDA, dont la densité est élevée dans la région Cyto-Architecturale CA1 de l'hippocampe. Ces interactions génèrent des potentiels post-synaptiques excitateurs et s'accompagnent d'entrée de calcium dans les neurones CA1, stimulant ainsi la CAM-kinase 2, protéine kinase activée par le couple Calcium-Calmoduline, et impliquée dans le phénomène lent, dit LTP, de potentialisation à long terme, qui constitue le mécanisme de base des processus d'apprentissage et de mémorisation.
– les interactions rapides, dues aux synapses à récepteurs de type GABAergiques (inhibiteurs), ou Dopamine-, AMPA- et kaïnat-ergiques (activateurs)[2]. Ces interactions permettent l'évocation mnésique, avec possibilité (I) de distinction entre des souvenirs proches dans leur séquence de codage et leur acquisition contextuelle, et (II) de complétion par une connaissance pré-existante associée. La région CA3 maintient la balance dynamique entre ces deux processus duaux[3].

7 Nous retiendrons la structure classique[4] des interactions rapides entre les régions CA1 et CA3, par l'intermédiaire de la boucle poly-synaptique passant par le Gyrus Denté (DG), le SuBiculum

(SB) et le cortex entorhinal (EC), que l'on peut résumer sur la Figure 1.

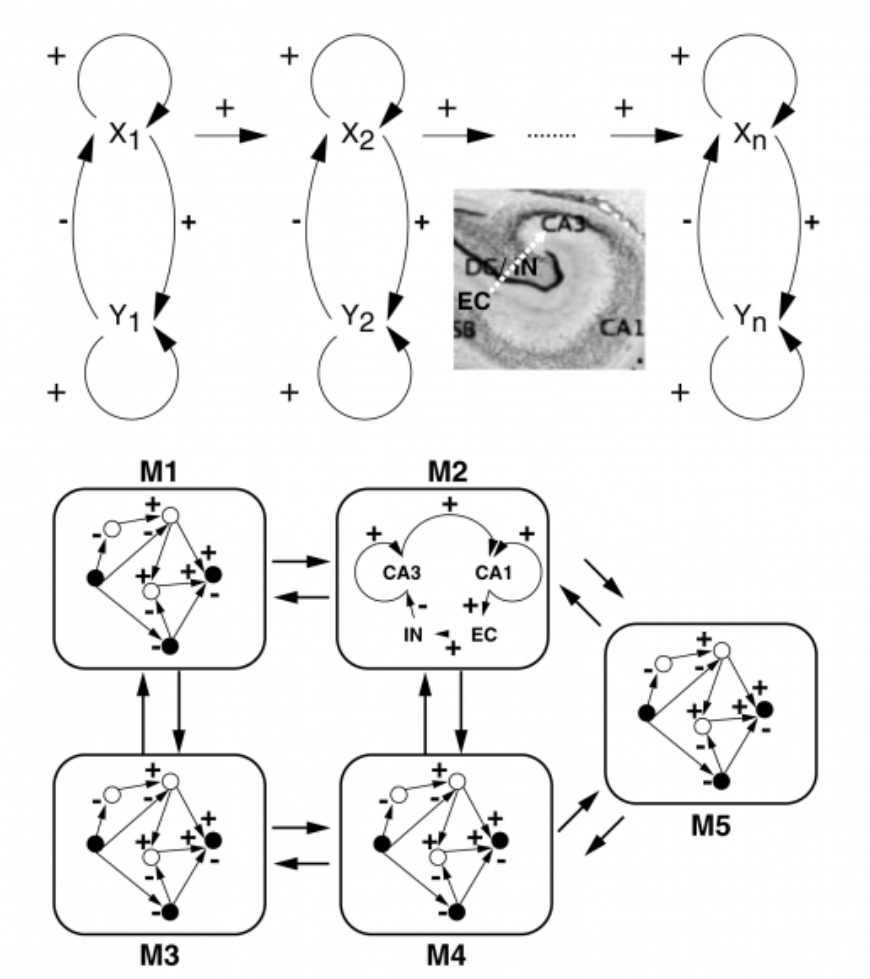


Figure 1a

En haut : représentation schématique de la boucle (Cortex Entorhinal EC/SB-SuBiculum-Gyrus Denté DG/IN InterNeurones-CA3-CA1-SB/EC) par une chaîne de régulons négatifs (circuit minimum comportant un circuit négatif et deux circuits positifs), où X_1 (resp. X_2) représente l'activité des neurones CA3 (resp. CA1). La photographie centrale représente l'anatomie de l'hippocampe dans son contexte, avec les groupes neuronaux CA3, CA1, DG/IN, EC et SB.

En bas : organisation modulaire, avec l'exemple d'un module M2 hippocampique.

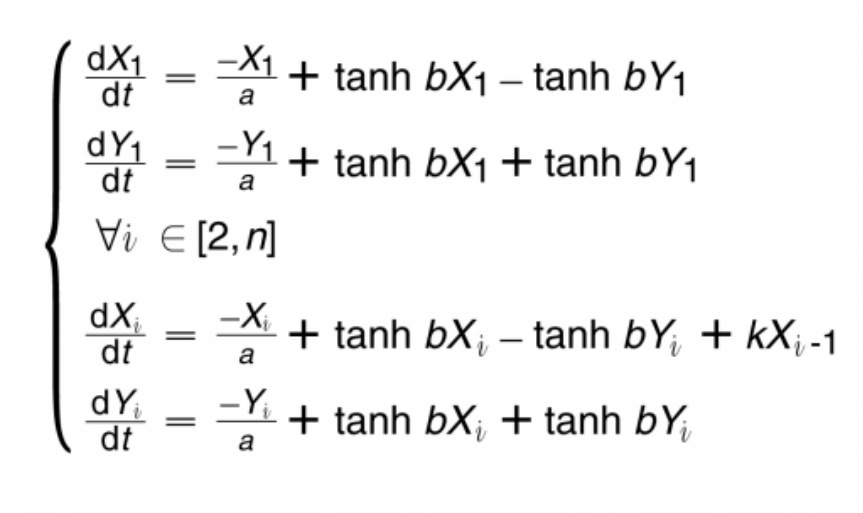


Figure 1b.

Système d'équations différentielles ordinaires, dites de Wilson-Cowan[5].

8 En l'absence de synchronisation[6], et en supposant que leurs conditions initiales ont été choisies uniformément au hasard, les modules régulons de la Figure 1 ne sont pas en phase, ce qui signifie que leur moyennisation « verticale », obtenue par passage à la limite thermodynamique (i.e., lorsque le nombre des modules tend vers l'infini, ce qui est une bonne approximation du grand nombre de modules observés dans l'hippocampe) donne un signal moyen égal à la moyenne temporelle « horizontale » de leur activité. Nous modéliserons cette activité, sur la variété rapide de l'espace d'état du système, par le système d'équations différentielles ordinaires, dites de Wilson-Cowan, donné sur la Figure 1b, dans lequel le terme k traduit le couplage entre les modules, b (resp. $1/a$) est un paramètre qui règle la raideur de la fonction tangente hyperbolique à l'origine, donc l'intensité des activités synaptiques inhibitrices ou activatrices (resp. l'intensité de la protéolyse) des neurotransmetteurs.

9 Ce système est proche, si l'on développe en série les fonctions tangentes hyperboliques (supposées quasi-linéaires à l'origine) pour les faibles valeurs de b (en supposant les activités X_i comprises entre 0 et 1), de la famille des systèmes dynamiques apparentés au système de Liénard[7] : dx/dt=y, dy/dt=g(x)+f(x)y, où g et f sont des polynômes. Cela permet de réunir les équations classiques de van der Pol, de FitzHugh-Nagumo (approximation 2D de Hodgkin-Huxley) et de Hopfield dans un même ensemble d'équations exprimant les interactions entre modules d'un système complexe, ici neuronal, et génétique dans les

paragraphes suivants (Figure 2).

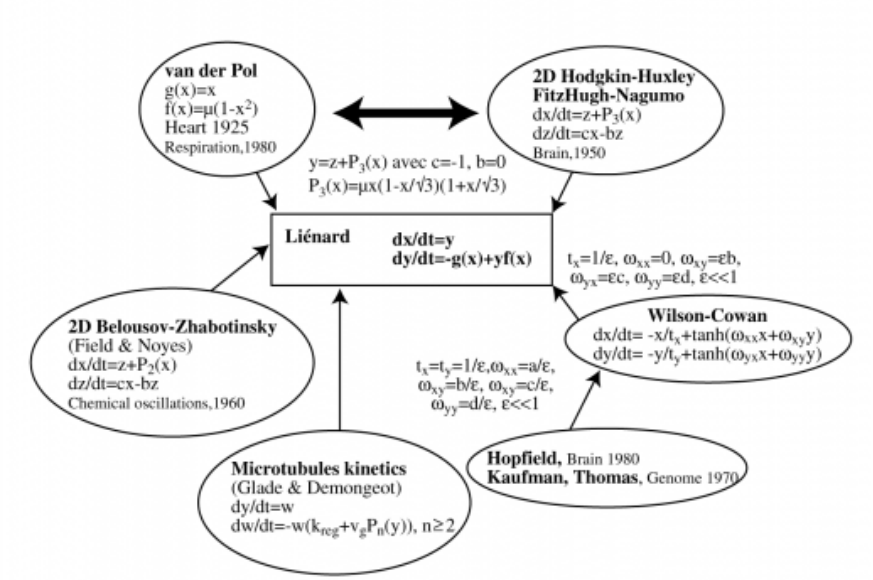


Figure 2.

Représentation simplifiée des systèmes dynamiques (en périphérie) apparentés au système de Liénard (au centre).

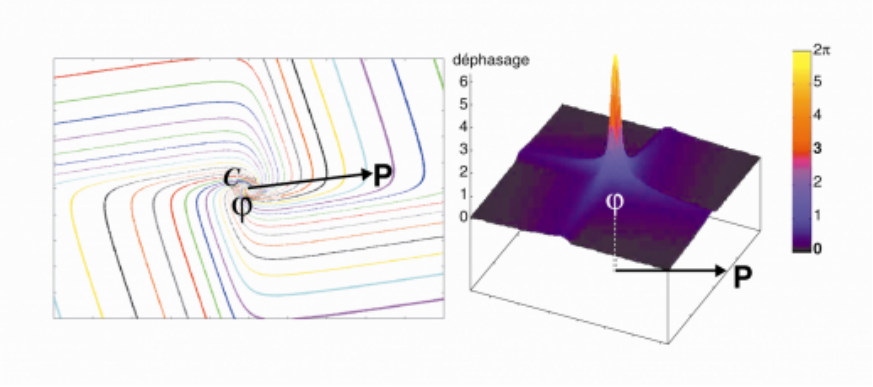


Figure 3.

Réseau isochrone du système de Wilson-Cowan (à gauche) et valeur (entre 0 et 2π) du déphasage consécutif à une perturbation P du point de phase φ du cycle limite *C* de l'oscillateur.

10 La synchronisation de modules identiques, dont la dynamique est gérée par les équations de Wilson-Cowan, est étudiée à l'aide de son réseau de fibres isochrones, ou isochron (Figure 3). L'isochron de phase φ d'un oscillateur de cycle limite *C*, est l'ensemble des conditions initiales ayant une dynamique asymptotique en phase avec la trajectoire issue, au temps 0, d'un point de *C* ayant une phase initiale φ, dans la période, ramenée à 2π, de l'oscillateur.

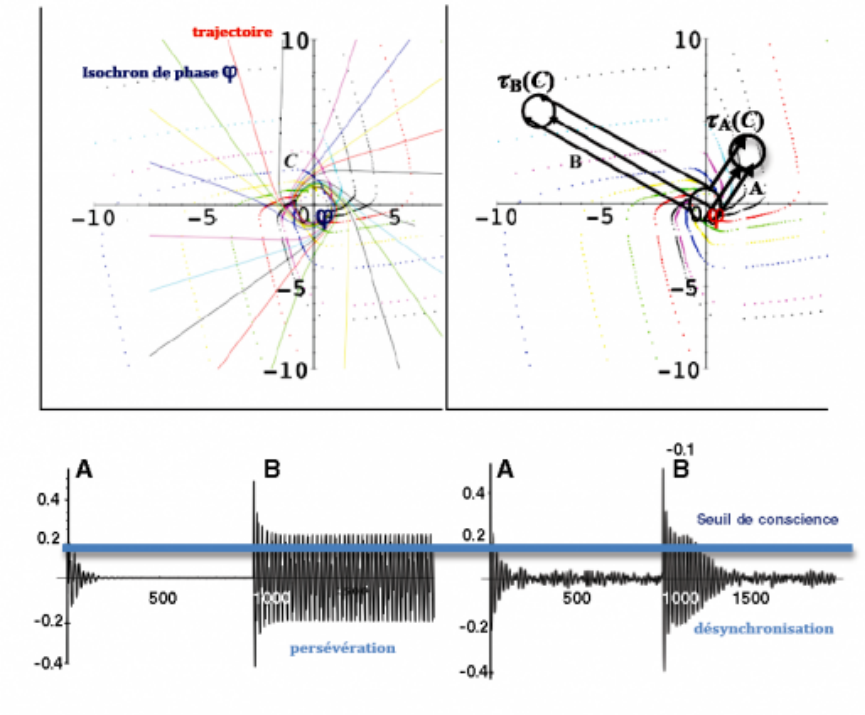


Figure 4.

Fibrations transverses isochrone et trajectorielle du système de Wilson-Cowan (en haut à gauche) ; perturbations instantanées désynchronisante (A) et synchronisante (B), avec les translatés correspondants du cycle-limite, $\tau_A(C)$ et $\tau_B(C)$, symbolisés par des cercles (en haut à droite) ; profils temporels en absence d'interaction intermodulaire, en cas de perturbations de type A et B (en bas à gauche) et en présence d'interaction intermodulaire (k = 0.1, en bas à droite), avec indication du seuil de conscience du comportement de persévération (à gauche) et de désynchronisation (à droite).

11 Une perturbation instantanée du système, représentée par un vecteur translation dans l'espace d'état de l'oscillateur, peut être synchronisante, si elle envoie tous les points du cycle limite C de l'oscillateur entre deux isochrons, ou au contraire désynchronisante si le translaté du cycle, noté $\tau(C)$, intersecte beaucoup d'isochrons (Figure 4). Supposons que cette synchronisation soit obtenue par une cascade de stimulations répétées régulièrement, aux bornes d'intervalles successifs de durée T. Soit le niveau de gris $G_{i,j}$ (supposé normalisé entre 0 et 2π) du pixel (i,j), issu des N pixels d'une image : $G_{i,j} = g_{i,j}2\pi/N$, où $g_{i,j}$ est un entier entre 0 et N. Supposons que ce niveau de gris $G_{i,j}$ soit codé dans un sous-ensemble de n oscillateurs identiques de Wilson-Cowan, par le $g_{i,j}^{ème}$ point, de phase $G_{i,j}2\pi/N$, dans l'ordre

de description temporelle des N points du cycle limite, de phases successives 0, 2π/N, 4π/N, 6π/N, ..., (N-1)2π/N).

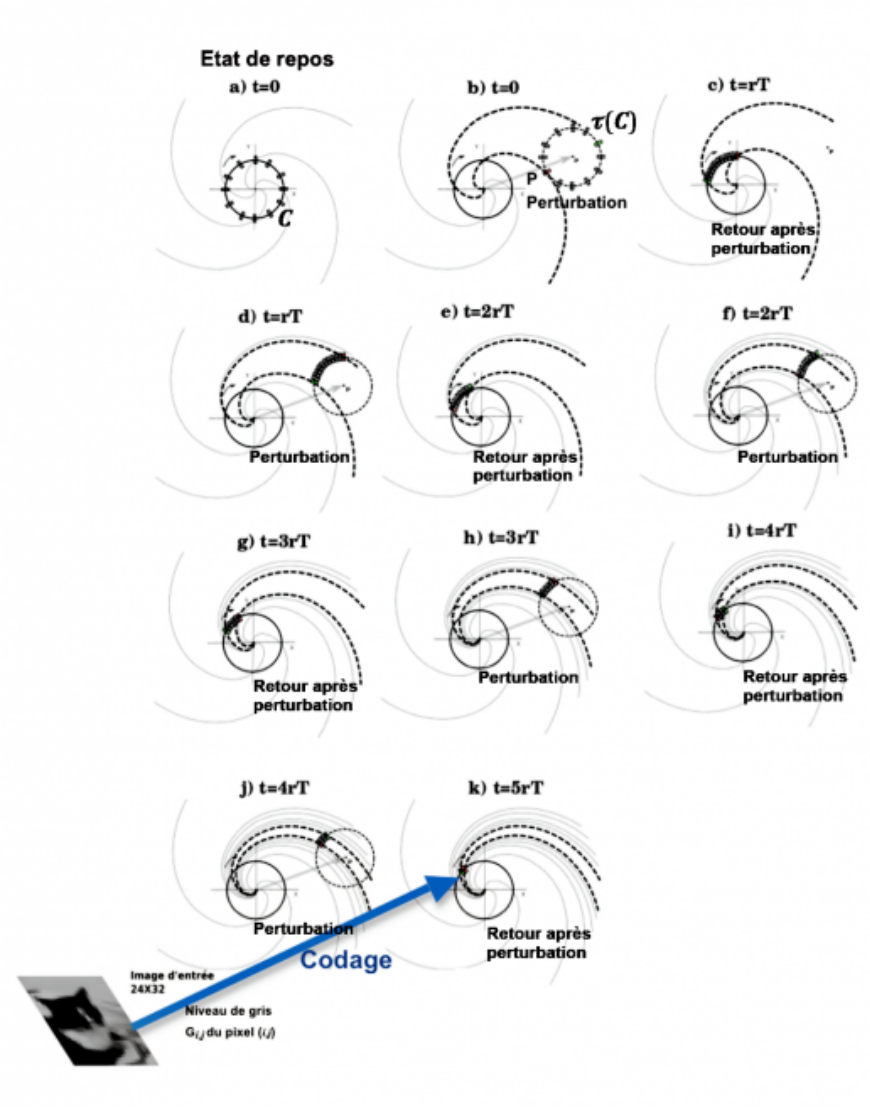


Figure 5.

Synchronisation progressive, par perturbations répétées de type B à intervalles de temps constants de durée T (en haut), avec codage du niveau de gris par la phase finale observée en retour après une perturbation instantanée de l'image d'un chat dans une grille d'oscillateurs de Wilson-Cowan, de cycle limite *C*.

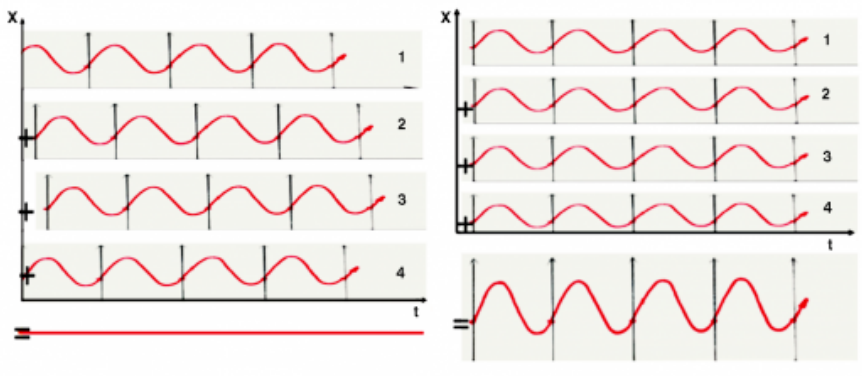


Figure 6.

Sommation, donnant un signal constant, des signaux temporels d'activation de modules désynchronisés (à gauche) ; sommation, en cas de synchronisation, révélant la « mélodie » sous-jacente (à droite).

12 Pour évoquer le niveau $G_{i,j}$ du pixel (i,j), on devra appliquer à l'état initial du réseau constitué des oscillateurs ayant des phases équi-réparties entre 0 et 2π (état désynchronisé de repos, Figure 6 à gauche), une stimulation synchronisante sur l'axe horizontal des X (intensité des neurones de la zone CA3, qui reçoivent et traitent, de manière multi-modale et rapidement apprise, les afférences sensorielles, en particulier celles de l'aire 17 optique[8]) : elle enverra chacun des oscillateurs au voisinage de l'isochron de phase $G_{i,j}$ (état synchronisé excité, Figure 6 à droite), selon le schéma de la Figure 5. On suppose donc que l'évocation de l'image de chat résulte de N stimulations (résultant de l'excitation par les N pixels de l'image) des n modules d'un système de Wilson-Cowan, initialement dans l'état de repos. La phase d'apprentissage de la « phasotopie » de la fibration isochrone peut se faire par présentation précoce répétée de l'image, stabilisant l'intensité de la réponse des neurones de CA3[9]. La désynchronisation sera obtenue naturellement par le couplage des oscillateurs des modules, qui seront remis à leur activité constante de repos (Figure 6).

13 L'exemple ci-dessus permet d'identifier le pouvoir « simplexificateur » de la synchronisation, qui « verrouille en phase » des activités neuronales codant des intensités de gris, dans des zones neuronales hippocampiques en homotopie de projection avec des zones de stimulation de l'aire 17 optique, elles-mêmes homologues des zones rétinotopiques périphériques. Un tel pouvoir devrait pouvoir faire l'objet de réalisations hardware neuromimétiques, dans lesquelles la recherche

d'information par RAM (Random Access Memory) dans des mémoires statiques serait remplacée par la perturbation (éventuellement en cascade, pour augmenter la synchronisation, donc l'intensité du signal mnésique) de mémoires dynamiques déphasées à l'état de repos et stimulées par une entrée de type image, son ou chaîne de caractères, le protocole d'I/O (Input/Output) étant numérisé et programmé, ou étant analogique.

14 La requête d'information par un tel processeur à mémoire dynamique aurait ainsi beaucoup d'analogie avec le processus d'évocation mnésique émergente hippocampique et permettrait de modéliser les processus attentionnels, en ne synchronisant qu'une partie du message (image ou signal) évoqué[10], qui est alors seul à passer le seuil de conscience (Figure 4). Il pourrait également simuler les phénomènes pathologiques de « persévération » dans la synchronie, qui ne permettent pas l'alternance physiologique rapide des phases de synchronisation et désynchronisation, nécessaires à la compréhension de stimuli successifs polysémiques, dans la fluidité d'évocations successives. De tels comportements de persévération sont observés dans des maladies dégénératives, comme la maladie d'Alzheimer.

4. Réduction dynamique en immunologie

15 Considérons, de manière générale[11], un réseau de contrôle génétique (ou métabolique) : il possède, comme sommets, des gènes (resp. protéines) et, comme arcs, des liens correspondant à des interactions d'inhibition (resp. activation), exercés par des complexes protéiques répresseurs (resp. inducteurs) de l'expression génique, ou par des substrats activateurs (resp. inhibiteurs) d'une activité fonctionnelle enzymatique, transporteuse, réceptrice, ...

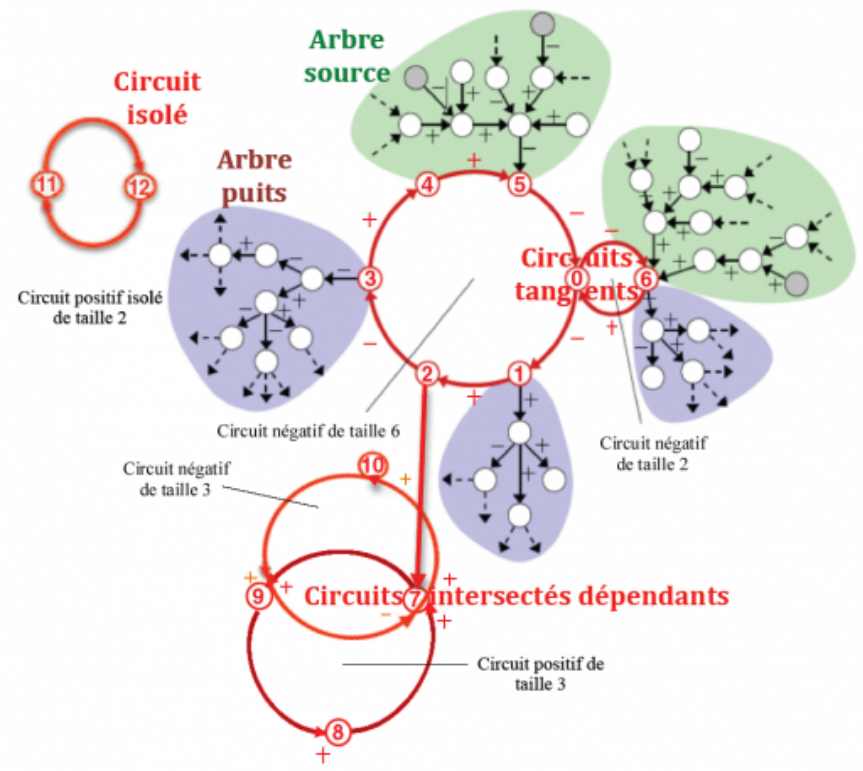


Figure 7.

Architecture générale d'un réseau de contrôle génétique, avec sa structure arborescente

16 L'architecture générale du réseau consiste en trois parties principales (Figure 7[12]) :
– une arborescence initiale, faite d'arbres sources, partant des gènes qui n'ont pas d'antécédents dans le réseau (gènes sources) ;
– une arborescence terminale, faite d'arbres puits, aboutissant à des gènes qui n'ont pas de successeurs dans le réseau (gènes puits) ;
– entre les deux, des composantes fortement connexes, c'est-à-dire connexes et maximales : connexes, car composées uniquement de circuits (chemins d'interaction fermés, dont les deux extrémités sont identiques) ; et maximales, car non contenues dans une composante connexe strictement plus grande. Elles sont faites d'un circuit isolé, ou de grappes de circuits tangents (un seul gène commun) ou intersectés (plus d'un gène commun) (Figure 7), pouvant être une arborescence de grappes dépendantes, situées hors de la composante source, mais non isolées.

17 Seules les composantes fortement connexes peuvent avoir plusieurs configurations attractantes, une fois fixées leurs conditions initiales et les états des gènes qui les précèdent dans

leurs arbre sources. Les états des arbres sources dépendent des états des gènes sources, et ceux des arbres puits dépendent des états des composantes fortement connexes. La syntaxe « grammaticale » dynamique générale d'un réseau est telle qu'un groupe contextuel, arbre source, instancie un attracteur du groupe « verbal », composante fortement connexe, qui agit sur un groupe objet, arbre puits, pour en fixer la configuration attractante, fixe ou périodique (cycle limite). Dans le cas d'un cycle limite, si l'on considère qu'il y autant d'oscillateurs que de cellules qui expriment cet attracteur, le signal global correspondant à une expression synchrone, par ces cellules, de la fonction codée par cet attracteur, est la résultante d'une mise en phase provoquée par une stimulation « évocatrice », comme dans le paragraphe précédent. La simulation de l'expression globale du réseau peut s'effectuer avec une règle de mise à jour des gènes parallèle ou séquentielle, ou guidée par l'horloge chromatinienne, qui est une mise à jour endogène, dépendant de l'état de certains gènes régulateurs temporels (histone-acétyltransférases, endonucléases, hélicases, réplicases, ligases, polymerases, ...).

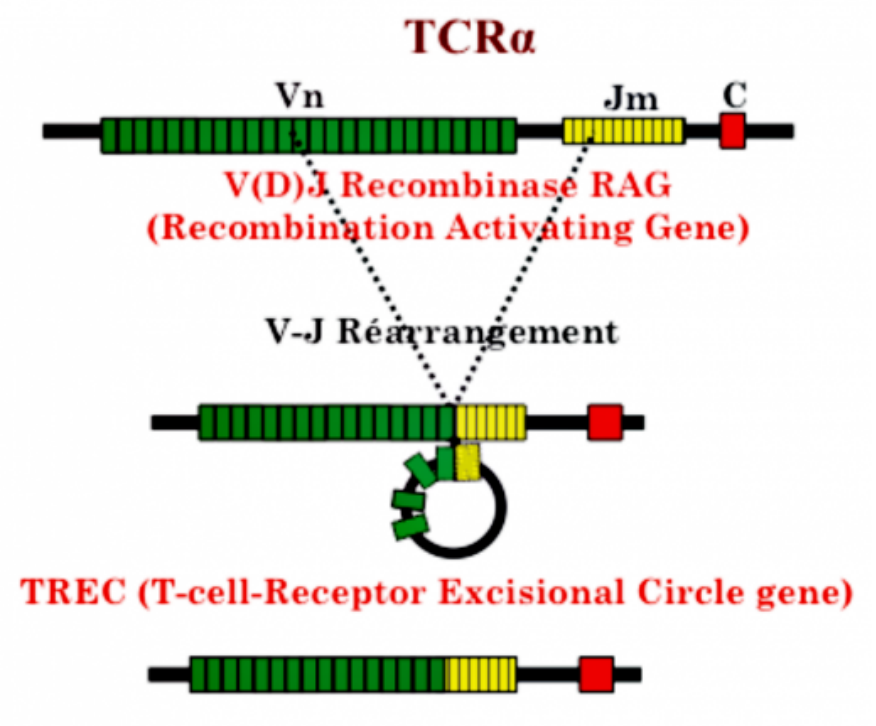


Figure 8.

Action de la recombinase, qui excise la boucle d'ADN contenant une partie des gènes V et J du chromosome 14 de la souris ou de l'homme, pour constituer un nouveau gène immun.

18 Considérons à présent le réseau de contrôle du gène

Recombinase RAG, qui assure la recombinaison des gènes de la zone V et de la zone J, pour donner les gènes des récepteurs α des lymphocytes T (TCRα) dans le chromosome 14 de la souris (Figure 8) et de l'homme[13] : la protéine RAG excise la boucle d'ADN contenant une partie des gènes V et J, pour reconstituer un nouveau gène immun, à condition que le triplet du gène V commençant ce gène et celui du gène de la zone J le terminant soient en phase de part et d'autre de la coupure.

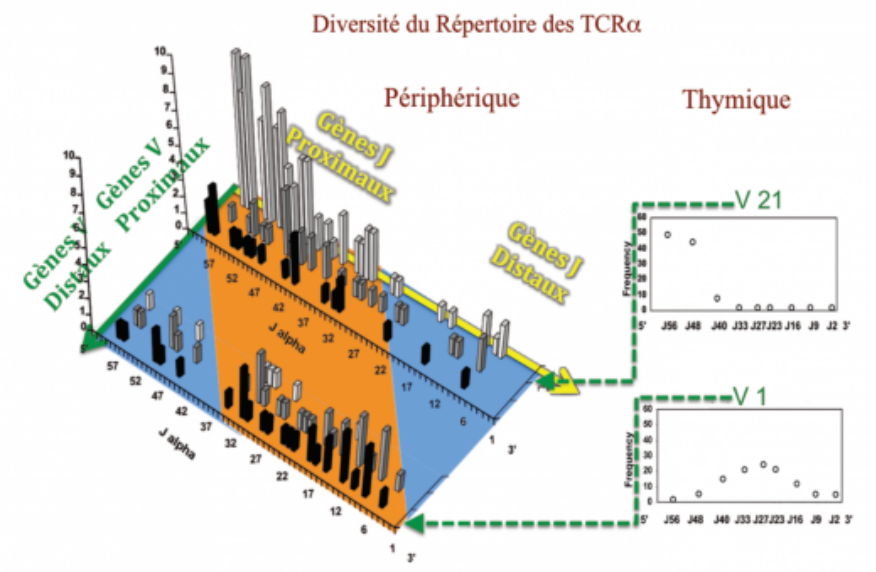


Figure 9.

Histogramme des fréquences de recombinaison entre les gènes V et J du chromosome 14 de la souris, dans les lymphocytes T circulants du système immunitaire périphérique (à gauche) et les thymocytes du système embryonnaire thymique (à droite).

19 La diversité de cette production de néo-gènes immuns est montrée sur la Figure 9 (gènes immuns de souris), où sont confrontées la diversité résiduelle du répertoire périphérique (dans les lymphocytes T circulants) et la distribution des recombinaisons V-J possibles, pour deux gènes V1 et V21, montrant la combinaison préférentielle des gènes V et J distaux (par rapport à la coupure), ainsi que celle des gènes V et J proximaux.

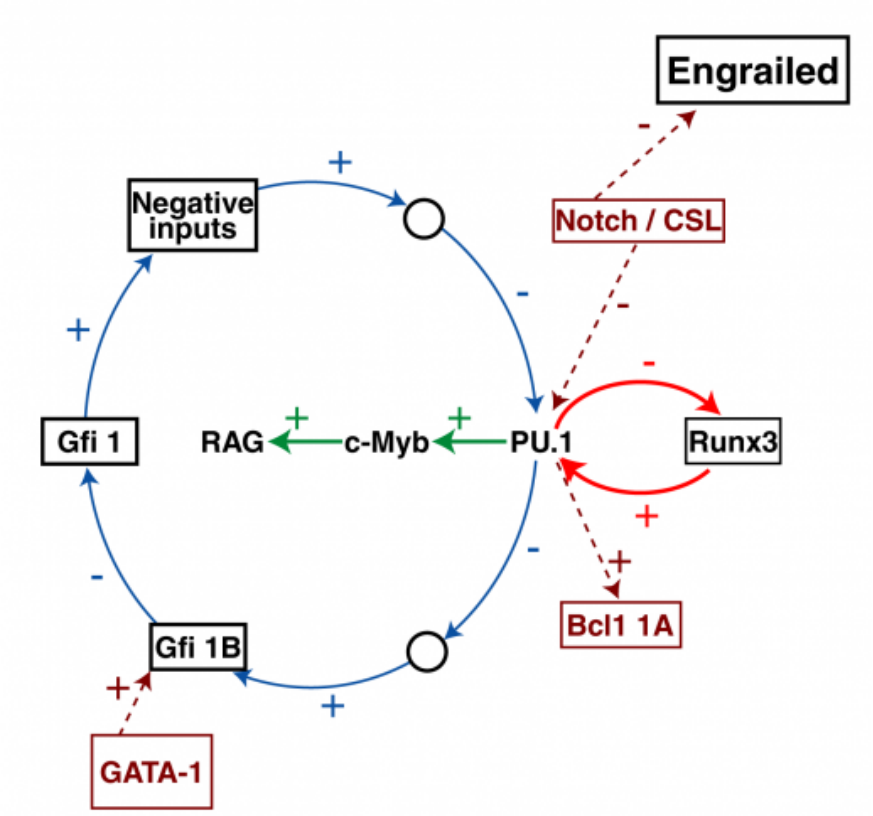
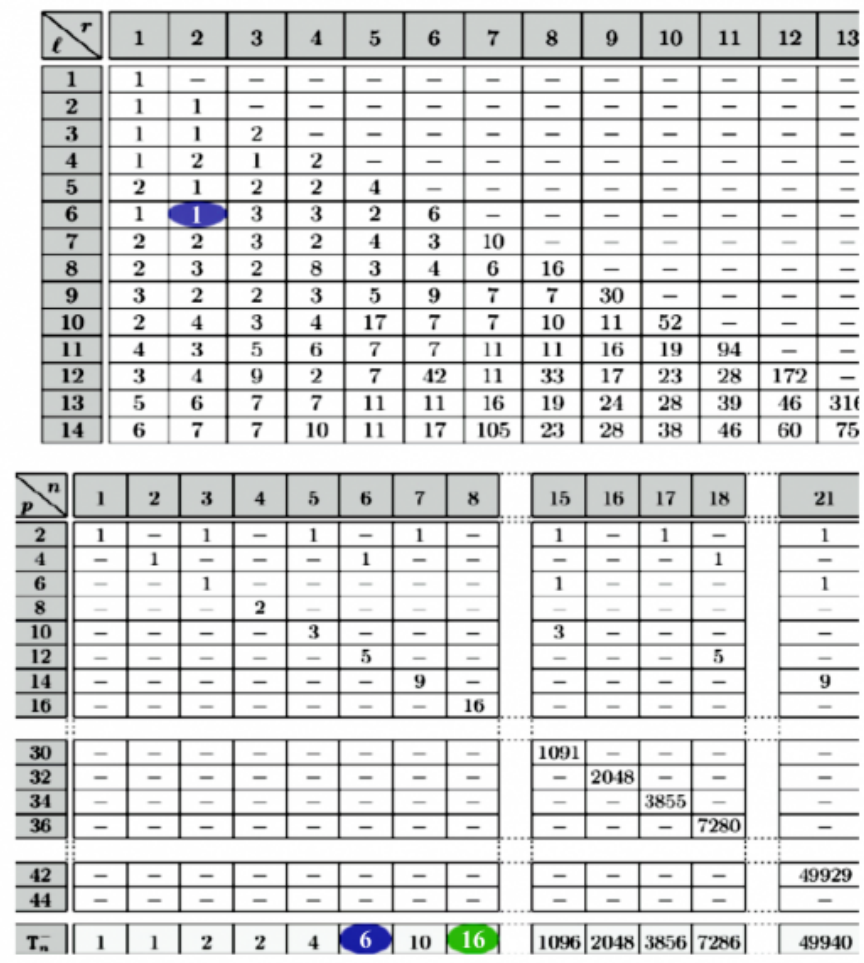


Figure 10.

Réseau de contrôle de la Recombinase RAG, avec 2 circuits négatifs tangents de taille 2 (en rouge) et 6 (en bleu) et 2 gènes de contrôle, inhibiteur Notch/CSL et activateur GATA-1 (en brun) [14].

20 Le réseau de contrôle de la Recombinase RAG possède, comme dans l'architecture générale de la Figure 7, une arborescence source, une arborescente puits, et deux circuits négatifs tangents, de taille respectivement 2 (en rouge) et 6 (en bleu), dans la composante fortement connexe isolée dans la Figure 10.



ℓ \ r	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	–	–	–	–	–	–	–	–	–	–	–	–
2	1	1	–	–	–	–	–	–	–	–	–	–	–
3	1	1	2	–	–	–	–	–	–	–	–	–	–
4	1	2	1	2	–	–	–	–	–	–	–	–	–
5	2	1	2	2	4	–	–	–	–	–	–	–	–
6	1	1	3	3	2	6	–	–	–	–	–	–	–
7	2	2	3	2	4	3	10	–	–	–	–	–	–
8	2	3	2	8	3	4	6	16	–	–	–	–	–
9	3	2	2	3	5	9	7	7	30	–	–	–	–
10	2	4	3	4	17	7	7	10	11	52	–	–	–
11	4	3	5	6	7	7	11	11	16	19	94	–	–
12	3	4	9	2	7	42	11	33	17	23	28	172	–
13	5	6	7	7	11	11	16	19	24	28	39	46	316
14	6	7	7	10	11	17	105	23	28	38	46	60	75

p \ n	1	2	3	4	5	6	7	8	15	16	17	18	21
2	1	–	1	–	1	–	1	–	1	–	1	–	1
4	–	1	–	–	–	1	–	–	–	–	–	1	–
6	–	–	1	–	–	–	–	–	1	–	–	–	1
8	–	–	–	2	–	–	–	–	–	–	–	–	–
10	–	–	–	–	3	–	–	–	3	–	–	–	–
12	–	–	–	–	–	5	–	–	–	–	–	5	–
14	–	–	–	–	–	–	9	–	–	–	–	–	9
16	–	–	–	–	–	–	–	16	–	–	–	–	–
30	–	–	–	–	–	–	–	–	1091	–	–	–	–
32	–	–	–	–	–	–	–	–	–	2048	–	–	–
34	–	–	–	–	–	–	–	–	–	–	3855	–	–
36	–	–	–	–	–	–	–	–	–	–	–	7280	–
42	–	–	–	–	–	–	–	–	–	–	–	–	49929
44	–	–	–	–	–	–	–	–	–	–	–	–	–
T_n	1	1	2	2	4	6	10	16	1096	2048	3856	7286	49940

Figure 11.

Nombre d'attracteurs de deux circuits négatifs tangents de longueurs 6 et 2, mis à jour en mode synchrone (en bleu, en haut) ; en bas, nombre d'attracteurs de période *p* d'un circuit négatif isolé de longueur *n*, mis à jour en mode synchrone, pour les longueurs 6 (en bleu, en bas) et 8 (en vert, en bas).

21 Le nombre d'attracteurs de la composante fortement connexe est donné (en bleu) dans la Figure 11[15] : il est égal à 1 dans le cas des circuits négatifs tangents de taille 6 et 2, et il aurait été de 6 dans le cas de circuits isolés, et de 16 dans le cas d'un unique circuit négatif de taille 6+2=8 (en vert). Cette unicité, dans le cas du réseau de contrôle de la RAG, permet donc d'observer seulement (I) un seul attracteur périodique 1100, constitué des états successifs de la RAG exprimé/exprimé/silencieux/silencieux, si les gènes de contrôle sont silencieux ; (II) un seul état stationnaire de Recombinase RAG, égal à 1 (état exprimé), si le gène de contrôle activateur GATA-1 est exprimé et l'inhibiteur Notch/CSL silencieux ; (III) un seul état stationnaire de la Recombinase RAG, égal à 0 (état silencieux), si GATA-1 et

Notch/CSL sont exprimés. Le « double-pédalage » spécifique du contrôle de la Recombinase RAG, joint aux propriétés combinatoires générales de l'architecture des circuits tangents, permet ainsi de simplifier la fonction[16], tout en assurant sa robustesse[17].

5. Optimisation en morphogenèse

22 Les processus de morphogenèse font intervenir des opérations endogènes complexes, sous le contrôle de réseaux de régulation génétique spécifiques, composés de morphogènes et sous l'influence de champs exogènes physico-chimiques (chimiotactiques, haptotactiques, électro-magnétiques, gravitationnels, lumineux, fluidiques, ...). La diffusion des éléments physico-chimiques internesconstituant le champ externe, jointe aux réactions propres aux processus vivants interne, constitue le moteur de la morphogenèse, qui remplit 3 objectifs, consistant à assurer :
– (I) le développement de l'organe (et donc de l'organisme), jusqu'à l'obtention de sa fonction, phase dans laquelle la prolifération et la diffusion dominent les processus réactionnels cellulaires endogènes spécifiques de cette fonction ;
– (II) le maintien de l'état adulte, dans une homéostasie appelée maintien du « moule intérieur » par Buffon, phase dans laquelle la prolifération est équilibrée par l'apoptose et où l'exérèse accidentelle du tout ou d'une partie d'un organe est intégralement ou partiellement compensée (méro- ou apo-catagenèse[18]) ;
– (III) le contrôle du vieillissement, en évitant les discontinuités des pertes totale ou partielle de la fonction, phase dans laquelle dominent les processus de réparation et de prévention de néogenèses anarchiques.

23 Dans la phase de développement, l'action des champs externes et la diffusion doivent équilibrer la production de structures dissipatives par les réactions cinétiques endogènes. Si u(s,a,t) dénote la densité des cellules d'âge a, situées en un point de l'espace s, à l'instant t, et si B(ε,a,t) dénote la courbe de niveau définie par u(s,a,t)=ε, alors la dynamique morphogénétique peut être définie par une équation générale du type[19] :

Croissance Diffusion Vieillissement Chimiotaxie Haptotaxie Convection Différentiatior

$$\partial u/\partial t = D\Delta u - \beta\partial u/\partial a + \nabla(u\nabla c(u)) + \nabla(u\nabla h(z)) - \nabla(uv) - 2(1\text{-}Q)\beta u,$$

où 1/D est la viscosité, (1-β) le taux de mortalité, v la vitesse du fluide externe, c la fonction de sécrétion chimiotactique, z la densité de la matrice extracellulaire, h la fonction haptotactique et (1-Q) un taux de différentiation. À la limite supérieure M de l'âge cellulaire, on utilise la loi de prolifération, appelée loi de Thom, ou règle de division en fonction du rapport Surface/Volume cellulaire[20] :

$$u(s,1,t) = \int_1^M 2Q\beta C_\varepsilon(s,a,t)u(s,a,t)da,$$

où $C_\varepsilon(s,a,t)$ est la courbure de $B(\varepsilon,a,t)$ en s.

24 Le fait que la division cellulaire soit liée, dans sa régulation épigénétique, à la courbure des frontières où se fait la nutrition/excrétion permet à l'organe d'adapter le flux nutritif au volume à nourrir (anabolisme) et/ou drainer (catabolisme). Cette adaptation de la forme organique, pour accomplir de manière optimale une fonction, peut être considérée comme un des phénomènes centraux de la simplexité morphogénétique. Ce processus est également soumis à un contrôle génétique, et l'exemple de la gastrulation est, sur ce plan, démonstratif.

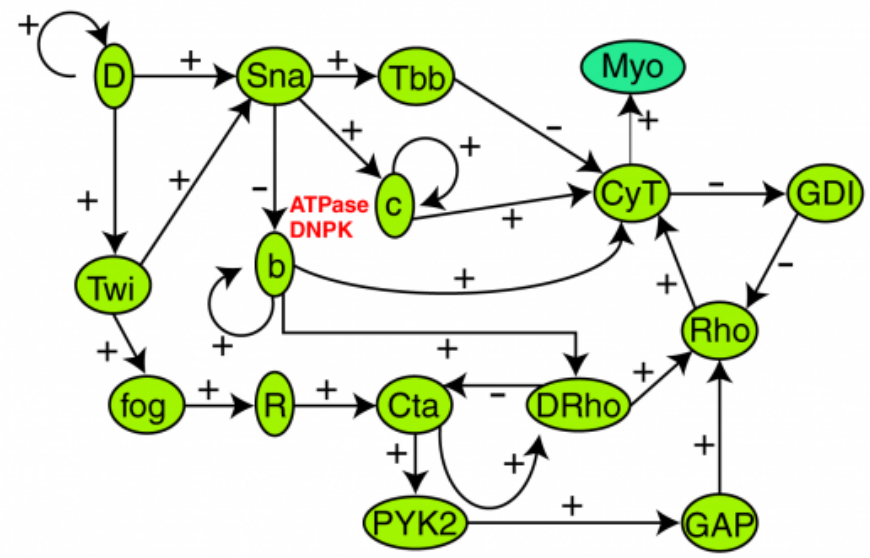


Figure 12.

Réseau de contrôle de la production de Myosine (Myo), avec deux circuits positifs de taille 1, un circuit négatif de taille 2 et un circuit positif de taille 3 dépendant de ceux-ci.

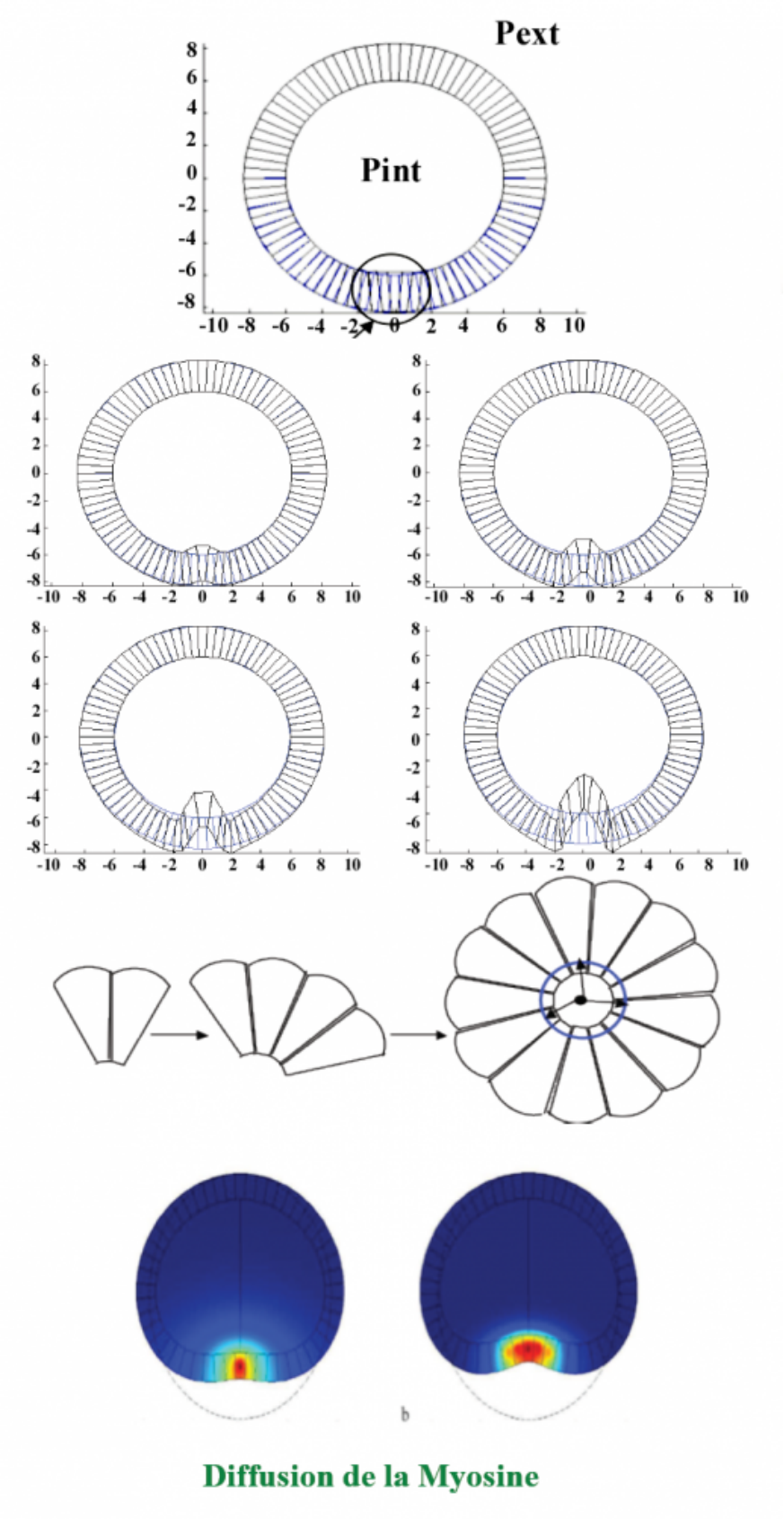


Figure 13.

Diffusion de la myosine (en haut, en bleu) provoquant la contraction cellulaire et la formation du sillon ventral gastrique

(au milieu). Division cellulaire dépendant de la courbure (en bas).

25 Nous suivrons les travaux expérimentaux de référence[21], qui décrivent la première étape de formation du sillon ventral chez la drosophile, préalable à la formation du tube gastrique. La production progressive de myosine, contrôlée par un réseau impliquant de nombreux gènes régulateurs (Figure 12), apparaît initialement comme une fluctuation aléatoire dans une zone centrale du cylindre embryonnaire, diffusant ensuite vers ses extrémités (Figure 13). La myosine provoque une contraction cellulaire, qui, compte tenu des contraintes exercées sur les parois de l'embryon, provoque une invagination, en l'absence de prolifération cellulaire, d'abord aux extrémités, puis ensuite au centre de la surface embryonnaire, phénomène que l'on peut simuler[22].

26 La régulation génétique de la gastrulation s'exerce à travers un circuit négatif de taille 2 (Figure 12), qui a un seul attracteur cycle limite de période 4, 1100 pour les états successifs de DRho, dont dépend un circuit positif de taille 3, qui a un comportement périodique dépendant de ce cycle limite (0110 pour Rho)[23]. La présence des 2 circuits positifs de taille 1 représente l'auto-activation des gènes de l'ATPase et de la DiNucléotide Phosphate Kinase (DNKP), ce qui permet d'avoir en tout 4 attracteurs, correspondant aux 4 états de différenciation des cellules constituant la paroi du futur tube gastrique, issu de la gastrulation. L'expression du réseau de contrôle de la gastrulation, comme celui du réseau précédent de contrôle de la recombinase RAG, présente donc un optimum de simplexification, en termes d'architecture de son graphe d'interaction.

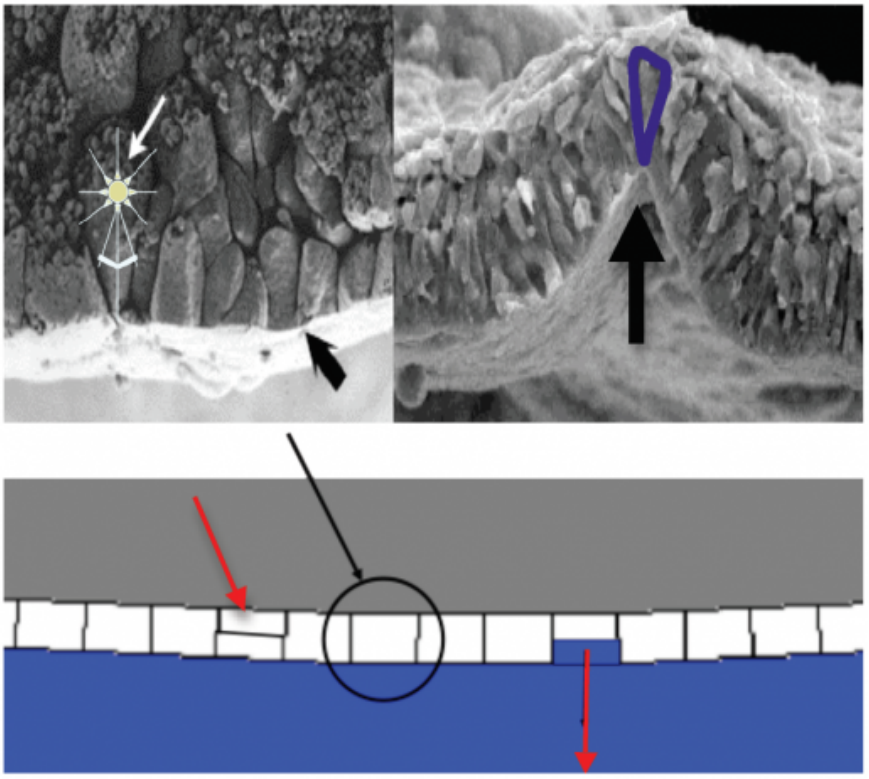


Figure 14a

Cellule en bouteille (flèche blanche) induisant une frontière concave, prélude à la prolifération axiale du sillon (flèches noires) et à la prolifération longitudinale faisant croître le stroma interne (flèches rouges).

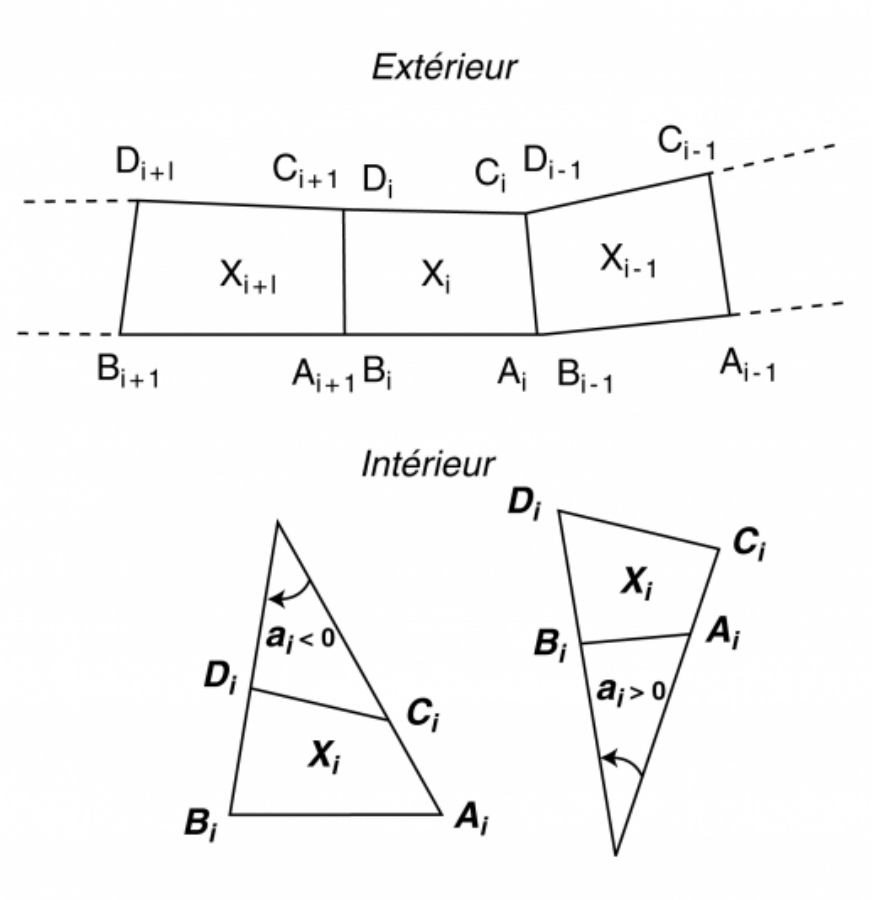


Figure 14b.

Schéma des contours cellulaires $A_iB_iC_iD_i$, dont la division dépend, dans les simulations, du signe de l'angle a_i : la prolifération a lieu, si ce signe est négatif (cellule « concave »).

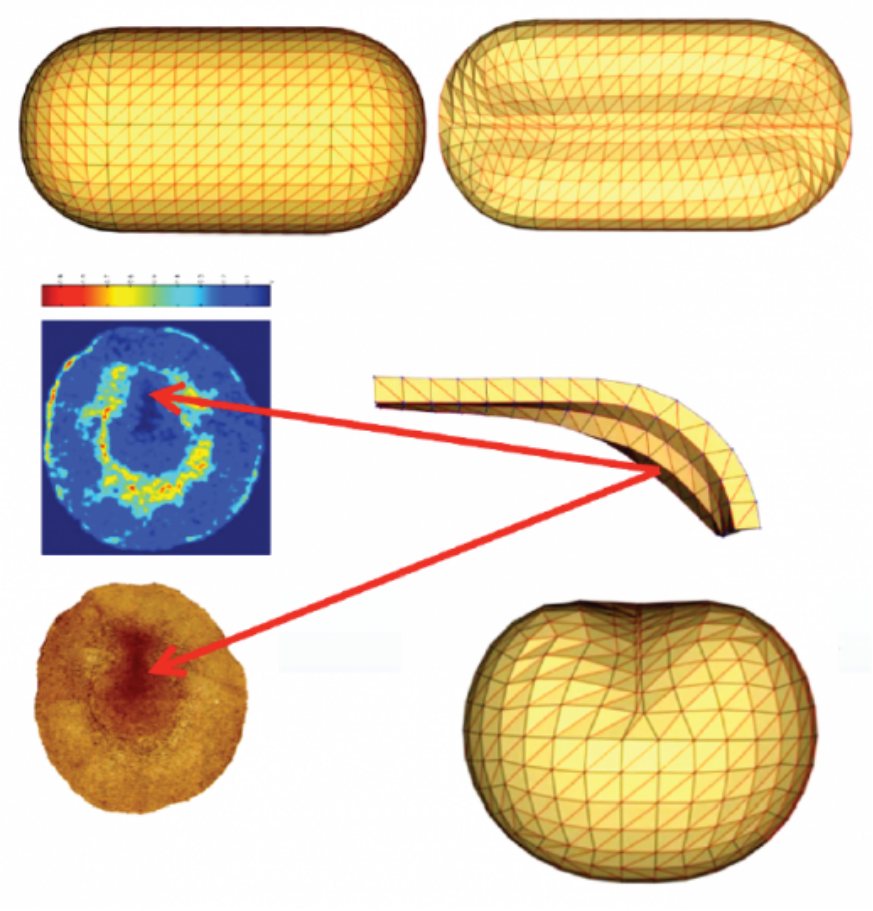


Figure 15

Apparition des premières divisions cellulaires (marquées en vert au BRDU, en haut à gauche) dans la zone la plus concave du sillon ventral gastrique (image microscopique en bas à gauche). À droite : simulations.

27 Dans la phase non proliférante, les simulations des Figures 13 et 15 sont effectuées en respectant les règles de diffusion sur les surfaces régulières, discrétisées dans un « mesh » prenant en compte chaque cellule à travers le cytosquelette de tubuline et du réseau d'acto-myosine. Le bilan des contractions cellulaires fait apparaître, dans les simulations comme dans la réalité, la formation du sillon ventral gastrique aux extrémités du cylindre embryonnaire, l'invagination s'étendant, dans un second temps, à la partie centrale. De plus, les cellules ectodermiques profondes des extrémités du sillon devenant concaves, elles se mettent à proliférer, en majorité par des divisions axiales (Figure 14), comme cela est prouvé expérimentalement par fluorescence du BRDU (5-bromo-2'-deoxyuridine) (Figure 15), provoquant alors la fermeture progressive du sillon, qui devient ainsi le tube gastrique[24].

28 L'exemple de la gastrulation montre que l'application d'un principe variationnel simple, consistant à optimiser la surface nutritive d'une cellule ou d'un organe en fonction du volume à nourrir, suffit à expliquer la formation d'une structure importante, qui représente la deuxième brisure de symétrie du processus de développement de l'embryon : la première est une rupture de symétrie sphérique, due au contact entre la coiffe (acrosome) du spermatozoïde et la membrane pellucide de l'œuf, permettant l'apparition de l'axe céphalo-caudal. La seconde est ici la rupture de la symétrie de révolution du cylindre embryonnaire, et l'apparition du premier tube de l'organisme, le tube gastrique, qui permettra sa nutrition, en changeant l'homotopie sphérique de l'embryon en une homotopie torique, spécialisant ainsi la surface externe pour les échanges gazeux et liquidiens de la thermorégulation (ectoderme) et la surface interne pour les échanges alimentaires (endoderme), les cellules mésodermiques entre ces surfaces étant responsables du maintien de la forme globale de l'organisme. La structure torique, qui n'est pas de révolution, mais appartient à une classe d'homotopie différente de celle du cylindre, sera ensuite brisée de nombreuses fois, en particulier par la seconde structure canalaire de l'organisme, le tube neural, nécessaire à la commande nerveuse centrale, permettant par exemple la locomotion contrôlée par la vision (activité sensorimotrice utilisant dans les deux sens le support de la moelle épinière, située dans ce tube neural), ainsi qu'aux activités réflexes médullaires. La succession de ces brisures de symétrie peut être considérée comme une phase cruciale du développement, relevant de la simplexification de la forme d'un organe, afin d'en optimiser la fonction.

6. Résonance entre réseaux

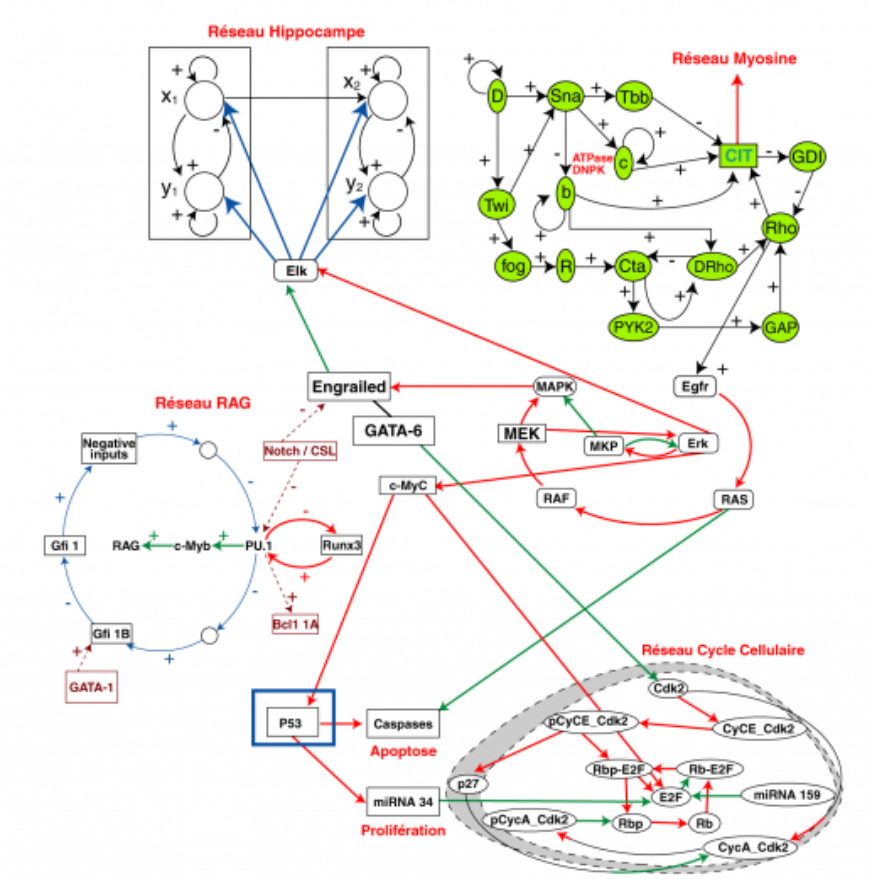


Figure 16.

Coalescence du réseau de contrôle central Engrailed et des réseaux RAG Myosine, Hippocampe et Cycle Cellulaire. Les flèches rouges (resp. vertes) correspondent à des activations (resp. blanc inhibitions).

29 Les réseaux de régulation qui ont été considérés dans cet article sont en fait interconnectés et leur coalescence globale en un réseau supérieur permet d'étudier l'intrication de fonctions primaires participant à une régulation d'ordre supérieur. Un exemple est donné par le contrôle de la taille d'un organe à l'état adulte : le réseau de contrôle du cycle cellulaire est activé par l'une des protéines anti-tumorales ubiquitaires majeures, p53 (Figure 16, rectangle bleu), qui active périodiquement le processus linéaire d'apoptose, de manière à juste compenser le processus exponentiel de mitose, afin de conserver le nombre de cellules d'un tissu[25]. Cette protéine est sous la dépendance du réseau central de régulation Erk-Engrailed, qui participe également, via le gène Elk, au contrôle des canaux potassiques permettant le fonctionnement du circuit négatif hippocampique de la Section 3, nécessaire à l'évocation mnésique. Enfin, Egfr est

activé par Rho[26] et Engrailed est inhibé par Notch chez la souris[27] et la drosophile[28]. Notch inhibe également le réseau de contrôle de la Recombinase RAG (Figure 9).

30 L'existence d'un petit nombre de gènes dans les arbres sources (en général inhibiteurs, que ce soient des gènes véritables ou de petits ARN, eux-mêmes inhibés par des ARN circulaires cytoplasmiques comme ciRS7), responsables du faible nombre final d'attracteurs du réseau global, permet de penser que la simplicité de l'architecture de haut niveau est la condition de la simplexité des processus de contrôle génétique mis en jeu. L'existence d'un contrôle inhibiteur de haut niveau dans le système nerveux central (comme l'inhibition corticale des noyaux de la base) invite à considérer que ce mécanisme est général et sélectionne finalement un faible nombre d'attracteurs, parmi ceux générés par les circuits positifs internes des réseaux (comme conjecturé par R. Thomas[29], puis démontré ensuite théoriquement[30], optimisant ainsi le nombre de fonctions liées à ces attracteurs (comme le proposait déjà M. Delbrück en 1949[31]), en supprimant les attracteurs inutiles et donc leurs fonctions associées superflues.

7. Conclusion

31 En conclusion , nous avons vu successivement le rôle de la simplexité dans l'émergence de nombreuses fonctions cruciales pour la survie d'un organisme :

– fonction cognitive de l'évocation mnésique ;

– contrôle de la diversité des gènes des récepteurs α des lymphocytes T, responsable de la richesse des capacités de réponse immunologique ;

– apparition de phénomènes dynamiques primitifs, comme la gastrulation, étape cruciale de la morphogenèse.

32 Le rôle des circuits à l'intérieur des systèmes complexes (neuronaux ou génétiques) mis en jeu dans cette émergence fonctionnelle, ainsi que dans l'apparition de comportements dynamiques robustes et optimaux des cellules et des organes, soumis à des principes variationnels d'optimisation, permet d'envisager la simplexité comme l'ultime avatar de la complexité, et invite à chercher d'autres procédures simplexifiantes que celles décrites dans cet article, rendant ainsi l'étude de la simplexité duale et naturellement complémentaire de celle de la complexité.

Bibliographie

Bibliographie

L. ABBAS, J. DEMONGEOT & N. GLADE. Synchrony in Reaction-diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves. *Phil. Trans. Royal Soc.* A, 367, 4829-4862 (2009).

L. ALMEIDA & J. DEMONGEOT. Predictive power of “a minima” models in biology. *Acta Biotheoretica*, 60, 3-19 (2012).

M.K. ANDERSON, G. HERNANDEZ-HOYOS, C.J. DIONNE, A.M. ARIAS, D. CHEN & E.V. ROTHENBERG. Definition of Regulatory Network Elements for T Cell Development by Perturbation Analysis with PU.1 and GATA-3. *Developmental Biology*, 246, 103–121 (2002).

J. ARACENA, M. GONZÁLEZ, A. ZUÑIGA, M. MÉNDEZ & V. CAMBIAZO. Regulatory network for cell shape changes during Drosophila ventral furrow formation. *Journal of Theoretical Biology*, 239, 49-62 (2006).

T.P. BAUM, N. PASQUAL, V. HIERLE, F. BELLAHCENE, D. CHAUME, M.P. LEFRANC, E. JOUVIN-MARCHE, P. MARCHE & J. DEMONGEOT. IMGT/GeneInfo : new gamma and delta chains for database V(D)J recombination. *BMC Bioinformatics*, 7, 224-228 (2006).

H. BEN AMOR, J. DEMONGEOT, A. ELENA & S. SENÉ. Structural Sensitivity of Neural and Genetic Networks. *Lecture Notes in Computer Science*, 5317, 973-986 (2008).

H. BEN AMOR, N. GLADE, C. LOBOS & J. DEMONGEOT. The isochronal fibration: characterization and implication in biology. *Acta Biotheoretica*, 58, 121-142 (2010).

H. 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*. Thèse, Université J. Fourier (2012).

H. BOSCH, R. MILANESE, A. LABBI & J. DEMONGEOT. Rate and Temporal Coding with a Neural Oscillator. *Lecture Notes in Computer Science*, 3697, 1021-1026 (1998).

F. CARAGUEL, M. TAYYAB, F. GIROUD & J. DEMONGEOT. Evolution of the genetic regulatory networks: the example of the cell cycle control network. From gastrulation modelling to apocatagenesis. In: IEEE AINA' 10 & BLSMC' 10, *IEEE Proceedings*, Piscataway, 767-774 (2010).

O. CINQUIN & J. DEMONGEOT. Positive and negative feedback : striking a balance between necessary antagonists. *J. Theor. Biol.*, 216, 229-241 (2002).

C. CUI, X. YANG, M. CHUAI, J.A. GLAZIER & C.J. WEIJER. Analysis of tissue flow patterns during primitive streak formation in the chick embryo. *Developmental Biology*, 84, 37-47 (2005).

M. DELBRÜCK. *Unités biologiques douées de continuité génétique*. Éditions du CNRS, Paris, pp. 33-34 (1949).

J. DEMONGEOT & J.P. FRANÇOISE. Approximation for limit cycles and their isochrons. *Comptes Rendus Biologies*, 329, 967-970 (2006).

J. DEMONGEOT, M. THELLIER & R. THOMAS. Storage and recall of environmental signals in a plant: modelling by use of a differential (continuous) formulation. *Comptes Rendus Biologies*, 329, 971-978 (2006).

J. DEMONGEOT, N. GLADE & L. FOREST. Liénard systems and potential-Hamiltonian decomposition. I-III. *Comptes Rendus Mathématique*, 344, 121-126, 191-194, 253-258 (2007).

J. DEMONGEOT, A. ELENA & S. SENÉ. Robustness in neural and genetic networks. *Acta Biotheoretica*, 56, 27-49 (2008).

J. DEMONGEOT, C. JEZEQUEL & S. SENÉ. Asymptotic behavior and phase transition in regulatory networks. I-II. Theoretical results. *Neural Networks*, 21, 962-979 (2008).

J. DEMONGEOT, H. BEN AMOR, P. GILLOIS, M. NOUAL & S. SENÉ. Robustness of regulatory networks. A Generic Approach with Applications at Different Levels: Physiologic, Metabolic and Genetic. *Int. J. Molecular Sciences*, 10, 4437-4473 (2009).

J. DEMONGEOT, E. GOLES, M. MORVAN, M. NOUAL & S. SENÉ. Attraction Basins as Gauges of Environmental Robustness in

Biological Complex Systems. *PloS ONE*, 5, e11793 (2010).

J. DEMONGEOT, A. ELENA, M. Noual, S. SENÉ & F. THUDEROZ. "Immunetworks", attractors and intersecting circuits. *J. Theor. Biology*, 280, 19-33 (2011).

J. DEMONGEOT & J. WAKU. Robustness in biological regulatory networks. I-IV *Comptes Rendus Mathématique*, 350, 221-228, 289-298 (2012).

J. DEMONGEOT, M. NOUAL & S. SENÉ. Combinatorics of Boolean automata circuits dynamics. *Discrete Applied Mathematics*, 160, 398-415 (2012).

J. DEMONGEOT, J. GAUDART, A. LONTOS, E. PROMAYON, J. MINTSA & M. RACHDI. Least diffusion zones in morphogenesis and epidemiology. *Int. J. Bifurcation and Chaos*, 22, 50028 (2012).

W. DENG, J.B. AIMONE & F.H. GAGE. New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory*?* *Nature Reviews Neuroscience*, 11, 339-350 (2010).

S.S. DESHMUKH & J.J. KNIERIM. Hippocampus, Overview. *Cognitive Science*, 3, 231-251 (2012).

L. FOREST, J. SAN MARTIN, F. PADILLA, F. CHASSAT, F. GIROUD & J. DEMONGEOT. Morphogenetic processes: application to cambial growth dynamics. *Acta Biotheoretica*, 52, 415-438 (2004).

L. FOREST & J. DEMONGEOT. Cellular modelling of secondary radial growth in conifer trees: application to Pinus radiata. *Bull. Math. Biol.*, 68, 753-784 (2006).

L. FOREST, S. MARTINEZ, F. PADILLA, J. DEMONGEOT & J. SAN MARTIN. Modelling of auxin transport affected by gravity and differential radial growth. *J. Theor. Biol.*, 241, 241-251 (2006).

L. FOREST, N. GLADE & J. DEMONGEOT. Liénard systems and potential-Hamiltonian decomposition. Applications. *C. R. Acad. Sci. Biologies*, 330, 97-106 (2007).

L. FOREST & J. DEMONGEOT. A general formalism for tissue morphogenesis based on cellular dynamics and control system interactions. *Acta Biotheoretica*, 56, 51-74 (2008).

C. GEORGESCU, W.J.R. LONGABAUGH, D.D. SCRIPTURE-ADAMS, E.S. DAVID-FUNG, M.A. YUI, M.A. ZARNEGAR, H. BOLOURI & E.V. ROTHENBERG. A gene regulatory network armature for T lymphocyte specification. *PNAS*, 105, 20100–20105 (2008).

L. GONZÁLEZ-MARISCAL, R. TAPIA & D. CHAMORRO. Crosstalk of tight junction components with signaling pathways. *BBA*, 1778, 729-756 (2008).

J.W. LACY, M.A. YASSA, S.M. STARK & C.E.L. STARK. Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning Memory*, 18, 15-18 (2011).

M. LEPTIN. Gastrulation in Drosophila: the logic and the cellular mechanisms. *The EMBO Journal*, 18, 3187–3192 (1999).

J. LERMA. Net(o) excitement for kainate receptors. *Nature Neuroscience*, 14, 808-810 (2011).

E.Y. MANTROVA & T. HSU. Down-regulation of transcription factor CF2 by DrosophilaRas/MAP kinase signaling in oogenesis: cytoplasmic retention and degradation. *Genes & Dev.*, 12, 1166-1175 (1998).

A.C. MARTIN, M. KASCHUBE & E.F. WIESCHAUS. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature*, 457, 495-499 (2009).

M.P. MATISE, A.L. JOYNER. Expression patterns of developmental control genes in normal and Engrailed-1 mutant mouse spinal cord reveal early diversity in developing interneurons. *J. Neurosci.*, 17, 7805-7816 (1997).

O. NÉROT. *Mémorisation par forçage neuronal des dynamiques chaotiques dans les modèles connexionnistes récurrents*. Thèse, Institut National Polytechnique, Grenoble (1996)

R.M. SANCHEZ, W. DAI, R.E. LEVADA, J.J. LIPPMAN & F.E. JENSEN. AMPA/Kainate Receptor-Mediated Downregulation of GABAergic Synaptic Transmission by Calcineurin after Seizures in the Developing Rat Brain. *The Journal of Neuroscience*, 25, 3442–3451 (2005).

A.M. SOKAC & E. WIESCHAUS, Zygotically controlled F-actin establishes cortical compartments to stabilize furrows during Drosophila cellularization, *J Cell Sci*., 121, 1815-1824 (2008a).

A.M. SOKAC & E. WIESCHAUS. Local actin-dependent endocytosis is zygotically controlled to initiate Drosophila cellularization, *Dev Cell*., 14, 775-786 (2008b).

C. SOULÉ. Graphic requirements for multistationarity. *Complexus*, 1, 123-133 (2003).

A.T. TANG, W.B. CAMPBELL & K. NITHIPATIKOM. ROCK1 feedback regulation of the upstream small GTPase RhoA. *Cell Signal*., 24, 1375-1380 (2012).

S. TAKASHIMA, H. YOSHIMORI, N. YAMASAKI, K. MATSUNO & R. Murakami. Cell-fate choice and boundary formation by combined action of Notch and Engrailed in the Drosophila hindgut. *Development, Genes and Evolution*, 212, 534-541 (2002).

M. TAYYAB, A. LONTOS, E. PROMAYON & J. DEMONGEOT. Modelling and image processing of constriction and proliferation in the gastrulation process of Drosophila melanogaster. In: IEEE AINA' 11 & BLSMC' 11, *IEEE Proceedings*, Piscataway, 473-477 (2011).

R. THOMAS. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer Series in Synergetics*, 9, 180-193 (1981).

F. THUDEROZ, M.A. SIMONET, O. HANSEN, A. DARIZ, T.P. BAUM, V. HIERLE, J. DEMONGEOT, P.N. MARCHE & E. JOUVIN-MARCHE. From the TCRAD rearrangement quantification to the computational simulation of the locus behavior. *PloS Comp. Biol*., 6, e1000682 (2010).

O.S. VINOGRADOVA. Hippocampus as Comparator: Role of the Two Input and Two Output Systems of the Hippocampus in Selection and Registration of Information. *Hippocampus*,11, 578–598 (2001).

H.R. WILSON & J.D. COWAN. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*., 12, 1–24 (1972).

M.F. YECKEL & T.W. BERGER. Feedforward excitation of the hippocampus by afferents from the entorhinal cortex: Redefinition of the role of the trisynaptic pathway. *Proc. Natl. Acad. Sci. USA*, 87, 5832-5836 (1990).

Notes

1. Lacy *et al.*, 2011 ; Deshmukh & Knierim, 2012 ; Deng *et al.*, 2010 ; Lerma, 2011 ; Sanchez *et al.*, 2005.

2. Sanchez *et al.*, 2005.

3. Lerma, 2011.

4. Deng *et al.*, 2010.

5. Wilson & Cowan, 1972.

6. Nous nous appuierons, dans ce qui suit, sur les publications suivantes : Demongeot & Françoise, 2006 ; Demongeot, Thellier & Thomas, 2006 ; Forest, Glade & Demongeot, 2007 ; Demongeot, Glade & Forest, 2007 ; Ben Amor, 2012 ; Wilson & Cowan, 1972 ; Yeckel & Berger, 1990 ; Vinogradova, 2001 ; Nérot, 1996 ; Bosch *et al.*, 1998.

7. Forest, Glade & Demongeot, 2007 ; Demongeot, Glade & Forest, 2007.

8. Yeckel & Berger, 1990 ; Vinogradova, 2001.

9. Vinogradova, 2001.

10. Nérot, 1996 ; Bosch *et al.*, 1998.

11. Demongeot, Elena & Sené, 2008 ; Demongeot, Jezequel & Sené, 2008 ; Ben Amor *et al.*, 2008 ; Demongeot, Ben Amor, Gillois, Noual & Sené, 2009 ; Ben Amor *et al.*, 2010 ; Demongeot, Noual & Sené, 2012 ; Demongeot & Waku, 2012 ; Baum *et al.*, 2006 ; Thuderoz *et al.*, 2010 ; Demongeot, Goles, Morvan, Noual & Sené, 2010.

12. Demongeot, Elena, Noual, Sené & Thuderoz, 2011.

13. Baum *et al.*, 2006 ; Thuderoz *et al.*, 2010.

14. D'après Georgescu *et al.*, 2008 ; et Anderson *et al.*, 2002.

15. D'après Demongeot, Noual & Sené, 2012.

16. Demongeot, Noual & Sené, 2012.

17. Demongeot & Waku, 2012.

18. Caraguel *et al.*, 2010.

19. Forest *et al.*, 2004 ; Forest & Demongeot, 2008.

20. Pour une utilisation dans la croissance tronculaire des arbres, cf. Forest & Demongeot, 2006 ; Forest *et al.*, 2006 ; Abbas *et al.*, 2009.

21. Leptin, 1999 ; Aracena *et al.*, 2006; González-Mariscal, 2008 ; Sokac & Wieschaus, 2008a, 2008b ; Martin *et al.*, 2009 ; Tang *et al.*, 2012.

22. Tayyab *et al.*, 2011 ; Demongeot, Gaudart, Lontos, Promayon, Mintsa & Rachdi, 2012 ; Almeida & Demongeot, 2012 ; Cui *et al.*, 2005.

23. Demongeot, Noual & Sené, 2012.

24. Cui *et al.*, 2005.

25. Almeida & Demongeot, 2012.

26. Mantrova & Hsu, 1998.

27. Matise & Joyner, 1997.

28. Takashima *et al.*, 2002.

29. Thomas, 1981.

30. Cinquin & Demongeot, 2002 ; Soulé, 2003.

31. Delbrück, 1949.

Auteurs

Jacques Demongeot

Université J. Fourier, AGIM CNRS 3405 & IUF

Hedi Ben Amor

Hana Hazgui

Athanasios Lontos

Référence électronique du chapitre

DEMONGEOT, Jacques ; et al. *La simplexité, ultime avatar de la complexité ?* In : *Complexité-Simplexité* [en ligne]. Paris : Collège de France, 2014 (généré le 06 décembre 2014). Disponible sur Internet : <http://books.openedition.org/cdf/3393>. ISBN : 9782722603301.

Référence électronique du livre

BERTHOZ, Alain (dir.) ; PETIT, Jean-Luc (dir.). *Complexité-Simplexité*. Nouvelle édition [en ligne]. Paris : Collège de France, 2014 (généré le 06 décembre 2014). Disponible sur Internet : <http://books.openedition.org/cdf/3339>. ISBN : 9782722603301.
Compatible avec Zotero

Acta Biotheor (2014) 62:243–284
DOI 10.1007/s10441-014-9229-5

REGULAR ARTICLE

Stability, Complexity and Robustness in Population Dynamics

J. Demongeot · H. Hazgui · H. Ben Amor · J. Waku

Received: 15 December 2013 / Accepted: 17 June 2014 / Published online: 9 August 2014

Abstract The problem of stability in population dynamics concerns many domains of application in demography, biology, mechanics and mathematics. The problem is highly generic and independent of the population considered (human, animals, molecules,…). We give in this paper some examples of population dynamics concerning nucleic acids interacting through direct nucleic binding with small or cyclic RNAs acting on mRNAs or tRNAs as translation factors or through protein complexes expressed by genes and linked to DNA as transcription factors. The networks made of these interactions between nucleic acids (considered respectively as edges and nodes of their interaction graph) are complex, but exhibit simple emergent asymptotic behaviours, when time tends to infinity, called attractors. We show that the quantity called attractor entropy plays a crucial role in the study of the stability and robustness of such genetic networks.

Keywords Structural stability · Liapunov stability · Asymptotic stability · Attractors · Genetic networks · Network robustness

J. Demongeot (✉) · H. Hazgui · H. Ben Amor
AGIM, FRE CNRS 3405, Faculty of Medicine of Grenoble, University J. Fourier,
38700 La Tronche, France
e-mail: Jacques.Demongeot@yahoo.fr; Jacques.Demongeot@agim.eu

H. Hazgui
e-mail: Hana.Hazgui@agim.eu

H. Ben Amor
e-mail: hedibenamor@gmail.com

J. Waku
LIRIMA-UMMISCO, Faculté des Sciences, Université de Yaoundé, BP 812, Yaoundé, Cameroun
e-mail: Jules.Waku@gmail.com

1 Introduction

Since the beginning of the sixties, the problem of stability of the dynamics of a population has been at the centre of the preoccupations of the demographists, biologists, mechanical engineers and mathematicians. The problem of stability and the related question of robustness of a dynamical system is generic and independent of the population considered (human, animals, molecules,…). In this paper, we will give in Sect. 2 the main definitions useful for understand the notions of stability and robustness in population dynamics, and in Sect. 3 some examples of population dynamics concerning nucleic acids interacting through direct nucleic binding with small or cyclic RNAs (alone or linked to RNA-binding peptides) acting on mRNAs or tRNAs as translation factors, and through protein complexes expressed by genes and linked to DNA as transcription factors. The networks made of these interactions between nucleic acids considered respectively as edges and nodes of their interaction graph are complex, but exhibit simple emergent asymptotic behaviours, when time tends to infinity, called attractor dynamics, or simply attractors. We show in Sect. 4 that the quantity called the attractor entropy can play a crucial role in the study of the stability and robustness of such genetic networks and we use this quantity in Sect. 5, in order to study the asymptotic stability and the robustness of the examples given in Sect. 3.

2 Different Definitions of the Notions of Attractor and Stability

Different acceptations of the notions of attractor and stability have been proposed since 30 years, we recall now.

2.1 Definitions of the Notion of Attractor (Cosnard and Demongeot 1985a)

A trajectory x(a,t) in a state space $E \subset \mathbb{R}^n$ is the set of all states observed as time t goes in time space T to infinity from the initial value 0, corresponding to the state $x(a,0) = a$. The set of states visited when t tends to infinity is called the limit set of the trajectory starting in a, and is denoted $L(a) = \{y \in E;\ \forall \varepsilon > 0,\ \forall t \in T,\ \exists s(\varepsilon, t) > t / d(x(a, s(\varepsilon, t)), y) < \varepsilon\}$. If the initial states lie in a set A, then L(A) is the union of all limit sets L(a), for a belonging to $A : L(A) = \cup_{a \in A} L(a)$. Conversely, B(A), called the attraction basin of A, is the set of all initial conditions outside A, whose limit set L(a) is included in A:$B(A) = \{y \in E \backslash A;\ L(y) \subset A)\}$ (Fig. 1).

The definition of the "shadow connectivity" comes from the notion of "shadow trajectory", defined first by (Bowen 1978). The definition of attractor as a shadow connected set A=L(B(A)) has a character of generality (available for all known dynamical systems, in agreement with the common sense) and genericity (meaningful for both discrete or continuous time sets T independently of the time schedule, if the dynamical system is autonomous with respect to time t), contrary to the other definitions, less general and more specific (Cosnard and Demongeot 1985a, b).

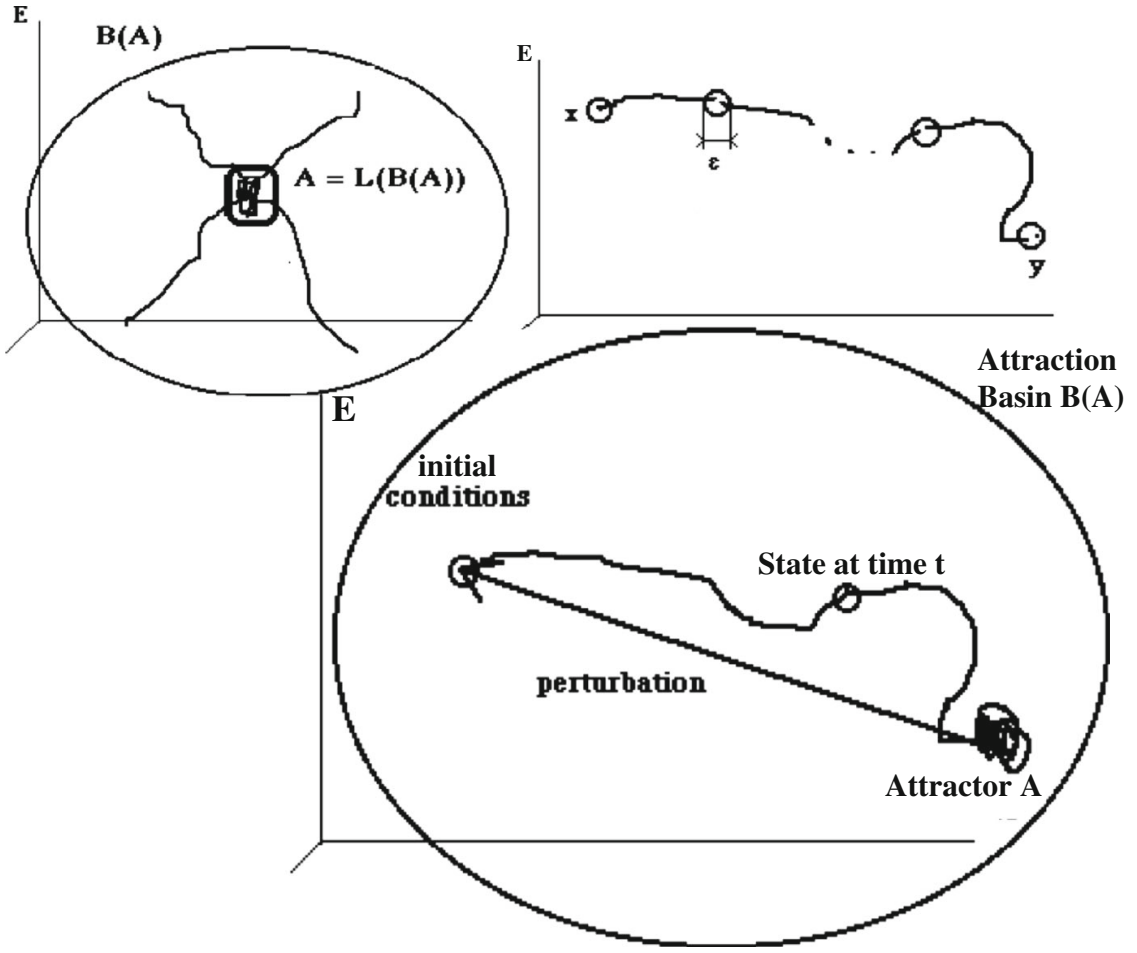


Fig. 1 *Top left* Attractor A as invariant set for the composed operator LoB. *Top right* shadow trajectory between x and y in E. *Bottom* the attraction basin B(A) as the set of states located outside A, reached from A after perturbation at time 0 and going back to A when t tends to $+\infty$

2.2 Definition of Liapunov (or Trajectorial) Stability and Asymptotic Stability (Liapunov 1992)

There exist several definitions of the stability of a dynamical system and we will give hereafter the most useful:

(1) a trajectory x(a,t) is called Liapunov stable, if no perturbation at any time t denoted $b = x(a,t) + e$ should be amplified: $\forall s > t,\ d(x(a,s), x(b, s-t)) \leq e$,

(2) a trajectory x(a,t) is called asymptotically stable, if any perturbation at any time t is asymptotically damped: if $b = x(a,t) + e$ denotes the perturbed state, $\lim_{s\to\infty} d(x(a,s), x(b, s-t)) = 0$.

2.3 Definition of the Structural Stability with Respect to a Parameter (Thom 1975)

A dynamical system whose trajectories $x_\mu(a,t)$ depend on a parameter μ is called structurally stable with respect to the parameter μ, if no perturbation of μ at any time

can provoke a change in number or nature of its attractors (fixed attractor called steady state, or periodic attractor called limit-cycle). The parameter μ may parametrize the state transition rule of the system or its architectural characteristics (number of elements, intensity of interactions).

2.4 Definition of the Structural Stability with Respect to Updating Modes (Aracena et al. 2013)

A dynamical system is called structurally stable with respect to updating modes, if no disturbance in the updating schedule, e.g., by passing from the sequential mode (in which nodes of the network are updated by the state transition rule one after the other in a given order) to the parallel one (in which all the nodes of the network are updated by the state transition rule at the same time) can change the number or nature of its attractors.

2.5 Definition of the Resistance to Boundary Perturbations or Resilience (Meyn 2008)

A dynamical system having a frontier separating them from its environment is resistant to boundary perturbations, if no perturbation in state or architecture of the environmental elements can provoke a change in number or nature of its attractors.

2.6 Definition of the Robustness (Waddington 1942)

A dynamical system whose trajectories $x_p(a,t)$ depend on a parameter p is said to be robust, if it is boundary resistant and structurally stable with respect to any parameter or updating schedule perturbation. We will give in the next Section some examples of robust and non robust biological networks, at different levels, genetic, metabolic and physiologic. These networks can be decomposed following their dynamical typology and we can distinguish between 4 categories of dynamics, whose definition will be given hereafter:

- Potential (or gradient, or purely dissipative)
- Hamiltonian (or conservative)
- Mixed potential-Hamiltonian (MPH)
- MPH with principal potential part.

2.7 Definition of a Potential (or Gradient) Dynamics (Schey 1992)

A dynamical system has a potential dynamics if the velocity along its trajectories is equal to the gradient of a scalar potential P defined on the state space E. If the system is governed by a differential equation defining its state transition rule, we have: $dx(a,t)/dt = -\nabla P = -\text{grad}P = -\partial P/\partial x$, where $-\partial P/\partial x$ is the vector $(-\partial P/\partial x_1, \ldots, -\partial P/\partial x_n)$ and the state x is a vector of dimension n: $x = (x_1, \ldots, x_n)$. The system is called dissipative, because the potential P decreases along trajectories until

attractors which are located on the minima of P. There exists a discrete version of this definition (Demongeot et al. 2006b).

2.8 Definition of a Hamiltonian dynamics (Audin and Babbitt 2008)

A dynamical system has a Hamiltonian dynamics if the velocity along its trajectories is tangent to the contour lines projected on E from the surface representative of an energy function H defined on E: $dx_\mu(a,t)/dt = \text{tangH}$. If the dimension of the system is 2, then tangH is equal to $(\partial H/\partial x_2, -\partial H/\partial x_1)$. The system is said conservative, because the energy function H is constant along a trajectory. There exists a discrete version of this definition (Demongeot et al. 2013e).

2.9 Definition of a Mixed Potential–Hamiltonian Dynamics (Demongeot et al. 2007)

A dynamical system has a mixed potential-Hamiltonian dynamics if the velocity along its trajectories can be decomposed into two parts, one potential and one Hamiltonian: $dx(a,t)/dt = -\text{gradP} + \text{tangH}$. If the set of minima of P is a contour line of the surface H on E, then its shadow connected components are attractors of the system.

2.10 Definition of a Principal Potential Part Dynamics (Krasnoselski 1968)

A mixed potential–Hamiltonian system has a principal potential part dynamics, if the ratio $\|\text{tangH}\|/\|\text{gradP}\|$ between the norms of the potential part and the Hamiltonian one, tends to 0, when $\|x\|$ tends to infinity.

3 Some Classical Results About Stability in Population Dynamics

3.1 Random Boolean Automata Networks and Markov Chains

3.1.1 Random Boolean Automata Networks. Definitions

During the eighties, numerous papers have been done about both deterministic and random Boolean automata networks, notably about the threshold Hopfield like automata networks. The application domain was initially neural networks, but progressively the genetic regulatory networks became the first topic using this mathematical tool, mainly under the appellation of genetic threshold Boolean random regulatory networks (getBrens). The random Boolean automata network (Ω, M) is defined by the state space $\Omega = \{0,1\}^m$ of its m Boolean automata and by a transition probability Markovian matrix M, defined for each pair of states (x, y) in Ω, by $M_{x,y}(t)$ equal to the probability to pass from the state x to the state y at time t. We suppose in the following that $M_{x,y}$ is independent of t, i.e., remains constant at each iteration of the network. We will use for defining M the classical framework of the threshold Hopfield like automata networks, which gives for each automaton $i (i \in \{1,\ldots,m\})$ the transition rule $P^{\beta}_{i,\gamma}$ giving the probability to get the state

$x_i(t+1) = \gamma$ in $\{0,1\}$ for the automaton i at time $t+1$, from the configuration $x(t) = \beta$ in Ω:

$$\forall \gamma \in \{0,1\}, \beta \in \Omega, P_{i,\gamma}^{\beta}(\{x_i(t+1) = \gamma | x(t) = \beta\}) = \exp\left[\gamma\left(\sum_{j\in N_i} w_{ij}\beta_j - \theta_i\right)\Big/ T\right]\Big/\left(1 + \exp\left[\left(\sum_{j\in N_i} w_{ij}\beta_j - \theta_i\right)\Big/ T\right]\right)$$

The matrix $W = \left(w_{ij}\right)_{i,j=1,\ldots,m}$ defines the interaction weights w_{ij} representing the influence the automaton j in the neighborhood N_i of i has on the automaton i, θ_i is the activation threshold of the automaton i and T (called temperature with reference to the models of statistical mechanics and neural networks) is a stochasticity parameter. When T tends to infinity, $P_{i,\gamma}^{\beta}$ tends to 1/2 and the invariant Markovian measure, notion which generalizes that of deterministic attractor (Goldstein 1981; Goldstein and Penrose 1981; Demongeot and Jacob 1989; Tuljapurkar 1993; Jensen et al. 2007; Elena and Demongeot 2008; Demongeot et al. 2012) tends to the uniform probability distribution on Ω. When T tends to 0, the behaviour of the Boolean random network becomes that of a deterministic Boolean network.

3.1.2 Relationships with Deterministic Boolean Automata Networks

The relation between random and deterministic Boolean automata is studied on a simple example in which $m = 4k^2$ automata are located on the 2d-grid $\{-k, k-1\}^2$, where k is an integer and $\Omega = \{0,1\}^m$. In Demongeot et al. (2008b) and Demongeot and Sené (2008), is described the behaviour of a Hopfield random Boolean automata network, notably in the case where interaction weights w_{ij} are isotropic and translation invariant on the 2d-grid, as in Ising model.

On Fig. 2, is drawn in blue the phase transition line, on which both simulations and theory proved the existence of multiple asymptotic behaviours, corresponding to distinct invariant measures defined by initial and boundary conditions, e.g., the coexistence of quincuncial configurations, when $T = +\infty$, $u_0 > 0$ and $u_1 < 0$ (cf. Fig. 2 bottom right).

The inhibitory ($u_1 < 0$) and activatory ($u_1 > 0$) cases of the 2d-Hopfield model correspond respectively to the repulsive and attractive cases of the 2d-Ising model. A simple way to obtain the correspondence between random and deterministic Boolean automata networks consists in forcing the interaction parameters $u_0 = w_{ii}/T$ and $u_1 = w_{ij}/T$ to tend to infinity, by pushing the stochasticity parameter T to 0. In these circumstances, the stochastic transition rule tends to the deterministic weighted threshold majority rule and all the results obtained by counting attractors provided by the couples of tangent circuits of this deterministic Boolean automata network (Demongeot et al. 2012e) are totally in agreement with the existence of a unique or multiple invariant measures in the random case: we see on Fig. 2 that there is no contradiction between theoretical calculations of the number of attractors of a couple of tangent circuits (Demongeot et al. 2012e) and the occurrence of simulated or theoretically predicted phase transition on the grid (Demongeot and Sené 2008; Demongeot et al. 2008b).

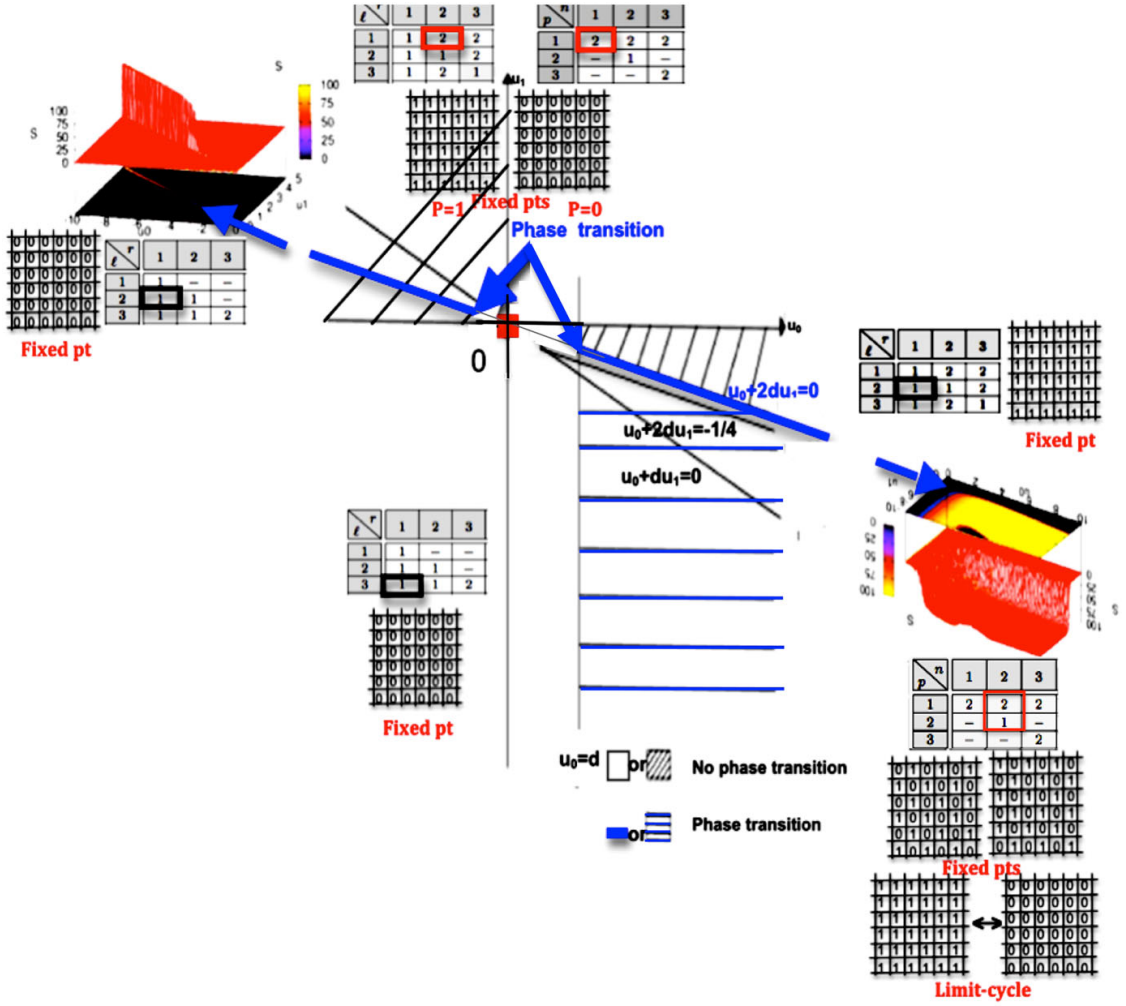


Fig. 2 Parametric conditions of phase transition (i.e., existence of multiple invariant measures) for the 2d-Hopfield model in the isotropic, translation invariant inhibitory ($u_1 < 0$) and activatory ($u_1 > 0$) cases [after Demongeot and Sené (2008) and Demongeot et al. (2008b)]. The deterministic attractors obtained for infinite values of the interaction parameters $u_0 = w_{ii}/T$ and $u_1 = w_{ij}/T$ are given on a $m = 6 \times 6$ 2d-grid, and, for the stochastic case, the representation (*in red*) of the difference S between the probabilities to have 1 or 0 at the central point of the grid (depending on the boundary conditions and parametric values) has been calculated for a 2d-grid of size $m = 131 \times 131$ (Demongeot and Sené, 2008) on the ($u_0 0 u_1$) parameter plane. The origin (*in red*) of the parametric plane ($u_0 0 u_1$) corresponds to a temperature tending to infinity and the invariant measure in this case is unique and equal to the uniform measure on $\Omega = \{0,1\}^m$. The blue line and the horizontally hatched zone correspond to the existence of phase transition. Tables [after Demongeot et al. (2012)] give the attractor numbers for couples of tangent circuits identified on the grid: n denotes the *gcd* of sizes of positive circuits and p the period of these attractors, when $u_0 + 2du_1 > 0$, u_1 and $u_0 > 0$ (where P=0 corresponds to an empty attraction basin for the fixed configuration) or, $u_0 + 2du_1 < 0$, $u_0 > 0$; r and l denote respectively the size of positive and negative circuits, when $u_0 u_1 < 0$, $u_0 + 2du_1 > 0$, and of negative circuits, when $u_0 < 0$, $u_0 + 2du_1 < 0$. (Color figure online)

3.2 Complexity and Stability in Markov Chains

In a Markov chain with transition matrix M, the complexity can be measured by the evolutionary entropy (Demetrius et al. 2004) equal to the Kolmogorov-Sinaï entropy of the chain, defined by:

$$\mathrm{E} = -\sum_{x,y\in\Omega} \mu_x M_{xy} \mathrm{Log} M_{xy},$$

where μ denotes the equilibrium invariant measure of M (defined by its eigenvector corresponding to the dominant eigenvalue 1). The concepts of stability and robustness pertain to the capacity of a network to maintain its functionality in case of random perturbations affecting respectively its dynamics (e.g., a change of initial condition) and its internal organization (e.g., a change of parameter value). The stability rate for the Euclidian norm between the initial and the perturbed invariant measures can be quantified by $-\mathrm{Log}\lambda_2$, where λ_2 is the subdominant eigenvalue of M. We give now Propositions about the estimation of E in the Markovian case, where we call attractors the final classes of the chain. We will denote in the following by $\mathrm{E}_\mu = -\Sigma_{x\in\Omega}\mu_x \mathrm{Log}\mu_x$ the entropy of the invariant measure μ, by $P_{x,y} = \mu_x M_{xy}$ the joined probability induced by the conditional probability M_{xy} and the invariant measure μ, and by $\mu_x \otimes \mu_y$ the Cartesian product of the invariant probabilities μ_x and μ_y. We introduce now the Kullback–Leibler divergence $\mathrm{d_{KL}}$ defined between two measures μ and ν by:

$$d_{\mathrm{KL}}(\mu, \nu) = -\sum_x \mu_x \mathrm{Log}(\nu_x/\mu_x)$$

Proposition 1 *We can calculate:* $E = E_\mu - d_{KL}(P_{x,y}, \mu_x \otimes \mu_y)$

Proof We have indeed:

$$\begin{aligned}\mathrm{E} &= -\sum_{x,y\in\Omega} \mu_x M_{xy} \mathrm{Log} M_{xy} = -\sum_{x,y\in\Omega} \mu_x\mu_y \left[P_{x,y}/(\mu_x\mu_y)\right]\left[\mathrm{Log}(P_{x,y}/(\mu_x\mu_y)) + \mathrm{Log}\mu_y\right] \\ &= -\sum_{x,y\in\Omega} P_{x,y}\left[\mathrm{Log}P_{x,y} - \mathrm{Log}(\mu_x\mu_y)\right] - \sum_{y\in\Omega} \mu_y \mathrm{Log}\mu_y = \mathrm{E}_\mu - \mathrm{d_{KL}}(P_{x,y}, \mu_x \otimes \mu_y)\end{aligned}$$

□

Let us do now 4 remarks:

(1) $\mathrm{d_{KL}}(P_{x,y}, \mu_x \otimes \mu_y)$ is an index of dependence between the successive random variables X_t and X_{t+1} constituting the Markov chain.

(2) When the Markov is a Bernoulli shift, then all the lines of the transition matrix $(M_{xy})_{y\in\Omega}$ are the same and X_t and X_{t+1} are independent. In this particular case, we have: $\mathrm{E} = \mathrm{E}_\mu$. It is for example the case when the Markov process represents the backward succession of the mother ages in the genealogies of a population dynamics formalized by the Leslie model (Abbas et al. 2009).

(3) Another circumstance in which $\mathrm{E} = \mathrm{E}_\mu$ is observed in getBrens, when T is tending to infinity, because both μ and $(M_{xy})_{y\in\Omega}$ are then the uniform distribution on Ω.

(4) On the contrary, $\mathrm{E} \approx 0 \Leftrightarrow \mathrm{E}_\mu \approx \mathrm{d_{KL}}(P_{x,y}, \mu_x \otimes \mu_y)$ when the asymptotic measure μ is uniform over the attractors, with a weak recurrence rate ε for the

not fixed configurations. For example, let us suppose that the genetic network is made of 2 genes, i.e., $\Omega = \{0,1\}^2$, and that the network has one attractor, the fixed point 01, with attraction basin {00}, and the invariant cycle (10,11). Then, if ε is sufficiently small:

$$\begin{array}{r} \text{final state} \rightarrow \\ \text{initial state} \downarrow \\ \\ M = \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} \varepsilon & 1-\varepsilon & 0 & 0 \\ \varepsilon & 1-\varepsilon & 0 & 0 \\ 0 & 0 & \varepsilon & 1-\varepsilon \\ 0 & 0 & 1-\varepsilon & \varepsilon \end{pmatrix} \end{array} \begin{array}{l} 00 \\ 01 \\ 10 \\ 11 \end{array}$$

, $\mu = (\varepsilon/3, (1-\varepsilon)/3, 1/3, 1/3)$ and $E \approx -\varepsilon \mathrm{Log}\varepsilon$.

Let consider now the symmetric Kullback–Leibler divergence d^*_{KL} between the measures μ and ν, defined by the following relationship:

$$d^*_{KL}(\mu,\nu) = \sum_{x\in\Omega} \mu_x[\mathrm{Log}\mu_x - \mathrm{Log}\nu_x] + \sum_{x\in\Omega} \nu_x[\mathrm{Log}\nu_x - \mathrm{Log}\mu_x].$$

Let suppose that μ and ν are close: $\mu_x = \nu_x(1+\varepsilon_x)$, where ε_x is small and $\Sigma_{x\in\Omega}\nu_x\varepsilon_x = 0$ and that ν is uniform over a subset C of cardinal C contained in Ω. Then, we can calculate an estimation of $d^*_{KL}(\mu,\nu)$ which shows its relationship with the classical quadratic distance between the vectors μ and ν.

Proposition 2 *We have:* $d^*_{KL}(\mu,\nu) \approx C\|\mu-\nu\|^2$, *where* $\|.\|$ *denotes the Euclidean norm.*

Proof Let suppose that: $\mu_x = \nu_x(1+\varepsilon_x)$, where ε_x is small. Then, we have:

$$d^*_{KL}(\mu,\nu) = \sum_{x\in\Omega} \mu_x[\mathrm{Log}\mu_x - \mathrm{Log}\nu_x] + \sum_{x\in\Omega} \nu_x[\mathrm{Log}\nu_x - \mathrm{Log}\mu_x] = \sum_{x\in\Omega} \nu_x\varepsilon_x \mathrm{Log}(\mu_x/\nu_x) \approx \sum_{x\in\Omega} \nu_x\varepsilon_x^2$$

but $\sum_{x\in\Omega} \nu_x\varepsilon_x = 0$, hence $\left(\sum_{x\in\Omega} \nu_x\varepsilon_x\right)^2 = 0$, $\|\mu-\nu\|^2 = \sum_{x\in\Omega} \nu_x^2\varepsilon_x^2 = -2\sum_{x,y\in\Omega, x\neq y} \nu_x\nu_y\varepsilon_x\varepsilon_y$ and $\left(\sum_{x\in\Omega} \nu_x\varepsilon_x\right)\left(\sum_{y\in\Omega} \varepsilon_y\right) = 0$, hence $d^*_{KL}(\mu,\nu) \approx \sum_{x\in\Omega} \nu_x\varepsilon_x^2 = -2\sum_{x,y\in\Omega, x\neq y} \nu_x\varepsilon_x\varepsilon_y$. Then, by supposing that $|C| = C$, we get:

$$\|\mu-\nu\|^2 = -2\left(\sum_{x,y\in\Omega} \varepsilon_x\varepsilon_y\right) \Big/ C^2 \approx d^*_{KL}(\mu,\nu)/C$$

□

Consider now a particular Markovian matrix $\boldsymbol{M}$ having only coefficients 1, 1/2 or 0 such as:

(1) the probability to stay in a fixed configuration x is $\boldsymbol{M}_{xx} = 1$,
(2) the probability to stay in a not fixed configuration y is $M_{yy} = 1/2$,
(3) there is a configuration $z \neq y$ such as the probability to go from y to z is $\boldsymbol{M}_{yz} = 1/2$

(4) there exists a succession of K such transient configurations before coming back to y, called stochastic limit-cycle or periodic confiner. If $\boldsymbol{M}_{yy} = 0$ and $\boldsymbol{M}_{yz} = 1$, this succession is called attractor of period K (Demongeot and Jacob 1989, 1990). Then we have:

Proposition 3 *If $v_k = v_0 \boldsymbol{M}$ is close to the invariant measure μ of $\boldsymbol{M}$, where v_0 is an initial probability measure on Ω and if λ_2 denotes the subdominant eigenvalue of $\boldsymbol{M}$, there are constants L, B and K' such as:*

$$d^*_{KL}(\mu, v_k) \leq K' C e^{2k\,Log\lambda_2} \approx L e^{-kBE}$$

Proof Suppose that the transition matrix $\boldsymbol{M}$ has p fixed configurations z, for which $\boldsymbol{M}_{zz} = 1$, n stochastic limit-cycles of order $T_k (k = 1, \ldots, n) : (z(1), \ldots, z(T_k))$, for which $\boldsymbol{M}_{z(i)z(i)} = 1/2$, and $\boldsymbol{M}_{z(i)z((i+1)\mathrm{mod}T_k)} = 1/2$ and q deterministic limit-cycles of order T'_j $(j = 1, \ldots, q) : \left(z(1), \ldots, z\left(T'_j\right)\right)$, for which $\mathrm{M}_{z(i)z((i+1)\mathrm{mod}T'_j)} = 1$, the m-p-n-q remaining lines corresponding to configurations lying in the attraction basins $B(A_i)$ of the previous attractors, with a part $B_1(A_i)$ made of stochastic attracted points and a part $B_2(A_i)$ made of deterministic attracted points. Then, we have:

$$\begin{aligned}\mathrm{Det}(\boldsymbol{M} - \lambda I) = (1-\lambda)^p \Pi_{k=1,n}\left[(1/2 - \lambda)^{T_k} - (1/2)^{T_k}\right] \Pi_{j=1,q} \\ \left[(1-\lambda)^{T'_j} - 1\right] \Pi_{i=1,p+n+q}(1/2 - \lambda)^{|B_1(A_i)|} \Pi_{j=1,p+n+q}(-\lambda)^{|B_2(A_i)|}\end{aligned}$$

Then, if at least one basin is not empty, the subdominant eigenvalue λ_2 of $\boldsymbol{M}$ is equal to 1/2 and $\mu_x = 0$, except for the $a = p + \sum_{k=1,\ldots,n} T_k + \sum_{i=1,\ldots,q} T'_j$ attracting configurations y (i.e., fixed and periodic configurations) for which $\mu_y = 1/a$. Hence, if $K = \sum_{k=1,\ldots,n} T_k$, we have:

$$\mathrm{E} = -K[1/2\mathrm{Log}(1/2) + 1/2\mathrm{Log}(1/2)]/a = -K\mathrm{Log}(1/2)/a = -K\mathrm{Log}\lambda_2/a$$

For k sufficiently large, from Perron–Fröbenius theory, we have: $\|\mu - v_k\|^2 \leq K' \mathrm{e}^{2k\,\mathrm{Log}\lambda_2}$ and from Proposition 2., C being equal to the union of all the attractors, we get: $L = K'C$ and $B = 2a/K$. □

Results showing relationships between entropy and eigenvalues, like in Propositions 2. and 3. can be found in the literature (Goldstein 1981; Goldstein and Penrose 1981; Jensen et al. 2007; Shieh 2011). They show here an explicit relationship between a stability ($-\mathrm{Log}\lambda_2$) and a complexity (E) index, which increases in general with the size and connectivity of the network.

3.3 Markov Sequential Dynamics of Random Boolean Automata Networks

The Markovian matrix corresponding to the sequential transition rule of a random Boolean network of m automata is a 2^mx2^mmatrix $M = \left(M_{xy}\right)_{x,y\in\Omega} = \Pi_{i=1,\ldots,m} M_i$, where M_i is the transition matrix corresponding to the step of updating the

automaton i. For each configurations x and $x\backslash i$ (for $x\backslash\{i\}$) in Ω (a configuration x being identified to the set of automata having state 1), the lines of M_i corresponding to x and $x\backslash i$ constitute a 2–2 submatrix $M_{x,i}$ defined by:

$$M_{x,i} = \begin{pmatrix} \exp\left[\left(\sum_{j\in N_i} w_{ij}x_j - \theta_i\right)\Big/ T\right]\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}\beta_j - \theta_i\right)/T\right]\right] \\ 1\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j - \theta_i\right)/T\right]\right] \\ \exp\left[\left(\sum_{j\in N_i} w_{ij}(x\backslash i)_j - \theta_i\right)\Big/ T\right]\Big/\left[1+\exp\left[\left(\sum_{j\in M_i} w_{ij}\beta_j - \theta_i\right)/T\right]\right] \\ 1\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}(x\backslash i)_j - \theta_i\right)/T\right]\right] \end{pmatrix}$$

The matrix M_i is made of lines $(M_{xy,i})_{y\in\Omega}$ and the 2–2 submatrix $M_{x,i}$ is built, for every configuration x possessing i, from the non zero coefficients of M_i, $(M_{xy,i})_{y\in\Omega}$ and $(M_{x\backslash i,y,i})_{y\in\Omega}$, corresponding respectively to the lines x and $x\backslash i$. Sub-matrices $M_{x,i}$ serve to the calculation of the eigenvalues of M_i, because we can calculate the characteristic polynomial of M_i as: $\mathrm{Det}(M_i - \lambda I) = \Pi_{i=1,\ldots,m}\mathrm{Det}(M_{x,i} - \lambda I)$, then the subdominant eigenvalue λ_2 of M_i comes by calculating eigenvalues of the sub-matrices $M_{x,i}$ (cf. Fig. 3). We will suppose now in the following that: $w_{ij} = a_{ij}w$, for $i \neq j$, and $w_{ii} = v$ where a_{ij} is equal to 0 or 1. Then, if we denote the potentials $H_{x,i} = w\left(\sum_{j\in N_i} a_{ij}x_j - \theta_i/w\right)/T$ and $H_{x\backslash i,i} = w\left(\sum_{j\in N_i} a_{ij}(x\backslash i)_j - \theta_i/w\right)/T$, where

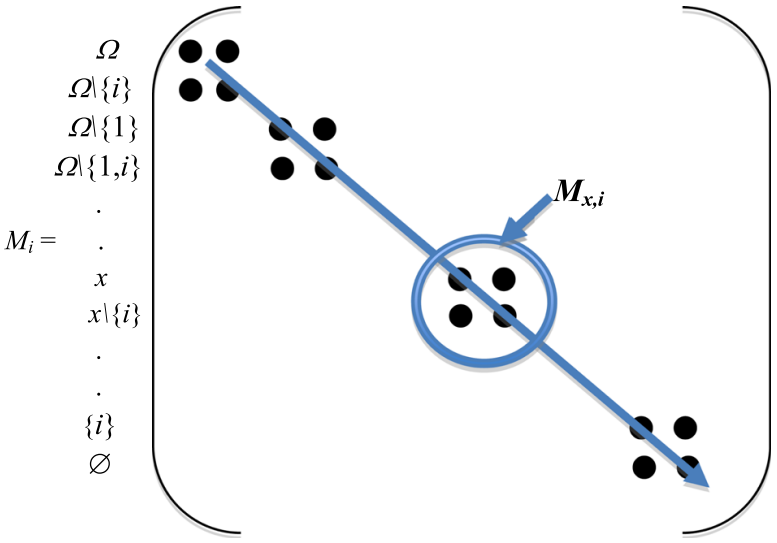


Fig. 3 Parametric schematic representation of the architecture of the matrix M_i and of the submatrices $M_{x,i}$, by using the order $\Omega, \Omega\backslash\{i\}, \Omega\backslash\{1\}, \Omega\backslash\{1,i\}, \Omega\backslash\{2\}, \Omega\backslash\{2,i\}, \ldots, \Omega\backslash\{i-1\}, \Omega\backslash\{i-1,i\}, \Omega\backslash\{i+1\}, \Omega\backslash\{i+1,i\}, \ldots, x, x\backslash\{i\}, \ldots, \{i\}, \emptyset$, where we identify a configuration to the set of automata in state 1

N_i is the set of neighbours of i (including i), we have for the characteristic equation of the matrix $M_{x,i}$:

$$\begin{aligned}\mathrm{Det}(M_{x,i}-\lambda I) &= \mathrm{Det}\begin{pmatrix} e^{Hx,i}/(1+e^{Hx,i})-\lambda & 1/(1+e^{Hx,i}) \\ e^{Hx\backslash i,i}/\left(1+e^{Hx\backslash i,i}\right) & 1/\left(1+e^{Hx\backslash i,i}\right)-\lambda \end{pmatrix} \\ &= \lambda^2 - \left[e^{Hx,i}/\left(1+e^{Hx,i}\right)+1/\left(1+e^{Hx\backslash i,i}\right)\right]\lambda \\ &\quad + \left(e^{Hx,i}-e^{Hx\backslash i,i}\right)/\left(\left(1+e^{Hx,i}\right)\left(1+e^{Hx\backslash i,i}\right)\right)\end{aligned}$$

3.4 The Example of the m-Switch

If $\theta_i = 0$, $w < 0$, $v > 0$ (resp. $w > 0$, $v < 0$) and $a_{ij} = 1$, the Boolean network is called local m-switch (resp. anti-m-switch), the local character referring to the neighbourhood N_i (see on Fig. 2 in bottom right quadrant a local m-switch, in which $|N_i| = 2\mathrm{d}+1$, the switch occurring between two asymptotic dynamical regimes depending on boundary conditions). Let suppose in the following that m-switch and anti-m-switch are global, i.e., N_i is equal to the set of all automata, and that x contains $d(x)$ (d in absence of ambiguity) nodes of $N_i\backslash i$, plus i. Then, we have:

$$\begin{aligned}\mathrm{Det}(M_{x,i}-\lambda I) = 0 \Leftrightarrow \lambda^2 &- \left[e^{dw+v}/\left(1+e^{dw+v}\right)+1/\left(1+e^{dw}\right)\right]\lambda \\ &+ \left(e^{dw+v}-e^{dw}\right)/\left(\left(1+e^{dw+v}\right)\left(1+e^{dw}\right)\right) = 0, \text{ hence:} \\ \lambda = &\left[\left(1+\left(e^{dw+v}-e^{dw}\right)/\left(\left(1+e^{dw+v}\right)\left(1+e^{dw}\right)\right)\right)\right. \\ &\pm\left[\left(\left(1+\left(e^{dw+v}-e^{dw}\right)/\left(\left(1+e^{dw+v}\right)\left(1+e^{dw}\right)\right)\right)^2\right.\right. \\ &\left.\left.-4\left(e^{dw+v}-e^{dw}\right)/\left(\left(1+e^{dw+v}\right)\left(1+e^{dw}\right)\right)\right)\right]^{1/2}\Big]\Big/2\end{aligned}$$

and λ is independent of i, as well as the subdominant eigenvalue λ_2 of M_i, such as:

$$\begin{aligned}\lambda_2 &= \sup_{x\ni i}\left[e^{d(x)w+v}-e^{d(x)w}\Big/\left(\left(1+e^{d(x)w+v}\right)\left(1+e^{d(x)w}\right)\right)\right] \\ &= \sup_{x\ni i}\left[(1-e^{-v})\Big/\left(\left(1+e^{-d(x)w-v}\right)\left(1+e^{d(x)w}\right)\right)\right] \\ &= (1-e^{-v})/\inf_x\left(1+e^{-v}+e^{-d(x)w-v}+e^{d(x)w}\right) \\ &= (1-e^{-v})/\left(1+e^{-v}+\inf_x\left(e^{-d(x)w-v}+e^{d(x)w}\right)\right).\end{aligned}$$

The Kolmogorov-Sinaï entropy E_i of M_i is equal to: $\mathrm{E}_i = -\sum_x \mu_x \sum_y M_{xy,i}\mathrm{Log}(M_{xy,i})$, where: $\forall x \in \Omega, \mu_x = \mu(\{x\})$, μ being the equilibrium invariant measure of M_i, then:

$$
\begin{aligned}
\mathrm{E}_i = -\sum_{x\ni i}\Big[\mu_x\Big[&\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\right.\\
&\mathrm{Log}\left(\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\Big/\left(1+\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\right)\right.\\
&+1\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\right]\\
&\mathrm{Log}\left(1\Big/\left[1+\exp\left[\left(\sum_{j\in N_i} w_{ij}x_j-\theta_i\right)\Big/T\right]\right]\right.\\
&+\mu_{x\setminus i}\left[\exp\left[\left(\sum_{j\in N_i} w_i(x\setminus i)_j-\theta_i\right)\Big/T\right]\right]\Big/\left[1+\exp\left(\left(\sum_{j\in N_i} w_{ij}(x\setminus i)_j-\theta_i\right)/T\right)\right]\\
&\mathrm{Log}\left(\exp\left[\left(\sum_{j\in N_i} w_i(x\setminus i)_j-\theta_i\right)\Big/T\right]\Big/\left[1+\exp\left(\left(\sum_{j\in N_i} w_{ij}(x\setminus i)_j-\theta_i\right)\Big/T\right)\right]\right)\\
&+1\Big/\left(1+\exp\left[\left(\sum_{j\in N_i} w_{ij}(x\setminus i)_j-\theta_i\right)/T\right]\right)\\
&\left.\left.\left.\mathrm{Log}\left(1\Big/\left(1+\exp\left[\left(\sum_{j\in N_i} w_{ij}(x\setminus i)_j-\theta_i\right)/T\right]\right)\right)\right)\right]\right]
\end{aligned}
$$

In the sequential update case, and if $\nu = 0$ $\mu_x = \exp[(\sum_{k\in x, j\in N_k} w_{kj}x_j - \theta_k)/T]/\sum_y \exp[(\sum_{k\in y, j\in N_k} w_{kj}x_j - \theta_k)/T]$ is the Gibbs measure common to all M_i's and $\mu_{x\setminus i} = \mu_x e^{-d(x)w}$. Hence, we will denote E_i by E and, if w>>0:

$$\mathrm{E} = -\sum_{x\ni i}\mu_x\left[(d(x)w+v)e^{d(x)w+v}/\left(1+e^{d(x)w+v}\right) - \mathrm{Log}\left(1+e^{d(x)w+v}\right)\right].$$

Let consider now a perturbation of intensity ε on the value $w_{ii} = v = 0$ of the self-interaction (called also external field), changing M_i in the perturbed matrix M_i^ε with new coefficients $M^\varepsilon_{xy,i} = M_{xy,i} + \varepsilon_{xy}$ and the Gibbs invariant measure μ of M_i in the perturbed measure μ_ε. Then, by using the Kullback–Leibler divergence $\mathrm{d_{KL}}(\mu,\mu_\varepsilon) = -\sum_x \mu_x \mathrm{Log}(\mu_{\varepsilon x}/\mu_x)$ between measures μ and μ_ε and by denoting $\boldsymbol{E}$ the expectation calculated for the invariant measure μ, we have:

$$
\begin{aligned}
\mathrm{d_{KL}}(\mu,\mu_\varepsilon) = -\sum_{x\ni i}\mu_x\Big[&e^{d(x)w}(\mathrm{Log}[e^{d(x)w+\varepsilon}/(1+e^{d(x)w+\varepsilon})]-\mathrm{Log}[e^{d(x)w}/\left(1+e^{d(x)w}\right)])\\
&\Big/\left(1+e^{d(x)w}\right)+(\mathrm{Log}[1/(1+e^{d(x)w+\varepsilon})]-\mathrm{Log}[1/\left(1+e^{d(x)w}\right)])/(1+e^{d(x)w})\Big],
\end{aligned}
$$

by supposing that ε is small:

$$\begin{aligned}\mathrm{d_{KL}}(\mu,\mu_\varepsilon) \approx & -\boldsymbol{E}[\varepsilon e^{d(x)w}/\left(1+e^{d(x)w}\right)^2 - \varepsilon e^{d(x)w}/\left(1+e^{d(x)w}\right)^2 - \varepsilon^2/2 + \varepsilon^2 e^{2d(x)w} \\ & /2\left(1+e^{d(x)w}\right)^2 + \mathrm{Log}\left[1/\left(1+e^{d(x)w}\right)\right] e^{d(x)w}/\left(1+e^{d(x)w}\right)]/2 \\ & = \boldsymbol{E}[\varepsilon^2(1-e^{2d(x)w}/\left(1+e^{d(x)w}\right)^2]/2 = \varepsilon^2\boldsymbol{E}[(1+\mathrm{A}(x))(1-\mathrm{A}(x))]/2,\end{aligned}$$

where $\mathrm{A}(x) = e^{d(x)w}/\left(1+e^{d(x)w}\right), e^{d(x)w} = \mathrm{A}(x)/(1-\mathrm{A}(x))$, because:

$$\begin{aligned}&\mathrm{Log}[e^{d(x)w+\varepsilon}/(1+e^{d(x)w+\varepsilon})] = \mathrm{Log}(e^{d(x)w+\varepsilon}) - \mathrm{Log}(1+e^{d(x)w+\varepsilon}) \approx \mathrm{Log}(e^{d(x)w} + \varepsilon e^{d(x)w}) \\ &- \mathrm{Log}(1+e^{d(x)w} + \varepsilon e^{d(x)w}) \approx d(x)w + \varepsilon - \varepsilon^2/2 - \mathrm{Log}\left(1+e^{d(x)w}\right) - \varepsilon e^{d(x)w} \\ &/\left(1+e^{d(x)w}\right) + \varepsilon^2 e^{2d(x)w}/2\left(1+e^{d(x)w}\right)^2 = \varepsilon/\left(1+e^{d(x)w}\right) + \varepsilon^2 e^{2d(x)w}/2\left(1+e^{d(x)w}\right)^2 \\ &+ \mathrm{Log}\left[e^{d(x)w}/\left(1+e^{d(x)w}\right)\right] \text{ and} \mathrm{Log}[1/(1+e^{d(x)w+\varepsilon})] = \mathrm{Log}\left(1/\left(1+e^{d(x)w}\right)\right) \\ &- \varepsilon e^{d(x)w}/\left(1+e^{d(x)w}\right) + \varepsilon^2 e^{2d(x)w}/2\left(1+e^{d(x)w}\right)^2\end{aligned}$$

In m-switch and anti-m-switch cases, the subdominant eigenvalue λ_2^ε of the matrix M_i^ε is given by:

$$\begin{aligned}\lambda_2^\varepsilon &= \sup_{x\ni i}[(1-e^{-\varepsilon})e^{d(x)w+\varepsilon}/((1+e^{d(x)w+\varepsilon})(1+e^{d(x)w}))] \\ &= \sup_{x\ni i}[(e^\varepsilon - 1)\mathrm{A}(x)/(1+e^\varepsilon \mathrm{A}(x)/(1-\mathrm{A}(x)))], \text{ hence, if } f(t) \\ &= t(1-t) \text{ and } \varepsilon \text{ is small} : \lambda_2^\varepsilon \approx \varepsilon \sup_{x\ni i}[\mathrm{A}(x)(1-\mathrm{A}(x))] = \varepsilon \sup_{x\ni i}[f(\mathrm{A}(x))].\end{aligned}$$

For the Kolmogorov-Sinaï entropy E of the transition matrix M_i, we have:

$$\begin{aligned}\mathrm{E} &= -\sum_x \mu_x \left[e^{d(x)w}\mathrm{Log}\left[e^{d(x)w}/\left(1+e^{d(x)w}\right)\right]/\left(1+e^{d(x)w}\right)\right. \\ &\quad \left. +\mathrm{Log}\left[1/\left(1+e^{d(x)w}\right)\right]/\left(1+e^{d(x)w}\right)\right] \\ &= -\boldsymbol{E}[\mathrm{A}(x)\mathrm{Log}(\mathrm{A}(x)) + (1-\mathrm{A}(x))\mathrm{Log}(1-\mathrm{A}(x))], \\ &\text{hence, if } |w| \text{ is sufficiently large, we have} : \mathrm{E} \approx E[-\mathrm{A}(x)Log(A(x))].\end{aligned}$$

Then, in m-switch and anti-m-switch cases, the perturbation observed for the Kullback–Leibler distance between initial and perturbed invariant measures is dominated, if $|w|$ is large, by the sub-dominant eigenvalue λ_2^ε, as well as by the entropy E through the following inequality:

$$\boldsymbol{E}(\mathrm{d_{KL}}(\mu,\mu_\varepsilon)) \approx \varepsilon^2\boldsymbol{E}[(1+\mathrm{A}(x))(1-\mathrm{A}(x))]/2 \le \varepsilon^2(1-\mathrm{E})/2 \le \varepsilon(\varepsilon+\lambda_2^\varepsilon)$$

If we consider the function $g_a(w) = e^{aw}/(1+e^{aw})$, Fig. 4 shows for example that, for $d(x) = 0,\ldots,4, g_{d(x)}(w)^2 = \mathrm{A}(x)^2 \approx 0$, when $w \le -1$, and the inequality rapidly holds for the m-switch case.

In the sequential update, the Gibbs measure μ is the common eigenvector related to the eigenvalue 1 both for the transition matrix M and for all matrices M_i. In the m-

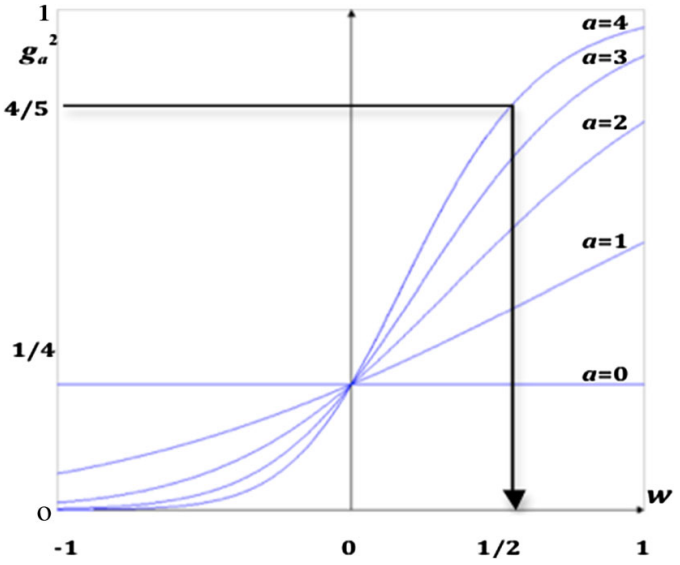


Fig. 4 Graph of the function $g_a^2(w) = e^{2aw}/(1+e^{aw})^2$

switch and anti-m-switch cases, the interaction architecture being the same for all automata, the eigenvalues of the M_i's are the same and the distance $d_{KL}(\mu, \mu_\varepsilon)$ between the Gibbs measures before and after a perturbation ε of the external field, is majored by $m\varepsilon^2(1-E)$ and $m\varepsilon\left(\varepsilon + \lambda_2^\varepsilon\right)$, which dominates the amplitude of change of the Gibbs measure μ. Hence, the robustness of M increases, when $-E$ or λ_2^ε decreases, and $-E$ or λ_2^ε can serve in this case as non-robustness index.

3.5 Evolutionary Entropy as Measure of the Stability

Let us give another example of the role of the evolutionary entropy as index of stability in Markov chains. Let consider a Leslie matrix L corresponding to a growth model of a hybrid population made of biological objects (animals, cells, molecules,…) belonging to 2 sub-populations A and B, characterized each by two ages (types or states) 1 and 2, whose size at time t and fecundity (reproducibility in the class 1) corrected by mortality (catabolism) are the same, but A having a higher entropy than B. Imagine that one could let type A (resp. B) reproduce with probabilities f_{1A}, f_{2A} (resp. f_{1B}, f_{2B}) at ages 1 and 2, die at age 1 with probability δ_A (resp. δ_B), or equivalently survive with probability $b_A = 1 - \delta_A$ (resp. $b_B = 1 - \delta_B$) between times t and $t+1$. Assume $f_{1A} > f_{1B}$ and keep equal the Malthusian parameters $r_A = \log\lambda_A$ and $r_B = \log\lambda_B$, defined from the dominant eigenvalues λ_A and λ_B of the Leslie matrices L_A and L_B defined by:

$$L_A = \begin{pmatrix} f_{1A} & f_{2A} \\ b_A & 0 \end{pmatrix} \quad \text{and} \quad L_B = \begin{pmatrix} f_{1B} & f_{2B} \\ b_B & 0 \end{pmatrix}$$

The growth rates of populations A and B are the same, but the population A is most successful in robustness because of its larger entropy E_A. Indeed, it is easy to calculate:

$$\lambda_A = f_{1A}/2 + (\varphi_{2A})^{1/2}(1 + f_{1A}^2/\varphi_{2A})^{1/2} = \lambda_B = f_{1B}/2 + (\varphi_{2B})^{1/2}(1 + f_{1B}^2/\varphi_{2B})^{1/2},$$

where $\varphi_{2A} = f_{2A} \cdot b_A$ is the fecundity of the class 2 corrected by the survival rate of the class 1. If $f_{1A} > f_{1B}$ and $\lambda_A = \lambda_B$, then $\varphi_{2A} < \varphi_{2B}$. The Markov chain ruling the backward genealogies (Demetrius et al. 2004; Demetrius and Ziehe 2007), i.e., the succession of the mother ages of individuals in the same lineage, is a Bernoulli shift (Abbas et al. 2009) defined by the Markov transition matrix M_A:

$$M_A = \begin{pmatrix} f_{1A}/\lambda_A & 1 - f_{1A}/\lambda_A \\ 1 & 0 \end{pmatrix}$$

The evolutionary entropy E_A of M_A (resp. E_B of M_B) is defined, if $A = \lambda_A/f_{1A}$(resp. $B = \lambda_B/f_{1B}$) is sufficiently large, from M_A (resp. M_B) and its stationary measure $\mu_A = (A/(2A-1), (A-1)/(2A-1))$ (resp. $\mu_B = (B/(2B-1), (B-1)/(2B-1))$ by:

$$\begin{aligned} E_A = &- A/(2A-1)[f_{1A}/\lambda_A \mathrm{Log}([f_{1A}/\lambda_A) + (\lambda_A - f_{1A})/\lambda_A \mathrm{Log}([(\lambda_A - f_{1A})/\lambda_A)] \\ &\approx f_{1A}/\lambda_A = 1/(2A-1)(\text{resp. } E_B = -B/(2B-1)[f_{1B}/\lambda_B \mathrm{Log}([f_{1B}/\lambda_B) \\ &+ (\lambda_B - f_{1B})/\lambda_B \mathrm{Log}([(\lambda_B - f_{1B})/\lambda_B)] \approx f_{1B}/\lambda_B = 1/(2B-1)). \end{aligned}$$

Because $f_{1A} > f_{1B}$ and $\lambda_A = \lambda_B$, then $A < B$ and $E_A > E_B$. If we perturb the stationary measure μ_A of M_A, defined by: $\mu_A = (A/(2A-1), (A-1)/(2A-1))$, then we obtain from the perturbed measure $\mu_{A\varepsilon}$, after one time step, the measure:

$$\begin{aligned} &(A/(2A-1) - \varepsilon, (A-1)/(2A-1) + \varepsilon)M_A \\ &= (A/(2A-1) + \varepsilon(A-1)/A, (A-1)/(2A-1) - \varepsilon(A-1)/A) \end{aligned}$$

The Kullback–Leibler divergence between $\mu_{A\varepsilon}$ and μ_A (resp. $\mu_{A\varepsilon}M_A$ and μ_A) is equal to:

$$\begin{aligned} d_{KL}(\mu_{A\varepsilon}, \mu_A) &\approx [(\mathrm{Log}(1 - \varepsilon(2 - 1/A)) + \mathrm{Log}(1 + \varepsilon(2 + 1/(A-1)))]/2 \\ &\approx \varepsilon(2A-1)/2A(A-1)(\text{resp. } d_{KL}(\mu_{A\varepsilon}M_A, \mu_A) \\ &\approx [\mathrm{Log}(1 + \varepsilon((A-1)(2A-1)/A^2) + \mathrm{Log}(1 - \varepsilon(2A-1)/A)]/2 \\ &\approx -\varepsilon(2A-1)/2A^2), \text{ then } |d_{KL}(\mu_{A\varepsilon}M_A, \mu_A)/d_{KL}(\mu_{A\varepsilon}, \mu_A)| \\ &\approx (A-1)/A \approx (1 - E_A)/(1 + E_A) < (B-1)/B \approx (1 - E_B)/(1 + E_B) \\ &\approx |d_{KL}(\mu_{B\varepsilon}M_{B\varepsilon}, \mu_B)/d_{KL}(\mu_{B\varepsilon}, \mu_B)| \end{aligned}$$

Hence, because the sub-population A is more stable than the sub-population B and $E_A > E_B$, the evolutionary entropy E is correlated with the stability rate and, hence, can serve for quantifying the distance to the stationary measure observed after a dynamical perturbation.

More generally, if the dynamics is defined in discrete (resp. continuous) population dynamics by:

$V(x,t) = x(t+1) - x(t) = (A - I)\, x(t)$ (resp. $dx/dt = F(t)$), it is possible to get a Markov matrix M acting on the backward genealogies: $M = (m_{ij}) = (a_{ij} \cdot m_j/\lambda m_i)$, where λ is the dominant eigenvalue of A supposed to be

positive definite and m is its normalized eigenvector (resp. $M = (m_{ij}) = (\lim_{\Delta F,\Delta x\to 0}\partial \mathrm{Log}\Delta F_i/\partial \mathrm{Log}\Delta x_j) = (J_{ij}F_j/\gamma_i)$, where J is the Jacobian matrix of the continuous system, $\gamma_i = d^2x_i/dt^2$ the acceleration and m_{ij} the control strength the subpopulation j of size x_j exerts on the growth of subpopulation i). M is used in discrete demography like in Leslie model (Demetrius 1978; Demongeot and Demetrius 1989) or in continuous kinetics like in enzymatic models (Reder 1988; Baconnier et al. 1993).

3.6 Evolutionary Entropy as Measure of the Connectivity

The evolutionary entropy E is a measure of the degree of connectivity between final states of the Markov chain, this property deriving from the Shannon-McMillan-Breiman theorem (Demetrius et al. 2004; Demetrius and Ziehe 2007), which says that, for arbitrarily small ε and $\delta > 0$, and for sufficiently large t, the set of the partial trajectories X of length t (i.e., the collection of states X_k of the system at time k observed between times $k = 0$ and $k = t - 1$) generated by the Markovian dynamics, can be divided into 2 classes S_1 and S_2 such that:

(a) for every X in S_1, we have: $|\mathrm{Log}\ \mu(X)/t + E| < \varepsilon$
(b) the sum of the probabilities of the partial trajectories X belonging to S_2 is less than δ.

All partial trajectories of the class S_1 have approximately the same probability, namely e^{-tE}. This means that the number of partial trajectories in S_1 is approximately e^{tE}, which specifies the number of "typical" partial individual trajectories and is positively correlated with E. (cf. also Donsker and Varadhan 1975; Freidlin and Wentzell 1984; Wentzell and Freidlin 1970)

4 Genetic Threshold Boolean Random Regulatory Networks (getBren)

We will now apply the results of the previous Sections to the genetic regulation (see Elena and Demongeot 2008; Demongeot et al. 2008b; Demongeot and Sené 2008; Demongeot et al. 2003, 2009a, b, 2010, 2011a, b, c) using the formalism of the genetic threshold Boolean regulatory networks (getBren), which is just the translation, in the genetic language, of the Hopfield like networks (Hopfield 1982; Hartwell et al. 1999; Weaver et al. 1999).

4.1 Biological Networks as Graphs

A biological network is described as a directed weighted graph G (digraph), defined in terms of nodes and arcs between the m nodes of G. The dynamical process defining the evolution of the network states invokes a configuration space Ω and a potential function denoted $H : \Omega \to \mathbb{R}$, built from interaction weights on the arcs of G. Configuration space Ω is the set of all doubly infinite paths on G. Let μ denote

the Gibbs equilibrium measure associated with the potential H. We will represent the biological network by the mathematical object (Ω, μ, H) and consider systems whose Gibbs measure μ is invariant under the shift τ on Ω. The main characteristics associated with this dynamical system is the Kolmogorov-Sinaï entropy E, called here evolutionary entropy (Demertius et al. 2004; Demetrius and Ziehe 2007).

4.2 Genetic Regulatory Networks

The expression of a gene, that is the production of the protein encoded by this gene, is regulated by the activity of other proteins. We will now describe the classical Boolean formalism coming from statistical mechanics (Hopfield 1982; Hartwell et al. 1999; Weaver et al. 1999; Demongeot et al. 2003) used for modelling the genetic networks, before applying it to particular networks like that ruling the cell cycle in different species. We will in particular show that, during the evolution, the robustness of this network has increased (e.g., by passing from *Cænorhabditis elegans* to *Drosophila melanogaster* and mammals).

5 Genetic Threshold Boolean Random Regulatory Networks (getBren) Definition

Let us start by defining a getBren N of size m: it is composed of m Boolean random automata interacting with each others over discrete time (the time space T is a subset of IN). The architecture of N is described by a digraph G in general no regular nor planar (Fig. 5a–c), contrarily to ideal gases, which are both regular and planar (Fig. 5d). Every random automaton i ($i = 1,\ldots,m$) of N owns a state x_i valued in {0,1} (0 means that gene i is inactivated and 1 that i is activated). Considering that $x_i(t)$ is the local state of the random automaton i at time t, we derive the notion of global state of the getBren at time t, called configuration, that is the vector: $x(t) = (x_i(t))_{i\in\{1,\ldots,m\}} \in \Omega = \{0,1\}^m$.

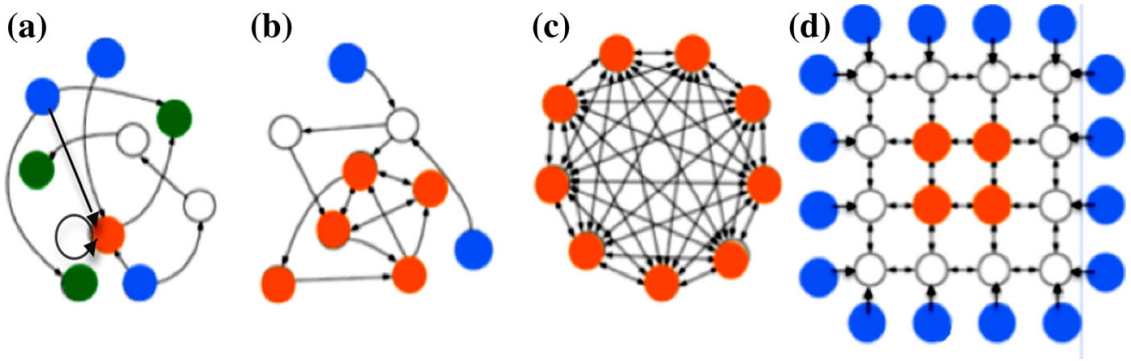


Fig. 5 Various networks whose interaction graph G has a unique strong connected component (scc) possessing central points (*red*), as well as sources (*blue*) and sinks (*green*) of this scc. Network connectivity increases: 10/9 (**a**), 16/9 (**b**), 72/9 (**c**) and 64/32 (**d**). **a**, **b** Realistic genetic network, **c** is a 9-switch used in morphogenesis and plant growth modelling and **d** represents a nearest neighbour ferromagnetic solid model. (Color figure online)

A getBren N of size m (cf. on Fig. 6, an example of size 5) is a triplet (W, Θ, P) where:

- W is a matrix of order m, whose coefficients $w_{ij}(w_{ij} \in \mathbb{R})$ represent the interaction weight a gene j has on gene i. Sign$(W) = (\text{sign}(w_{ij}))$ is adjacency matrix of network interaction graph G
- $\Theta = \{\theta_i\}$ is an activation threshold vector of dimension m, θ_i being activity threshold of gene i
- $P : P(\Omega) \rightarrow [0, 1]^m$, where $P(\Omega)$ is the set of all subsets of Ω and P a transition probability vector, i.e., a set of m local probability transitions P_i: $P(\Omega) \rightarrow$ [0,1], such as $P^{\beta}_{i,\gamma}$ is the probability for the gene i to go to state γ at time $t + 1$ from configuration β at time t (cf. Fig. 6):

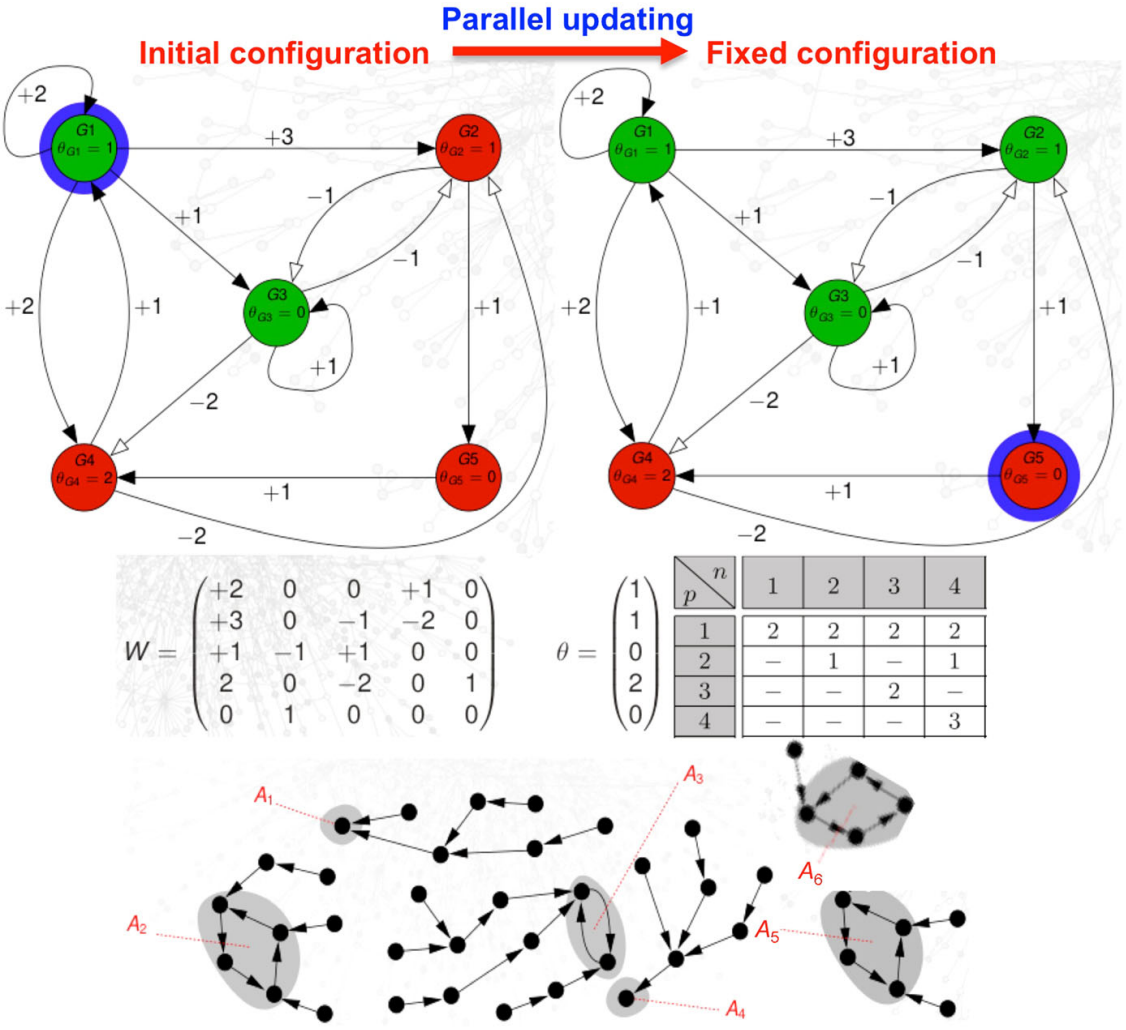


Fig. 6 A getBren composed of 5 genes G_i (i = 1,…,5) with their interaction digraph (whose positive circuit of length 4 provides 6 parallel attractors given below) and states (*green* for 1; red for 0) evolution in case of sequential (gene iterated and iteration number in *blue*) and parallel (*orange arrow*) iterations under a Boolean weighted majority deterministic rule with thresholds θ_i equal to 0, 1 or 2, showing the final fixed configuration (11101) observed from the initial configuration (10100). (Color figure online)

$$\forall \gamma \in \{0,1\}, \beta \in \Omega, P_{i,\gamma}^{\beta}(\{x_i(t+1)=\gamma | x(t)=\beta\}) = \exp\left[\gamma\left(\sum_{j\in Ni} w_{ij}\beta_j - \theta_i\right)/T\right] / \left(1+\exp\left[\left(\sum_{j\in Ni} w_{ij}\beta_j - \theta_i\right)/T\right]\right),$$

where N_i is the neighbourhood of gene i in getBren N, i.e., the set of genes j such that $w_{ij} \neq 0$.

P depends on the updating mode, and for extreme values of the randomness parameter T, we have:

(1) if $T = 0$, the getBren becomes deterministic and the transition $P_{i,\gamma}^{\beta}$ can be written as:

$$x_i(t+1)=1, \quad \text{if } \sum_{j\in Ni} w_{ij}\beta_j - \theta_i > 0; \quad x_i(t+1)=0,$$
$$\text{if } \sum_{j\in Ni} w_{ij}\beta_j - \theta_i < 0; \quad P(\{x_i(t+1)=1\}) = 1/2, \quad \text{elsewhere.}$$

The corresponding deterministic dynamics $x_i(t+1) = \mathrm{h}(\sum_{j\in Ni} w_{ij}\beta_j - \theta_i)$ is defined by the Heaviside function h (except for the value 0, where random rule holds: $x_i(t+1) = 1$ with probability 1/2),

(2) when T tends to infinite, then the transition probability $P_{i,\gamma}^{\beta}$ equals 1/2.

5.1 Entropy and Robustness in getBrens

Let $M = (M_{xy})$ denote the Markov matrix giving the transition probabilities of the getBren N, defined by an updating mode between configurations x and y of Ω, and $\mu = (\mu_x) = (\mu(\{x\}))_{x\in\Omega}$ be its stationary distribution on Ω. Evolutionary entropy E can be calculated for a getBren as:

$$\mathrm{E} = -\sum_{x,y\in\Omega} \mu_x M_{xy} \mathrm{Log} M_{xy}.$$

In the sequential updating mode, where the updating order of the nodes is the integer ranking 1,…,m, we have, by denoting $\boldsymbol{I} = \{1,\ldots,i-1\}$, $\boldsymbol{M\backslash I} = \{i,\ldots,m\}$ and identifying x with the set of the indices i such that $x_i = 1$:

$$M_{xy} = \Pi_{i=1,\ldots,m}\left[P_{i,1}^{[x\cap(M\backslash I)]\cup[y\cap I]} 1_{\{i\in y\}} + P_{i,0}^{[x\cap(M\backslash I)]\cup[y\cap I]} 1_{\{i\notin y\}}\right]$$

and μ is the Gibbs measure (Hopfield 1982) defined from the potential $H(x) = (\sum_{i\in x, j\in Ni} w_{ij}x_i x_j - \theta_i)/T$ by:

$$\forall x \in \Omega, \ \mu_x = \exp(H(x))/Z, \quad \text{where } Z = \sum_{y\in\Omega} \exp(H(y)).$$

When $T = 0$, μ is concentrated on the K ($\leq 2^m$) attractors of the deterministic dynamics and $E = 0$; when $T = +\infty$, μ is scattered uniformly over Ω and $E = m\text{Log}2$. Let define in this case one important feature of $B(A)$, the attraction basin of the attractor A (i.e., the set of all initial conditions, outside of A, having A as asymptotic behaviour): its Attraction Basin Relative Size, defined by $\text{ABRS}(A) = \sum_{x \in B(A) \cup A} \mu_x$. An example of calculation of ABRS is given in Figs. 5 and 8. We will in the following estimate in some cases E from a new quantity, the attractor entropy $E_{attractor}$ defined by: $E_{attractor} = -\Sigma_{k=1,\ldots,K}\, \text{ABRS}(A_k)\, \text{LogABRS}(A_k)$. In the following example, $E_{attractor}$ plays the same role as the divergence from independence, equal to $d_{KL}(P_{x,y}, \mu_x \otimes \mu_y)$ (cf. Proposition 1).

Proposition 4 *Let suppose that the invariant measure μ is uniform over all the $K \leq 2^m$ attractors of the deterministic dynamics on the configuration space Ω. Then, if for any configurations x,y of the attractor A_k or its basin $B(A_k)$, M_{xy} is defined with a local temperature T_{A_k} sufficiently large such as M_x. be scattered uniformly over $B(A_k) \cup A_k$, with $ABRS(A_k) = |B(A_k) \cup A_k|/2^m$, we have:*

$$E \approx \sum_{k=1,\ldots,K} ABRS(A_k) Log(2^m ABRS(A_k)) = mLog2 - E_{attractor} = E_\mu - E_{attractor}$$

Proof For any $k = 1,\ldots,K$, all transition matrix coefficients M_{xy} equal $1/|B(A_k) \cup A_k|$ on lines corresponding to configurations x belonging to $B(A_k) \cup A_k$ and μ is uniform over Ω. Then we can write: $E = -\sum_{k=1,\ldots,K} |B(A_k) \cup A_k|/2^m \text{Log}(2^m/(2^m |B(A_k) \cup A_k|)) = m\text{Log}2 - E_{attractor}$ □

Remarks

(1) a simple way to obtain a transition matrix M satisfying the conditions of the Proposition 4 is to impose to the transition probability to vanish outside the attraction basin of any attractor A_k and to be equal to a getBren transition inside, with a local stochasticity parameter T_{A_k}. Then, when all the parameters T_{A_k} increase, the invariant measure μ of M becomes the uniform measure over Ω.

(2) $E_{attractor}$ varies in opposite direction to E, as the index of dependence $d_{KL}(P_{x,y}, \mu_x \otimes \mu_y)$, then it can be considered as a parameter anti-correlated with the robustness.

We will now calculate in the framework of getBrens the sensitivity of the entropies E and E_μ (hence of $E_{attractor}$) to a variation of the parameter weights w_{ij}. Let us suppose for the sake of simplicity that $w_{ii} = 0$ and $\theta_i = 0$ for all the genes i and that $w_{ij} = \alpha_{ij} w$, where α_{ij} is the sign of the interaction of gene j on gene i and equals 0, −1 or +1, and $w > 0$ denotes the absolute value (supposed to be the same) of all non zero weights. Let consider now the random variable defined by:

$$U(x) = \sum_{i,j \in \{1,\ldots,m\}} \alpha_{ij} x_i x_j / T = (Q_+(N) - F(x))/T,$$

where $Q_+(N)$ is equal to the number of positive edges in the interaction graph G of the network N and $F(x)$ is the global self-frustration of x, *i.e.*, the number of pairs (i,j) where the values of x_i and x_j, are contradictory with the sign α_{ij} in the following sense: $F(x) = \sum_{i,j\in\{1,m\}} F_{ij}(x)$, where F_{ij} is the local self-frustration (Demongeot and Waku 2012a, b, c, d) of the pair (i,j) defined by:

$$F_{ij}(x) = 1, \text{ if } \alpha_{ij} = 1 \text{ and } x_i x_j = 0, \text{ or } \alpha_{ij} = -1 \text{ and } x_i x_j = 1, \text{ and } F_{ij}(x) = 0, \text{ elsewhere.}$$

The notion of frustration is very important for counting the attractors in circuits. For example, we can define an Hamiltonian function **H** on a circuit C = x(t) from the frustration F by:

$$\mathbf{H}(x(t)) = \sum_{i\in C}(x_i(t) - x_i(t-1))^2 = \sum_{i\in C}(\alpha_{i(i-1)}[x_{i-1}(t-1)] - x_i(t-1))^2 = F(x(t-1)),$$

if the transition rule is $\alpha_{i(i-1)i}[x_{i-1}(t-1)] = x_{i-1}(t-1)$ (resp.$1 - x_{i-1}(t-1)$), if $\alpha_{i(i-1)} = 1$ (resp.$\alpha_{i(i-1)} = -1$) and the local self-frustration is defined by:

$$F_{ij}(x) = 1, \text{ if } \alpha_{ij} = 1 \text{ and } x_i + x_j = 1, \text{ or } \alpha_{ij} = -1 \text{ and } x_i + x_j = 2 \ (\text{mod}2), \text{ and } F_{ij}(x) = 0, \text{ elsewhere.}$$

The Hamiltonian function **H**, conserved on the trajectories of configurations defined on the circuit C, equals twice the sum of discrete kinetic energies over all the nodes of C, *i.e.*, the frustration F, equal to the number of couples of successive nodes whose states are such as the value of the second is not equal to the value predicted by the first one. A continuous analogue of systems conserving the kinetic energy is for example Hamiltonian systems like the simple pendulum, for which the dynamical differential equations are: $dx/dt = y$ and $dy/dt = -x$ and the Hamiltonian function $\mathbf{H}(x) = (x^2 + y^2)/2$ is exactly the kinetic energy. For example, for a positive circuit of size $m = 8$, the values of frustration F are even (odd for a negative circuit). Table 1 gives the number of corresponding configurations, and the number and length of corresponding attractors (fixed state configurations or limit-cycles of state configurations):

For a positive circuit of size $m = 6$, we have the following even values of frustration F, the number of corresponding configurations (Table 2), and the number and length of corresponding attractors (fixed state configurations or limit-cycles of state configurations).

For a positive circuit of size $m = 4$, we have the following even values of frustration F, the number of corresponding configurations (Table 3), and the number and length of corresponding attractors (fixed state configurations or limit-cycles of state configurations).

Then, we have the following equalities, where the symbol ∂ denotes the partial derivative, all the functions considered depending on the two variables w and T.

Table 1 Values of the Hamiltonian **H** and number of corresponding configurations and attractors, with their length ($m = 8$)

H=F (frustration)	Configuration Number	Attractor Number	Attractor Length
0	1	2 fixed points	1
2	$4 = C^{H-1}_{n-H+2}/2$	7	8
4	$10 = C^{H-1}_{n-H+2}/2$	3	4 (symmetrized)
4		16	8
6	$4 = C^{n-H-1}_{H+2}/2$	7	8
8	1	1	2

Table 2 Values of the Hamiltonian **H** and number of corresponding configurations and attractors, with their length ($m = 6$)

H=F (frustration)	Configuration Number	Attractor Number	Attractor Length
0	1	2 fixed points	1
2	$3 = C^{H-1}_{n-H+2}/2$	5	6
4	$3 = C^{n-H-1}_{H+2}/2$	2	3 (symmetrized)
4		4	6
6	1	1	2

Table 3 Values of the Hamiltonian **H** and number of corresponding configurations and attractors, with their length ($m = 4$)

H=F (frustration)	Configuration Number	Attractor Number	Attractor Length
0	1	2 fixed points	1
2	$2 = C^{H-1}_{n-H+2}/2$	3	4
4	1	1	2

p \ n	1	2	3	4
1	2	2	2	2
2	–	1	–	1
3	–	–	2	–
4	–	–	–	3

The right hand side Table (Demongeot et al. 2012e) shows the number of attractors of length p for a positive circuit of size n

Proposition 5 *In sequential updating mode, we have:* $\partial E_\mu/\partial w = -wVarU = -wVarF/T^2$ *and* $\partial E_\mu/\partial T = wVarF/T^4$. *Hence,* $\left|\partial E_\mu/\partial w\right|$ *and* $\partial E_\mu/\partial T$ *decreases, when* $w > 0$ *decreases or when the temperature* T *increases,* μ *tending to the uniform distribution, for which* E_μ *is maximum.*

Proof We have: $E_\mu = -\sum_{x\in\Omega} \mu_x \mathrm{Log}\ \mu_x$, then $\partial E_\mu/\partial w = -\sum_{x\in\Omega} \partial\mu_x/\partial w\ \mathrm{Log}\mu_x - \sum_{x\in\Omega} \mu_x \partial \mathrm{Log}\mu_x/\partial w$, where $\partial\mu_x/\partial w = \partial[\exp((\sum_{i\in x, j\in Ni} w_{ij}x_i x_j - \theta_i)/T)/[\sum_{y\in\Omega} \exp((\sum_{j\in y,k\in N_j} w_{jk}y_j y_k - \theta_j)/T)]]/\partial w$. If $Z = \sum_{y\in\Omega} \exp((\sum_{j\in y,k\in Nj} w_{jk}y_j y_k - \theta_j)/T)$, then $\partial Z/\partial w = \sum_{y\in\Omega} (\sum_{i\in y,j\in Ni} \alpha_{ij}y_i y_j/T)\mu_y Z$, and:

$$\partial\mu_x/\partial w$$
$$= [\partial\left[\exp\left(\left(\sum_{i\in x,j\in Ni} w_{ij}x_ix_j - \theta_i\right)/T\right)/\partial w]/Z - \exp\left(\left(\sum_{i\in x,j\in Ni} w_{ij}x_ix_j - \theta_i\right)/T\right)(\partial Z/\partial w)/Z^2\right]$$
$$= \left(\sum_{i\in y,j\in N_i} \alpha_{ij}x_ix_j/T\right)\mu_x - \sum_{y\in\Omega}\left(\sum_{i\in y,j\in Ni} \alpha_{ij}y_iy_j/T\right)\mu_y\mu_x = \mu_x\partial \mathrm{Log}\ \mu_x/\partial w$$

and

$$\partial\mu_x/\partial w \mathrm{Log}\mu_x$$
$$= w\left(\sum_{i\in x,j\in Ni} \alpha_{ij}x_ix_j/T\right)^2\mu_x - \sum_{y\in\Omega}\left(\sum_{i\in y,j\in Ni} \alpha_{ij}y_iy_j/T\right)w\left(\sum_{i\in x,j\in Ni} \alpha_{ij}x_ix_j/T\right)\mu_y\mu_x$$
$$- \left(\sum_{i\in x,j\in Ni} \alpha_{ij}x_ix_j/T\right)\mu_x\mathrm{Log}Z + \sum_{y\in\Omega}\left(\sum_{i\in y,j\in Ni} \alpha_{ij}y_iy_j/T\right)\mu_y\mu_x\mathrm{Log}Z,$$

but:

$$\sum_{x\in\Omega}\partial\mu_x/\partial w = \partial(\textstyle\sum_{x\in\Omega}\mu_x)/\partial w = 0, \text{ therefore } \partial \mathrm{E}_\mu/\partial w = -\textstyle\sum_{x\in\Omega}\partial\mu_x/\partial w\ \mathrm{Log}\mu_x$$
$$= -wVarU = -wVarF/T^2$$

and

$$\partial \mathrm{E}_\mu/\partial T = -\sum_{x\in\Omega}\partial\mu_x/\partial T\ \mathrm{Log}\mu_x = VarU/T^2 = VarF/T^4.$$

□

By defining the local cross-frustration by: $C_{ij}(x,y) = 1$, if $\alpha_{ij} = 1$ and $x_iy_j = 0$, or $\alpha_{ij} = -1$ and $x_iy_j = 1$, and $C_{ij}(x,y) = 0$, elsewhere, then the global cross-frustration is defined by:

$$C(x,y) = \sum_{i,j\in\{1,\ldots,m\}} C_{ij}(x,y) = \Sigma_{i\in x,j\in y}\alpha_{ij}x_iy_j/T.$$

By denoting $\mathrm{E}_x = -\mu_x\sum_{y\in\Omega}M_{xy}\mathrm{Log}M_{xy}$, we get:

Proposition 6 *In the parallel updating mode, we have:* $\partial E_x/\partial w = -wVar_xC$, *where* Var_x *is related to the conditional measure defined by:*

$$M_{xy} = \exp\left(\left(\sum_{i\in x,j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right)/\sum_{y\in\Omega} exp\left(\left(\sum_{i\in x,j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right).$$

Proof The proof is the same as in Proposition 5, by replacing F by C. □

Proposition 7 *In the parallel updating mode, we have:* $\partial E/\partial w = -wVarC + Cov(E_M,C)$, *where Var and Cov are calculated for the random variables* $C(x,y)$ *and* $E_M(x,y) = -M_{xy}LogM_{xy}$, *by using the measure on* Ω^2 *defined by* $\mu_x M_{xy}$, *where the invariant measure* μ_x *for the parallel updating is defined as:*

$$\mu_x = \sum_{y\in\Omega} exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right)\Big/\sum_{x\in\Omega, y\in\Omega} exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right).$$

Proof We have: $\partial \mathrm{E}/\partial w = \partial E(\mathrm{E}_M)/\partial w = \Sigma_{x\in\Omega}(\mu_x \partial \mathrm{E}_x/\partial w + \mathrm{E}_x \partial \mu_x/\partial w)$. Hence, from Proposition 6, we get: $\partial E(\mathrm{E}_M/\partial w) = -wVarC + \Sigma_{x\in\Omega}\, \mathrm{E}_x\, \partial \mu_x/\partial w$, where:

$$\partial \mu_x/\partial w = \left[\sum_{y\in\Omega}\left[\partial \exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right)\Big/\partial w\right]Z\right.$$
$$\left. - \left[\sum_{y\in\Omega}\exp\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right](\partial Z/\partial w)\right]/Z^2,$$

where $Z = \sum_{x\in\Omega, y\in\Omega}\exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)/T\right)$ and $\partial Z/\partial w = \sum_{x\in\Omega, y\in\Omega}\left(\sum_{i\in x, j\in y}\alpha_{ij}x_iy_j/T\right)\exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)/T\right)$, then we have:

$$\partial \mu_x/\partial w = \sum_{y\in\Omega}\left(\sum_{i\in x, j\in y}\alpha_{ij}x_iy_j\Big/T\right)\exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right)\Big/$$
$$- \mu_x \sum_{x\in\Omega, y\in\Omega}\left(\sum_{i\in x, j\in y}\alpha_{ij}x_iy_j\Big/T\right)\exp\left(\left(\sum_{i\in x, j\in y} w_{ij}x_iy_j - \theta_i\right)\Big/T\right)$$
$$\Big/ Z = \mu_x(E_y(\mathrm{C}) - E(\mathrm{C}))$$

Finally, we get: $\partial \mathrm{E}/\partial w = E(\partial \mathrm{E}_M/\partial w) = -wVarC + \sum_{x\in\Omega}\mathrm{E}_x \partial \mu_x/\partial w = -wVarC + Cov(\mathrm{E}_M, C)$. □

We have shown, in Propositions 5, 6 and 7 above, a direct link between the sensitivity to w or T of E_μ and E, hence of $\mathrm{E}_{attractor} = \mathrm{E}_\mu - \mathrm{E}$, and the variability of the frustration of the network. We could show similar properties in the case of feed-forward block sequential updating, for which the invariant measure is also known (Demongeot et al. 2008b; Demongeot and Sené 2008).

6 Entropy and Robustness in Some Particular Examples

Let $\mu = \{\mu_\beta\}_{\beta\in\Omega}$ be the stationary distribution or invariant measure of the Markov probability transition matrix of a getBren. Then, the evolutionary entropy E can be

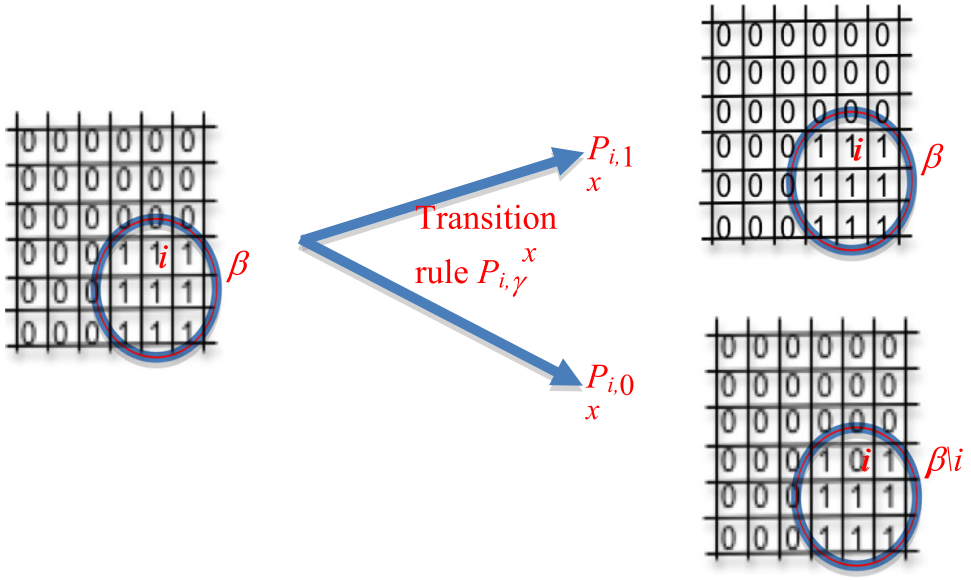


Fig. 7 Transition rule with the two possible choices for the state of the automata i decided under the probability $P^{\beta}_{i,\gamma}$ (the configuration β is identified to the set of the automata in state 1)

explicitely calculated: $\mathrm{E} = -\sum_{\gamma,\beta\in\Omega} \mu_\beta P^\beta_\gamma \mathrm{Log} P^\beta_\gamma$, as well as $\mathrm{E}_{attractor} = \mathrm{E}_\mu - \mathrm{E} = -\sum_{k=1,K} \mathrm{ABRS}(A_k)\mathrm{LogABRS}(A_k)$, from the ABRS of all attractors (cf. Fig. 8).

The main complexity parameter of $B(A)$, the attraction basin of an attractor A (i.e., the set of all initial conditions having A as asymptotic behaviour) is its Attraction Basin Relative Size (ABRS), defined by: $\mathrm{ABRS}(A) = \sum_{x\in B(A)} UA\mu_x$. There exists an intermediary temperature T between 0 and $+\infty$, for which for any $\beta \in B(A)$, the transition probability to γ, that is P^β_γ, is scattered rather uniformly over $B(A)UA$, then after proposition 5:

$$\mathrm{E} \approx \sum_{k=1,\ldots,m<2^m} \mathrm{ABRS}(A_k)\ \mathrm{Log}(2^m \mathrm{ABRS}(A_k)) = m\mathrm{Log}2\ -\ \mathrm{E}_{attractor}$$

Then, the attractor entropy $\mathrm{E}_{attractor}$ of the network controlling the feather morphogenesis in chicken (Abbas et al. 2009), whose interaction graph is pictured in Fig. 7 top panel, may be calculated for the parallel updating: $\mathrm{E} = 16\times 0.3 - \mathrm{E}_{attractor} = 4.8{-}9.9 \times 10^{-3}$. We can notice that $\mathrm{E}_{attractor}$ decreases from sequential to parallel schedule (from 7×10^{-2} to 9.9×10^{-3}), showing that the updating mode has an influence on the robustness of the network: synchrony increases robustness.

6.1 The Global m-Switch in Parallel Updating Case

Consider the case of the parallel update schedule of a global m-switch having an autocatalysis compensating the threshold ($w_{ii} = \theta_i \geq 0$) and a constant inhibitory weight $w < 0$ for the interactions between each pair of distinct nodes of the network. If the inhibitory interactions vanish at a certain distance of a node as in the

plant—like Araucaria—growth modelling (Cinquin and Demongeot 2002a, b; Demongeot et al. 2006a, b) this global m-switch can become local (cf. Section 2). We can calculate explicitly the transition probability M_{xy}, the invariant measure μ, the evolutionary entropy E and the attractor entropy $E_{attractor}$, when w is small (or T large):

$$\forall x, y \in \Omega, M_{yx} = \Pi_{i=1,\ldots,m}\left[P^y_{i,1}1_{\{i\in x\}} + P^y_{i,0}1_{\{i\notin x\}}\right],$$

where $\forall \gamma \in \{0,1\}, \beta \in \Omega, P^{\beta}_{i,\gamma} \quad (\{x_i(t+1) = \gamma | x(t) = \beta\}) = \exp(\gamma \sum_{j\in N_i} w\beta_j)/[1 + \exp(\sum_{j\in N_i} w\beta_j)]$,

$$\mu_x = \sum_{y\in\Omega} \exp\left(\sum_{i\in x, j\in y} wx_iy_j\right)/Z, \quad \text{where } Z = \sum_{z\in\Omega}\sum_{y\in\Omega} \exp\left(\sum_{k\in z, j\in y} wz_ky_j\right),$$

and:

$$\mathrm{E} = -\sum_{x,y\in\Omega} \mu_x M_{xy} \mathrm{Log}\, M_{xy} = m\mathrm{Log}2 \; - \; \mathrm{E}_{attractor}$$

Proposition 8 *For a global m-switch with $w_{ii} = \theta_i$ and a constant inhibitory interaction weight $w < 0$ sufficiently small, we have the following estimation of the evolutionary entropy E, attractor entropy $E_{attractor}$ and sensitivity $\partial E/\partial w$:*

$$E \approx mLog2 + wm(m-1)/2, E_{attractor} \approx -wm(m-1)/2 \text{ and } \partial E/\partial w = -\partial E_{attractor}/\partial w = C^2_m + O(w).$$

Proof In the case of the parallel update schedule, if $|x|$ denotes the number of 1 in configuration x, we can write:

$$M_{xy} = e^{w(|x|.|y|-|x\cap y|)}/[(1+e^{w|x|})^{m-|x|}(1+e^{w(|x|-1)})^{|x|}]$$

and:

$$\begin{aligned}\mu_x &= \sum_{j=0,|x|} \mathrm{C}^j_{|x|} \sum_{k-j=|y|-j=0,m-|x|} \mathrm{C}^{k-j}_{m-|x|} e^{w(|x|.(k-j+j)-j)}/Z \\ &= \sum_{j=0,|x|} \mathrm{C}^j_{|x|} e^{w(|x|-1)j}(1+e^{w|x|})^{m-|x|}/Z = (1+e^{w(|x|-1)})^{|x|}(1+e^{w|x|})^{m-|x|}/Z,\end{aligned}$$

where the normalisation constant Z verifies:

$$\mathrm{Z} = \sum_{z\in\Omega} (1+e^{w(|z|-1)})^{|z|}(1+e^{w|z|})^{m-|z|}$$

Let us denote $L(w,x) = (1+e^{w|x|})^{m-|x|}(1+e^{w|x|-1})^{|x|}$, then: $\mu_x = L(w,x)/\Sigma_{z\in\Omega}L(w,z)$ and

$$\begin{aligned}\mathrm{E}_x L(w,x) &= -\sum_{j=0,|x|} \mathrm{C}^j_{|x|} \sum_{k=|y|-j=0,m-|x|} \mathrm{C}^k_{m-|x|} e^{w(|x|(k+j)-j)} [w(|x|(k+j)-j) - \mathrm{Log}L(w,x)] \\ &= \sum_{j=0,|x|} \mathrm{C}^j_{|x|} e^{wj(|x|-1)} (1+e^{w|x|})^{m-|x|} [\mathrm{Log}L(w,x) - wj(|x|-1)] \\ &\quad -(1+e^{w|x|})^{m-|x|-1}(m-|x|)w|x|\Big]\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{E}_x &= L(w,x)\,[\mathrm{Log}L(w,x) - w|x|(|x|-1)(1+e^{w(|x|-1)})^{-1} \\ &\quad - w|x|(m-|x|)(1+e^{w|x|})^{|m|-2}]\,/L(w,x) \\ &= \mathrm{Log}L(w,x) - w|x|(|x|-1)(1+e^{w(|x|-1)})^{-1} - w|x|(m-|x|)(1+e^{w|x|})^{|m|-2}.\end{aligned}$$

Then, we can calculate E:

$$\begin{aligned}\mathrm{E} &= -\sum_{x,y\in\Omega} \mu_x M_{xy} \mathrm{Log} M_{xy} = \sum_{x\in\Omega} L(w,x)\ \mathrm{E}_x / \sum_{z\in\Omega} L(w,z) \\ &= \sum_{x\in\Omega} L(w,x)[-\mathrm{Log}L(w,x) + w|x|(|x|-1)(1+e^{w(|x|-1)})^{-1} \\ &\quad + w|x|(m-|x|)(1+e^{w|x|})^{-1}] \,/ \sum_{z\in\Omega} L(w,z) \\ &= -\sum_{i=0,m} \mathrm{C}^i_m L(w,i) \left[-\mathrm{Log}L(w,i) + wi(i-1)\Big(1+e^{w(i-1)}\Big)^{-1}\right. \\ &\quad \left. + wi(m-i)\big(1+e^{wi}\big)^{-1}\right] / \sum_{z\in\Omega} L(w,z)\end{aligned}$$

Therefore, if w is small, we have:

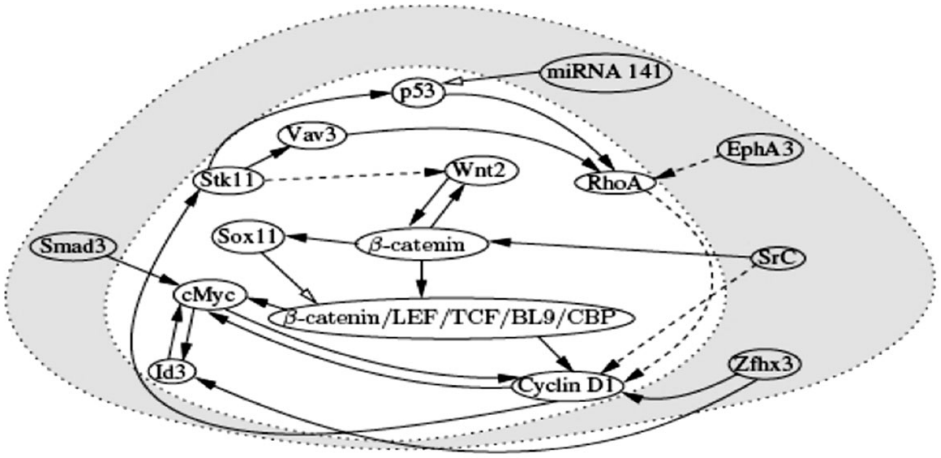
$$\begin{aligned}L(w,i) &= \big(1+e^{wi}\big)^{m-i}\Big(1+e^{w(i-1)}\Big)^{i} \approx 2^m[1 + i(m-i)w/2 + i(i-1)w/2] \\ &= 2^m[1 + i(m-1)w/2]\end{aligned}$$

and

$$Z = \sum_{z\in\Omega} L(w,z) \approx 2^{2m},\ \mu_x \approx [1 + i(m-1)w/2]/2^m,$$

$$\mathrm{Log}L(w,i) \approx m\mathrm{Log}2 + i(m-1)w/2,$$

then:



	Sequential updating		Parallel updating	
Nature	Attractor	ABRS	Attractor	ABRS
Fixed point 1	0000000000000000	1.56%	0000000000000000	≈ 0.00%
Fixed point 2	0011111001101110	96.88%	0011111001101110	99.66%
Fixed point 3	0000010000001100	1.56%	0000010000001100	≈ 0.00%
Limit cycle 1	None	–	0011010001101100 0000111000001110	0.34%

Fig. 8 *Top* Interaction signed digraph modelling the genetic regulation network of feather morphogenesis in chicken. Activation (resp. inhibition) is represented by a *black-headed* (resp. *white-headed*) arc. *Bottom* Attractors of the network for sequential and parallel schedules, with $T = 0$. Nodes order is: miRNA 141, EphA3, p53, Vav3, Stk11, Wnt2, RhoA, Smad3, SrC, Id3, Cyclin D1, Zfhx3, Sox11, β-catenin, cMyc and β-catenin/LEF/TCF/BL9/CBP

$$\begin{aligned}
\mathrm{E} &\approx -\sum_{i=0,m} \mathrm{C}_m^i \left[-\mathrm{Log}L(w,i) + 2^{-1}wi(i-1)(1-w(i-1)/2) \right. \\
&\quad \left. +2^{-1}wi(m-i)(1-wi/2) \right] / 2^m \\
&\approx \left(m2^m \mathrm{Log}2 + wm(m-1)2^{m-2} - wm(m-1)2^{m-2} \right) / 2^m \\
&= \left(m2^m \mathrm{Log}2 + wm(m-1)2^{m-1} \right) / 2^m = m\mathrm{Log}2 + wm(m-1)/2
\end{aligned}$$

Hence, we have: $\mathrm{E}_{attractor} \approx -wm(m-1)/2, \partial \mathrm{E}/\partial w = C_m^2 + \mathrm{O}(w)$ and $\partial \mathrm{E}_{attractor}/\partial w = -C_m^2 + \mathrm{O}(w)$. □

The robustness increases with m when w is small (or T large), because the number of fixed configurations in a deterministic m-switch equals m, then, when m tends to infinity, the number of attraction basins tends also to infinity, contributing to increase E and $\partial \mathrm{E}/\partial w$. It is interesting also to notice that in the case of the global m-switch, the derivative $\partial \mathrm{E}/\partial w$ is not depending on the frustration of the network like in Propositions 5–7 above. This observation comes from the fact that global m-switches are discrete Boolean potential systems (Tonnelier et al. 1999; Demongeot et al. 2006a, b, 2007, 2011a, b) contrarily to networks with only isolated circuits, which are purely Hamiltonian (Demongeot et al. 2011a, b, c, 2012e). Potential systems are dynamical systems defined on the state space Ω by:

$$x_i(t+1) = \mathbf{h}(-\Delta \boldsymbol{P}/\Delta x_i + x_i(t)) \text{ or } \Delta x_i/\Delta t = \mathbf{h}(-\Delta \boldsymbol{P}/\Delta x_i + x_i(t)) - x_i(t), \quad (1)$$

where $\Delta t = 1$, $\boldsymbol{P}$ is a real function (e.g., a polynomial with real coefficients valued in $\mathbb{R}$) on Ω and $\mathbf{h}$ a function from $\mathbb{R}$ to Ω, with boundary conditions ensuring that the flow remains in Ω.

The second formulation is the discrete equivalent of continuous differential equation $dx_i/dt = -\partial \boldsymbol{P}/\partial x_i$. In the Boolean case, we will choose for $\mathbf{h}$ the Heaviside function: $\mathrm{h}(s) = 1$, if $s > 0$, and $\mathrm{h}(s) = 0$, if not. If $\boldsymbol{P}(x) = \Sigma_k({}^t x A_k x)x_k + {}^t xWx + Bx$, where $A = (a_{ijk})$ is an interaction tensor (cf. Cosnard and Goles (1997)) expressing nonlinear triplet interactions with $A_k = \left(a_{ij}\right)_k$ as marginal matrices and $a_{iii} = 0, W = (w_{ij})$ is an interaction matrix and $\Theta = (\theta_i)$ a threshold line vector, the partial space derivatives of $\boldsymbol{P}$ are:

$$\begin{aligned}\Delta \boldsymbol{P}/\Delta x_i = {} & \Sigma_{j,k}\left(a_{ijk} + a_{jik} + a_{jki}\right)x_j x_k + \Sigma_j\left(w_{ij} + w_{ji}\right)x_j - \theta_i + [w_{ii} \\ & + \Sigma_{j\neq i}\left(a_{ijj} + a_{jij} + a_{jji}\right)x_j]\Delta x_i \end{aligned} \tag{2}$$

Then, the discrete potential system associated to $\boldsymbol{P}$ is defined in (Demongeot et al. 2006b), where the discrete velocity $\Delta x_i/\Delta t$ equals the opposite of the gradient of the potential:

$$\begin{aligned}\Delta x_i/\Delta t = \Delta x_i &= -\Delta P/\Delta x_i \\ &= -\Sigma_{j,k}\left(a_{ijk} + a_{jik} + a_{jki}\right)x_j x_k - \Sigma_j\left(w_{ij} + w_{ji}\right)x_j + \theta_i - [w_{ii} \\ &\quad + \Sigma_{j\neq I}\left(a_{ijj} + a_{jij} + a_{jji}\right)x_j]\Delta x_i\end{aligned}$$

$$\Delta x_i = -\left[\sum_{j,k}(a_{ijk} + a_{jik} + a_{jki})x_j(t)x_k(t) + \Sigma_j(w_{ij} + w_{ji})x_j(t) - \theta_i\right] \Big/ \left[1 + w_{ii} + \Sigma_{j\neq i}\left(a_{ijj} + a_{jij} + a_{jji}\right)x_j(t)\right]$$

and $x_i(t+1) = \mathrm{h}(\Delta x_i + x_i(t)) = \mathrm{h}(-\Delta P/\Delta x_i + x_i(t))$, where h is the Heaviside function. From (2), we derive (3):

$$\begin{aligned}x_i(t+1) = \mathrm{h}(-&\left[\Sigma_{j,k}\left(a_{ijk} + a_{jik} + a_{jki}\right)x_j(t)x_k(t) + \Sigma_j\left(w_{ij} + w_{ji}\right)x_j(t) - \theta_i\right] \\ &/\left[1 + w_{ii} + \Sigma_{j\neq i}\left(a_{ijj} + a_{jij} + a_{jji}\right)x_j(t)\right] + x_i(t))\end{aligned}$$

Proposition 9 *Consider a global m-switch as a Boolean potential system where $A = 0$, the potential function $\boldsymbol{P}(x) = {}^t xWx + Bx$, with $w_{ii} = \theta_i$ and each sub-matrix of W on any subset J of indices in $\{1,\ldots,m\}$ is non positive. Then $\boldsymbol{P}$ decreases on the trajectories of the potential system defined by: $x_i(t+1) = h(-\Delta \mathbf{P}/\Delta x_i + x_i(t))$ for any updating mode of the dynamics (sequential, block sequential and parallel). This system is a getBren, whose stable fixed configurations correspond to the minima of $\boldsymbol{P}$.*

Proof It is easy to check that from (2): $\Delta \boldsymbol{P}/\Delta x_i = \sum_j \left(w_{ij} + w_{ji}\right)x_j - \theta_i + w_{ii}\Delta x_i$, and from (3) and $w_{ii} = \theta_i$:

$$\begin{aligned} x_i(t+1) &= \mathrm{h}(-\Delta \boldsymbol{P}/\Delta x_i + x_i(t)) \\ &= \mathrm{h}\left(-\left[\sum_j \left(w_{ij} + w_{ji}\right) x_j(t) - \theta_i \right] / [1 + w_{ii}] + x_i(t) \right) \\ &= \mathrm{h}\left(-\sum_{j \neq i} (w_{ij} + w_{ji}) x_j(t + \theta_i) \right) \end{aligned}$$

Then, for any updating mode: $\boldsymbol{P}(x(t+1)) - \boldsymbol{P}(x(t)) = -\sum_{i \in J} \Delta x_i^2 (1 + w_{ii}) + \sum_{(i,j \in J \mathrm{x} J)} w_{ij} \Delta x_i \Delta x_j \leq 0$, the result coming from the non-positivity of the sub-matrices of W (Cosnard and Goles 1997). □

The results obtained by Cosnard and Goles (1997) concern Liapunov functions, but certain can be extended to potential systems (cf. also Fogelman-soulié et al. 1982; Goles and Olivos 1980; Goles 1981, 1985, 1986; Goles et al. 1985). In the case of Liapunov functions, it could be interesting to see if the following conjecture holds: for a large class of Boolean systems, there exists a general theorem of decomposition of the flow (or discrete velocity) in a potential part and a Hamiltonian part; such a system would be the discrete equivalent of the mixed potential-Hamiltonian continuous systems like the Liénard differential equations (Demongeot et al. 2007).

The following simple example of the Fig. 9 shows how a simple energy function, the frustration, could be used for solving the above conjecture.

The network of size 5 represented on Fig. 8 is ruled by the Boolean dynamical transition functions $\wedge$ (AND), $\vee$ (inclusive OR) and $\oplus$ (exclusive OR or XOR, or modulo 2 addition), updated synchronously, and its dynamics shows two fixed configurations and one limit-cycle (represented in grey boxes on Fig. 9). The values of the Liapunov function decrease along the trajectories toward the attractors on which the function vanishes (as indicated in red on the trajectory graph of the Fig. 9).

We can define indeed on the network trajectories an energy function, *i.e.*, the decreasing Liapunov function L, for any configuration x(t), by the global dynamical frustration of order 4, denoted D_4:

$$\mathrm{L}(\mathrm{x}(\mathrm{t})) = \mathrm{D}_4(\mathrm{x}(\mathrm{t})) + \mathrm{D}_4(\mathrm{x}(\mathrm{t}+1)), \quad \text{with } \mathrm{D}_4(\mathrm{x}(\mathrm{t})) = \sum_{\mathrm{i}=1,\ldots,\mathrm{n}} \mathrm{D}_{4,\mathrm{i}}(\mathrm{x}(\mathrm{t})),$$

where $\mathrm{D}_{4,\mathrm{i}}$ is the local frustration equal to twice the kinetic energy, i.e., the square of the velocity of order 4 between $\mathrm{x_i(t)}$ and $\mathrm{x_i(t+4)}$:

$$\mathrm{D}_{4,\mathrm{i}}(\mathrm{x}(\mathrm{t})) = (\mathrm{x_i(t+4)} - \mathrm{x_i(t)})^2.$$

6.2 The Genetic Regulatory Network Controlling the Cell Cycle in Mammals

We will study now a simple example of sub-network extracted from the genetic regulatory network dedicated to an important function, the control of the cell cycle in mammals, in particular the control of the cell proliferation at the G1/S transition check point (Kohn 1999). We give in Fig. 10 its interaction graph G and describe its

attractors in both parallel and sequential modes of updating, with a temperature T equal to 0. As an example, the evolutionary entropy E of the network pictured in Fig. 10, representing the core of the regulation of the cell cycle in mammals (Kohn 1999), may be calculated for sequential updating, under assumptions of Proposition 4., using Eq. (2) and calculations of ABRS's: $E = m\text{Log}2 - E_{\text{attractor}} = 12\times 0.69 - 0.86 = 7.5$. E increases from sequential to parallel updating (from 7.5 to 8.3), showing that the synchronous mode increases the robustness of the network. E decreases when boundary constraints grow (Fig. 10), and inhibition by miRNA 159 contributes to decrease the robustness of the genetic network, as in the case of the local m-switch (see Sect. 3), by increasing $E_{a\text{ttractor}}$ (from 0.024 to 0.654). More generally, an important inhibition noise from miRNAs (Fig. 10) can cause a diminution or in opposite a multiplication of the attractor number, by breaking a couple of circuits or in opposite a circuit, hence leading to an increase or a decrease of $E_{\text{attractor}}$, a decrease or an increase of E and, then, a diminution or an augmentation of the robustness: in the present case, miRNA 159 breaks the couple done by a couple of positive circuits of size 3 and 4, leaving only a circuit of size 4, which corresponds to an augmentation of the attractor number from 2 to 6, in the parallel case (see Table from (Demongeot et al. 2012e) on Fig. 10).

The control network of Cdk2 (red on Fig. 10) shows the complex inhibition role played by several miRNAs (miRs 21, 372 & 373) and by ubiquitous protein p53 in the expression of Cdk2, an important protein controlling the dynamics of G1/S transition in cell cycle. If the inhibition of Rbp by Cdk2 is sufficiently important, then the break of the order 4 positive circuit in the couple of tangent circuits leaves an order 3 positive circuit, leading to an augmentation of the attractor number from 2 to 4, in the parallel case (Fig. 10). Let us consider now a hybrid network, mixing neural and genetic regulatory sub-networks: the top of the Fig. 11 exhibits the neural part of the Engrailed network, under the control of the gene Elk whose expressed protein is necessary for the correct functioning of the K^+ ion-channels in excitable cells. Indeed, Elk controls positively the ability of CA3 cells (denoted Y_1 and Y_2 on the interaction graph of Fig. 11) to express their negative feedback upon the CA1 cells (denoted X_1 and X_2 on Fig. 11) inside the hippocampus (Tonnelier et al. 1999).

The Elk gene is activated by the gene Erk depending positively on the gene MEK, an intermediary gene in the pathway from the gene Egfr to the gene MAPK. In the centre of the Engrailed network on Fig. 11, the genes GATA-6 and c-Myc are regulating the proliferation with both positive and negative actions on the proliferation through the E2F expression in the mitotic regulatory box (Kohn 1999), as well as on the apoptosis through p53 (Hermeking and Eick 1994; Sheaff et al. 1997; Pelengaris et al. 2002; Ivanov and Hei 2005; Adhikary and Eilers 2005; He et al. 2007; Demongeot 2009; Suzuki and Miyazono 2010; Demongeot et al. 2013d). The triple action (accelerate, stop and slow down cell cycle) on the proliferation process is exerted negatively by the gene GATA-6, which is inhibited 1 time out of 2 by MAPK, and successively positively and negatively by the gene c-MyC which is activated 1 time out of 2 by Erk. The limit-cycle of order 4 brought by the negative circuit of size 2 MKP/Erk (cf. Table on Fig. 11 bottom right) leads the genes MKP, Erk, MAPK, Engrailed, GATA-6, c-MyC, p53, miRNA34, Cdk2,

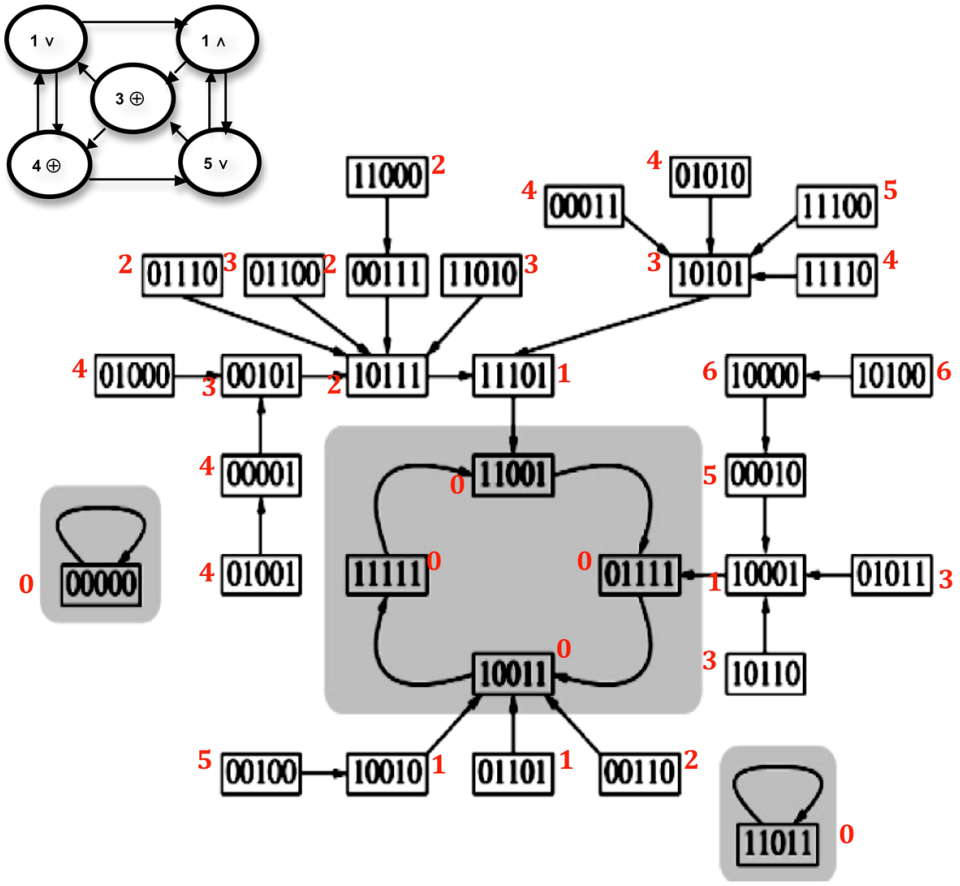


Fig. 9 Network of size 5 with its Boolean transition functions (*top left*) updated synchronously, and its trajectories (*bottom right*), with 2 fixed configurations and one limit-cycle represented in *grey boxes*. Values of the Liapunov function L, which decreases along trajectories toward attractors on which L vanishes, are indicated in red on trajectory graph. (Color figure online)

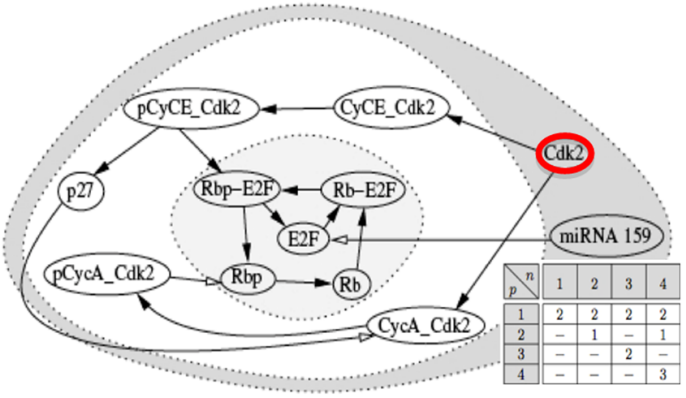
E2F, RAS and caspases to enter in the limit-cycle: (000000110110, 011000010010, 111101000011, 100111100011).

Then, the E2F/proliferation box is activated 1 time out of 4 and the apoptosis box 1 time out of 2, such as we observe the following cyclic behaviour: 4 cells, 4 cells, 3 cells, 2 cells, 4 cells,… This dynamical behaviour allows the exponential growth of proliferation to be compensated in a tissue by the linear decay of apoptosis, 2 daughter cells replacing 2 dead cells during one period of order 4 of the limit-cycle, hence ensuring the conservation of the tissue volume and function. Any disequilibrium of this balance leads either to a tumour growth or to a tissue rarefaction. Finally, concerning network N of Fig. 11, we exploit the conclusion of the Proposition 5. We have, if $Q_+(N)$ (resp. $Q_-(N)$) is equal to the number of positive (resp. negative) edges in interaction graph G of N:

$$Q_+(N) = 39, Q_-(N) = 11, F(00\ldots0) = Q_+(M) \text{ and } F(11\ldots1) = Q_-(M),$$

all other values of $F(x)$ falling inside the interval $[0, q]$, where $q = \sup(Q_-(M),Q_+(M))$. If we consider that the distribution of the values of $F(x)$ is uniform over this interval, then we get: $\mathrm{Var}U = \mathrm{Var}F = q^2/2\text{-}(q/2)^2 = q^2/4$, and here $q = 39$. Then, for the network of the Fig. 11 where all of the weights are

miRNA 159=0	Sequential updating		Parallel updating	
Nature	Attractor	ABRS	Attractor	ABRS
Fixed point 1	000000000000	6.25%	000000000000	0.5%
Fixed point 2	000000011111	56.25%	000000011111	99.5%
Limit cycle 1	000000001000 000000010111	37.5%	None	



p \ n	1	2	3	4
1	2	2	2	2
2	–	1	–	1
3	–	–	2	–
4	–	–	–	3

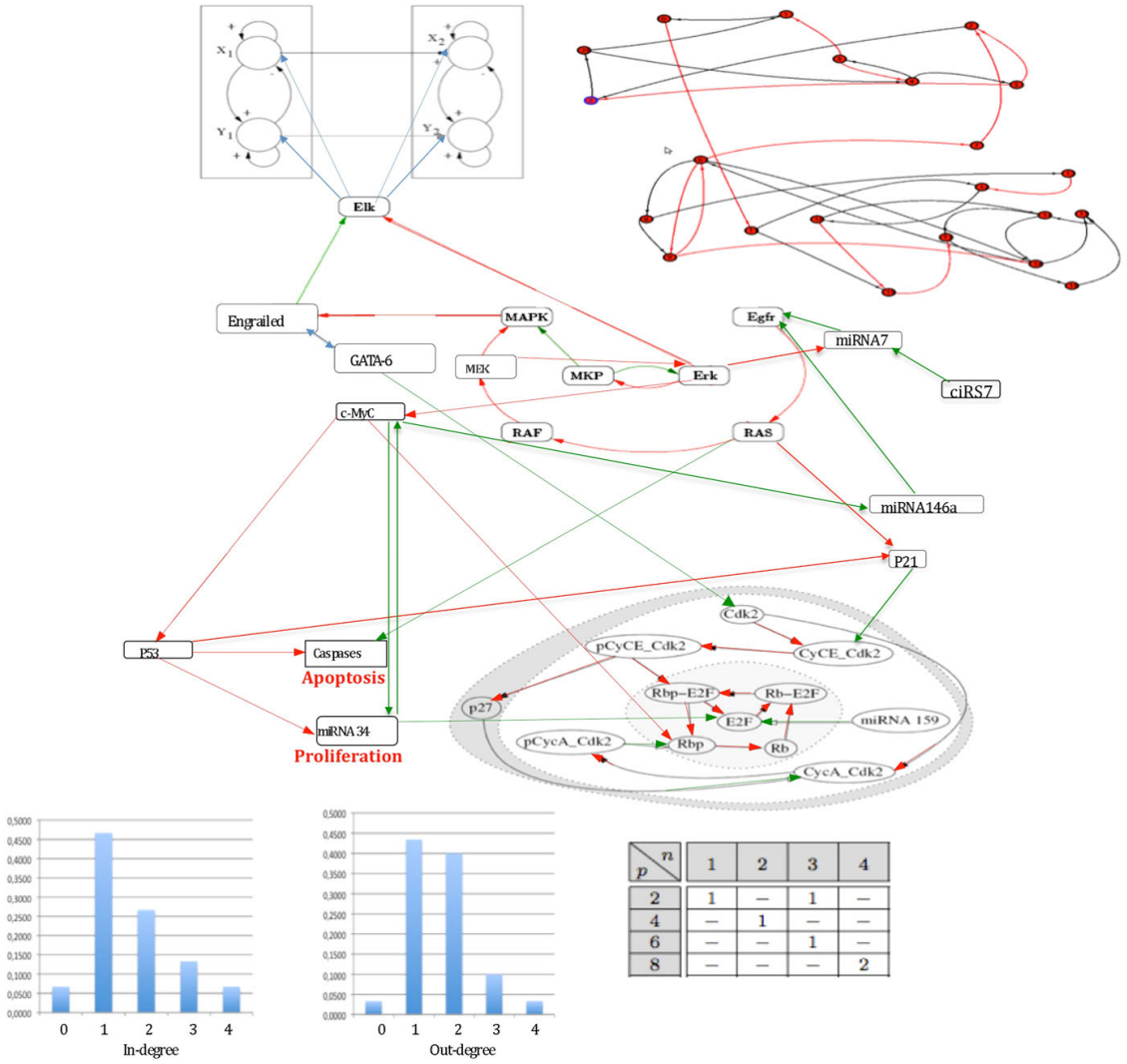
miRNA159=1 Parallel updating

Nature	Attractor	ABRS
Fixed point 1	000010000000	0.12%
Fixed point 2	000010011011	0.54%
Limit cycle 1	000010001010 000010010001	10.6%
Limit cycle 2	000010010011 000010001011 000010011001 000010011010	26.6%
Limit cycle 3	000010010010 000010000011 000010001001 000010011000	32.4%
Limit cycle 4	000010000010 000010000001 000010001000 000010010000	23.7%

Fig. 10 *Top* Attractors and attraction basins relative sizes (ABRS) of the cell cycle network, when miRNA 159 is fixed to 0, in case of sequential and parallel schedules, with $T = 0$. The nodes are ordered as follows: p27, Cdk2, pCyCE_Cdk2, CyCE_Cdk2, miRNA 159, pCycA_Cdk2, CycA_Cdk2, Rbp-E2F, Rb-E2F, E2F, Rbp and Rb. *Top right* Attractors and attraction basins relative sizes (ABRS), when miRNA 159 is fixed to 1, in case of parallel schedule, with $T = 0$. *Bottom left* Interaction signed digraph modelling the genetic regulation network of the cell cycle in mammals (Kohn 1999). Black (resp. *white*) *arrows* represent activations (resp. inhibitions). *Table bottom middle* (Demongeot et al. 2012e) indicates attractor number of period p observed for a positive circuit of length n, e.g., 4, if E2F circuit is inhibited by miRNA 159. (Color figure online)

supposed to have the same absolute value 1, we get: $\mathrm{Var}U = 39^2/4 \approx 380$ and the robustness can be quantified by this value 380. A systematic calculation of $E_{attractor}$ allows to quantify the robustness of the networks ensuring dedicated regulatory functions in different species: for example, the increase of the number of sources in up-trees converging on a conserved (among species during evolution) sub-graph of a genetic network, e.g., on the core regulating the cell cycle in *C. elegans*, *D. melanogaster* and mammals (cf. Fig. 12), causes a decrease of its attractor number, hence a decrease of $E_{attractor}$, hence an augmentation of the robustness.

Then, the progressive appearance during the evolution of many upstream controllers of the cell cycle network, as mammals microRNAs inhibiting p27 (Sheaff et al. 1997) and p53 (cf. Fig. 11 Bottom left, He et al. (2007) and Suzuki and Miyazono (2010)) provides a robust regulation (Dimova and Dyson 2005; Moriya et al. 2006; Ben Amor et al. 2008) of Rb-E2F, ensuring integrity of a crucial cellular function: cell proliferation at G1/S transition. A last example of the influence of the microRNAs on the robustness of genetic regulatory networks (Demongeot et al. Demongeot et al. 2009b, 2010, 2011b) lies in the appearance in the not-coding genome of mammals of inhibitions by microRNAs like miR-17-5p-20a-106a on the gene RUNX1 (involved, in human myeloid cells, in the control of TCRα recombination in immunology, see Baum et al. (2004)) or in the control by several



p \ n	1	2	3	4
2	1	–	1	–
4	–	1	–	–
6	–	–	1	–
8	–	–	–	2

Fig. 11 *Top* Engrailed hybrid network with multiple connections to apoptosis and proliferation control boxes, and to the hippocampus neural networks. Activation (resp. inhibition) *arrows* are drawn in *red* (resp. *green*). Black arrows in neural networks correspond to neural activations and inhibitions, and *blue arrows* correspond to activation links toward other functions than the strict Engrailed network. *Top right* reduced interaction graph G of the Engrailed network simulated in Fig. 13, with inhibitions in *red* and activations in *black*. *Bottom left* histograms of the in-degree and out-degree of the 50 interactions of the Engrailed network. *Bottom right*: Table (from Demongeot et al. 2012e) giving the number of the attractors of length p for the isolated negative circuits of length n. (Color figure online)

microRNAs (miR-150, miR-155 and let-7 family) of the T cell fate determination (Xiao et al. 2007; O'Connell et al. 2010; Almanza et al. 2010).

The reduced network of Fig. 11 has an attractor entropy $E_{\text{attractor}} = 0.67$ and by adding the circular RNA ciRS 7 (Demongeot et al. 2014) this entropy increases, because the negative circuits tangent at the gene Erk, which are three of respective size 2, 6 and 7 and have 1 attractor in parallel updating mode (Demongeot et al. 2012e), are replaced by two of respective size 2 and 7 with 2 attractors, contributing

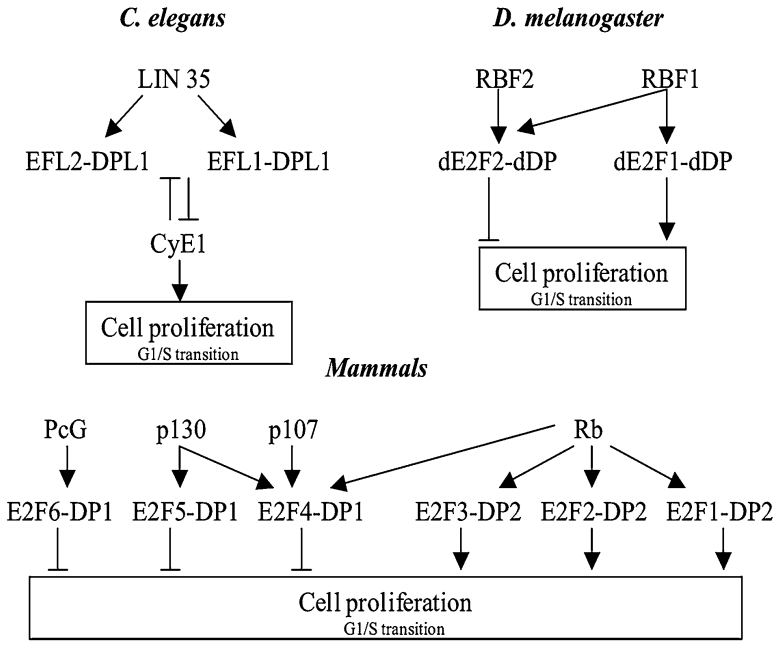


Fig. 12 Evolution of the structure of control of the cell proliferation box (cf. Fig. 9 and Kohn (1999)) for different species, *Cœnorhabditis elegans, Drosophila melanogaster* and mammals

to decrease the robustness of the network. A systematic study of microRNA inhibition by the circular RNAs, called "microRNA sponges" in (Hansen et al. 2013) (cf. also Memczak et al. 2013), is necessary to make more precise their stabilizing role (see also (Denzler et al. 2014; Broderick and Zamore 2014) (Fig. 13).

7 Perspectives and Conclusion

We have proposed in the present paper several examples of practical usage of the complexity and stability parameter E (evolutionary entropy) and of the parameter $E_{attractor}$ (attractor entropy), by applying the theoretical results to specific genetic networks, as those regulating feather morphogenesis in chicken and cell cycle in mammals. More generally, the quantification of the complexity, stability and robustness in biological regulatory networks allows us to explain in which way they have been adapted during the evolution, e.g., by exploiting the opportunities offered by the increasing number of small RNAs as translation factors inhibiting connected components in particular circuits of the getBrens resulting in an increase of the efficiency of their own regulation mechanisms on all the sides through which the biological regulation networks (neuronal, genetic, metabolic, physiologic and social) improved their efficiency from a long lasting evolution in the framework of an adaptive selection, by acquiring for a given complexity a triple optimality (Lesne 2008; Glade et al. 2011; Demongeot et al. 2013a, b) in terms of:

- robustness against parametric structural perturbations affecting the interaction weights, the external field and/or the stochasticity parameter, hence changing the

N°	Name of the gene	Fixed Point 1	Fixed Point 2	Limit Cycle	
1	Engrailed/GATA-6	0	0	0	0
2	MAPK	0	0	0	0
3	Egfr	0	0	0	0
4	MKP	0	0	0	0
5	Erk	0	0	0	0
6	MiRNA7	0	0	0	0
7	RAS/RAF/MEK	0	0	0	0
8	c-MyC	0	1	0	1
9	MiRNA146a	0	0	0	0
10	p53	0	1	1	0
11	miRNA 34	0	0	0	1
12	p27/pCyCE-Cdk2	0	0	0	0
13	Cdk2	0	0	0	0
14	CyCE-Cdk2	0	0	0	0
15	p21	0	1	0	1
16	Rbp-E2F	0	1	1	1
17	Rb-E2F	0	1	1	1
18	E2F	0	1	1	1
19	Rbp	0	1	1	1
20	Rb	0	1	1	1
21	CycA_Cdk2	0	0	0	0

Fixed Point 1: 000000000000000000000
Attraction Basin Relative Size(ABRS): $12032/2^{21} = 5‰$
Fixed Point 2:
000000010100001111111062
Attraction Basin Relative Size(ABRS): $1311488/2^{21} = 62.5\%$
Limit Cycle:
000000000100000111110 ⇒ 000000010010001111110
Attraction Basin Relative Size(ABRS): $773632/2^{21} = 37\%$

Fig. 13 *Left* list of the genes involved in the reduced network of Fig. 11 with indication of their state in the 3 attractors of the network. *Right* attraction basins size of the attractors in parallel updating mode of the reduced network of the Fig. 11

invariant measure μ of the Markov matrix M used for quantifying the state transition in the stochastic case into a perturbed measure μ_ε (cf. Fig. 14)

- stability with respect to a dynamical change of initial condition.
- resilience to environmental modifications of boundary conditions or updating rules.

For studying the robustness, we have defined a new quantity, the attractor entropy anti-correlated to a complexity parameter, the evolutionary entropy entropy (Demetrius et al. 2004) previously proposed for quantifying the stability (correlated in some cases to the logarithm of the inverse of the sub-dominant eigenvalue of the transition matrix). For the resilience, future studies are needed to define indices measuring the effect of boundary or updating perturbations (Demongeot and Sené 2008; Demongeot et al. 2008b, 2013a). The influence of the mode of updating (sequential, parallel or block-sequential) and of the nature of the dynamical state transition (potential, Hamiltonian or mixed) on the network robustness has to be carefully investigated in these future studies, in particular to take into account both the chromatin clock and the slow evolution dynamics of the interactions (like the progressive appearance of microRNAs inhibitions influencing in general the robustness), analogue to the slow Hebbian dynamics of the synaptic weights in neural networks. The study of the relationship between complexity and stability notions is old: Cohen and Newman (1985) have been the first to prove counter-examples to the mathematical statement (Hastings 1982), claiming that the stability of ecological systems was positively correlated with their connectivity. Here we have also shown some examples where this statement can be false (e.g., the evolution of the proliferation network), where the robustness is contrarily decreasing with connectivity. The general problem of the relationships between complexity, stability and robustness remains open and requires many further

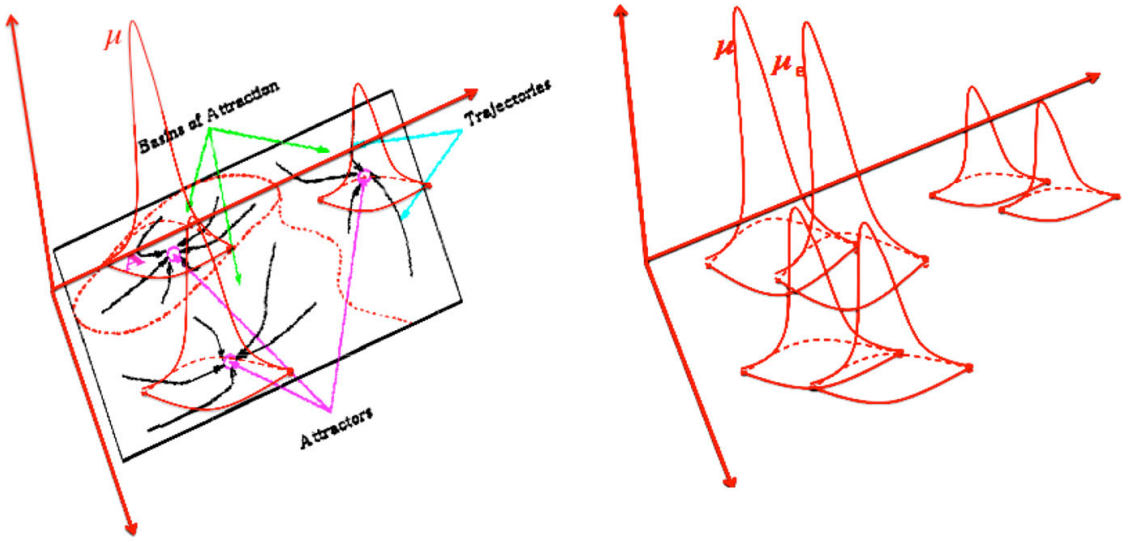


Fig. 14 *Left* Visualisation of the invariant measure μ of the Markov matrix *M* in the stochastic case over the basins of stability of the deterministic case (for which the temperature T = 0); *Right* Perturbation μ_ε of the invariant measure μ. (Color figure online)

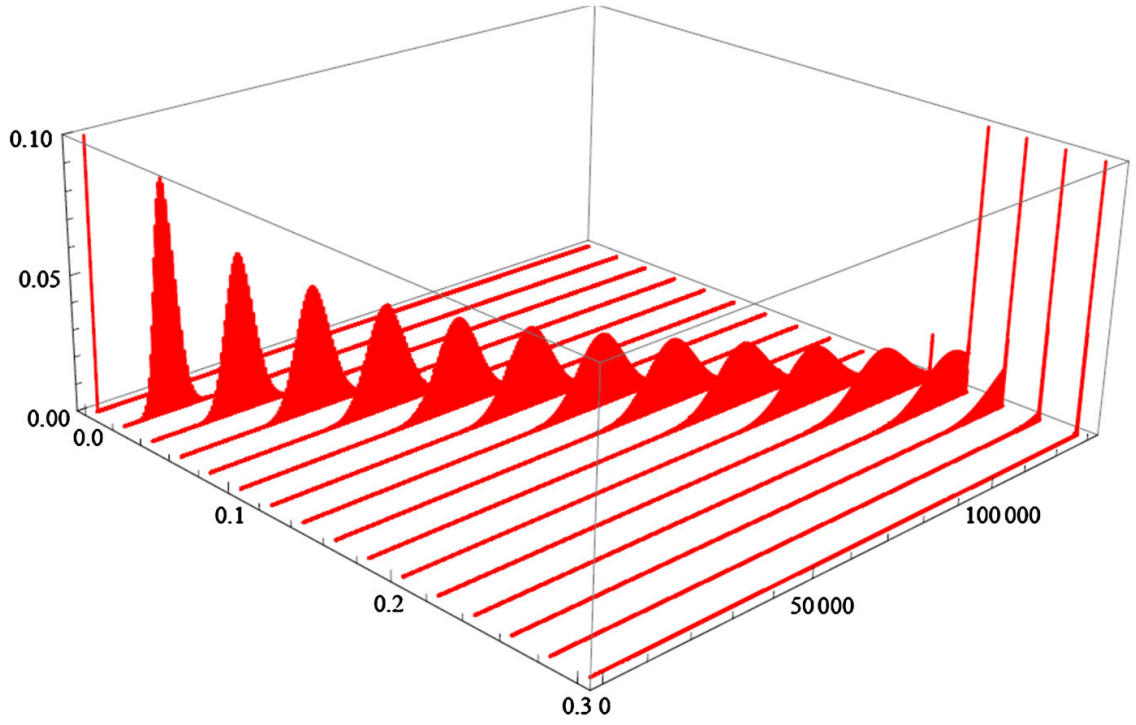


Fig. 15 3D visualisation [after Demongeot et al. (2013c)] of the evolution of a Markov chain, from an initial measure which is a Dirac on state 0 (all genes of the network of Fig. 11 at state 0) until a final measure, which is a Dirac on the final class of the chain, made of a unique last element (the second fixed configuration of the network of Fig. 11)

studies, the corresponding research being still active, as shown by recent works (Thébault and Fontaine 2010; Liu and Wang 2011; Allesina and Tang 2012).

Eventually, to ensure a useful restitution of information concerning the evolution of the getBrens, we have to develop new user-friendly tools representing in

particular the evolution of the probability measures on the network, for identifying their most frequent states experimentally observable (cf. Fig. 15 and Demongeot et al. 2013c) and giving information about the duration of the transient stage, the proximity to the asymptotic behaviour, attractor or confiner (Cosnard et al. 1988; Demongeot 1985, 1987; Demongeot and Fricot 1986; Demongeot et al. 2008a; Ribeiro et al. 2009, 2012; Demongeot et al. 2013c), and the complexity of interaction architecture, e.g., existence of circuits, nonlinear interactions or external fields (Demongeot and Sené 2011, 2012; Antonopoulos et al. 2013; Wainrib and Touboul 2013; Bandiera et al. 2013; Demongeot and Sené 2014).

Acknowledgments We thank VHP NoE (EC) and MEC Grant from CONICYT (Chile) for financially aiding our research.

References

Abbas L, Demongeot J, Glade N (2009) Synchrony in reaction-diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves. Phil Trans Royal Soc A 367:4829–4862

Adhikary S, Eilers M (2005) Transcriptional regulation and transformation by Myc proteins. Nat Rev Mol Cell Biol 6:635–645

Allesina S, Tang S (2012) Stability criteria for complex ecosystems. Nature 483:205–208

Almanza G, Fernandez A, Volinia S, Cortez-Gonzalez X, Croce CM, Zanetti M (2010) Selected MicroRNAs Define cell fate determination of murine central memory CD8 T cells. PLoS ONE 5:e11243

Antonopoulos C, Basios V, Demongeot J, Nardone P, Thomas R (2013) Linear and nonlinear arabesques: a study of closed chains of negative 2-element circuits. Int J Bifurc Chaos 23:30033

Aracena J, Demongeot J, Fanchon E, Montalva M (2013) On the number of different dynamics in Boolean networks with deterministic update schedules. Math Biosci 242:188–194

Audin M, Babbitt DG (2008) Hamiltonian systems and their integrability. AMS, Providence

Baconnier PF, Pachot P, Demongeot J (1993) An attempt to generalize the control coefficient concept. J Biol Syst 1:335–347

Bandiera S, Matégot R, Demongeot J, Henrion-Caude A (2013) MitomiRs: delineating the intracellular localization of microRNAs at mitochondria. Free Radic Biol Med 64:12–19

Baum TP, Pasqual N, Thuderoz F, Hierle V, Chaume D, Lefranc MP, Jouvin-Marche E, Marche P, Demongeot J (2004) IMGT/GeneInfo: enhancing V(D)J recombination database accessibility. Nucleic Acids Res 32:51–54

Ben Amor H, Demongeot J, Elena A, Sené S (2008) Structural sensitivity of neural and genetic networks. L N Comp Sci 5317:973–986

Bowen R (1978) On axiom a diffeomorphisms. Reg. Conf Series Math 35, AMS, Providence.

Broderick JA, Zamore PD (2014) Competitive endogenous RNAs cannot alter microRNA function in vivo. Mol Cell 54:711–713

Cinquin O, Demongeot J (2002a) Positive and negative feedback: striking a balance between necessary antagonists. J Theor Biol 216:229–241

Cinquin O, Demongeot J (2002b) Positive and negative feedback: mending the ways of sloppy systems. C R Biol 325:1085–1095

Cohen JE, Newman CM (1985) When will a large complex system be stable? J Theor Biol 113:153–156

Cosnard M, Demongeot J (1985a) Attracteurs: une approche déterministe. C R Acad Sci Maths Série I 300:551–556

Cosnard M, Demongeot J (1985b) On the definitions of attractors. Lect Notes Math 1163:23–31

Cosnard M, Goles E (1997) Discrete states neural networks and energies. Neural Netw 10:327–334

Cosnard M, Goles E, Moumida D (1988) Bifurcation structure of a discrete neuronal equation. Discret Appl Math 21:21–34

Demetrius L (1978) Adaptive value, entropy and survivorship curves. Nature 275:213–214

Demetrius L, Ziehe M (2007) Darwinian fitness. Theor Popul Biol 72:323–345

Demetrius L, Gundlach M, Ochs G (2004) Complexity & demographic stability in population models. Theor Popul Biol 65:211–225
Demongeot J (1985) Random automata and random fields. In: Demongeot J et al (eds) Dynamical systems and cellular automata. Academic Press, New York, pp 99–110
Demongeot J (1987) Random automata. In: Soulie Fogelman F (ed) Automata networks in computer science: theory & applications. Princeton University Press, Princeton, pp 47–57
Demongeot J (2009) Biological boundaries and biological age. Acta Biotheor 57:397–419
Demongeot J, Demetrius L (1989) La dérive démographique et la sélection naturelle. Population 2:231–248
Demongeot J, Fricot J (1986) Random fields and renewal potentials. Springer NATO ASI Series in Systems & Computer Science F 20:71–84
Demongeot J, Jacob C (1989) Confineurs: une approche stochastique. C R Acad Sci Série I 309:699–702
Demongeot J, Jacob C (1990) Confiners, stochastic equivalents of attractors. In: Tautu P (ed) Stochastic modelling in biology. Relevant mathematical concepts and recent applications. World Scientific Publication, Singapore, pp 309–327
Demongeot J, Sené S (2008) Asymptotic behaviour and phase transition in regulatory networks. II Simulations. Neural Netw 21:971–979
Demongeot J, Sené S (2011) The singular power of the environment on nonlinear Hopfield networks. In: Proceedings CMSB'11. ACM Proceedings, New York, pp 55–64
Demongeot J, Sené S (2014) Nonlinear threshold Boolean automata networks and phase transitions. arxiv.org/pdf/1011.4675
Demongeot J, Waku J (2012a) Robustness in biological regulatory networks. I Mathematical approach. C R Math 350:221–224
Demongeot J, Waku J (2012b) Robustness in biological regulatory networks. II Application to genetic threshold Boolean random regulatory networks (getBren). C R Math 350:225–228
Demongeot J, Waku J (2012c) Robustness in biological regulatory networks. III Application to genetic networks controlling the morphogenesis. C R Math 350:289–292
Demongeot J, Waku J (2012d) Robustness in biological regulatory networks. IV Application to genetic networks controlling the cell cycle. C R Math 350:293–298
Demongeot J, Aracena J, Thuderoz F, Baum TP, Cohen O (2003) Genetic regulation networks: circuits, regulons and attractors. C R Biol 326:171–188
Demongeot J, Thellier M, Thomas R (2006a) Storage and recall of environmental signals in a plant: modelling by use of a differential (continuous) formulation. C R Biol 329:971–978
Demongeot J, Elena A, Weil G (2006b) Potential-Hamiltonian decomposition of cellular automata. Application to degeneracy of genetic code and cyclic codes III. C R Biol 329:953–962
Demongeot J, Glade N, Forest L (2007) Liénard systems and potential-Hamiltonian decomposition. I methodology. C R Math 344:121–126
Demongeot J, Elena A, Sené S (2008a) Robustness in neural and genetic networks. Acta Biotheor 56:27–49
Demongeot J, Jezequel C, Sené S (2008b) Asymptotic behaviour and phase transition in regulatory networks. I theoretical results. Neural Networks 21:962–970
Demongeot J, Ben Amor H, Gillois P, Noual M, Sené S (2009a) Robustness of regulatory networks. A generic approach with applications at different levels: physiologic, metabolic and genetic. Int J Mol Sci 10:4437–4473
Demongeot J, Drouet E, Moreira A, Rechoum Y, Sené S (2009b) Micro-RNAs: viral genome and robustness of the genes expression in host. Phil Trans Royal Soc A 367:4941–4965
Demongeot J, Goles E, Morvan M, Noual M, Sené S (2010) Attraction basins as gauges of environmental robustness in biological complex systems. PLoS ONE 5:e11793
Demongeot J, Elena A, Noual M, Sené S (2011a) Random Boolean networks and attractors of their intersecting circuits. In: Barolli L et al (eds) Proceedings AINA'11. IEEE Press, New York, pp 483–487
Demongeot J, Henrion-Caude A, Lontos A, Promayon E (2011b) General architecture of a genetic regulation network. Applications to embryologic and immunologic control. In: Lenaerts T et al (eds) Proceedings ECAL'11. MIT Press, Cambridge, MA, pp 1–8
Demongeot J, Elena A, Noual M, Sené S, Thuderoz F (2011c) "Immunetworks", attractors & intersecting circuits. J Theor Biol 280:19–33
Demongeot J, Noual M, Sené S (2012) Combinatorics of Boolean automata circuits dynamics. Discret Appl Math 160:398–415

Demongeot J, Cohen O, Doncescu A, Henrion-Caude A (2013a) MitomiRrs and energetic regulation. Proceedings AINA'13. IEEE Press, New York, pp 1501–1508

Demongeot J, Hazgui H, Vuillerme N (2013b) MicroRNAs: unspecific inhibitory regulation in immunologic control and in mitochondrial respiration. In: Barolli L et al (eds) Proceedings AINA'13. IEEE Press, New York, pp 1509–1516

Demongeot J, Vallejos R, Barria M, Taramasco C (2013c) Information design of biological networks: application to genetic, immunologic, metabolic and social network. In: Barolli L et al (eds) Proceedings AINA'13. IEEE Press, New York, pp 1533–1540

Demongeot J, Hazgui H, Bandiera S, Cohen O, Henrion-Caude A (2013d) MitomiRs, ChloromiRs and general modelling of the microRNA inhibition. Acta Biotheor 61:367–383

Demongeot J, Cohen O, Henrion-caude A (2013e) MicroRNAs and robustness in biological regulatory networks. A generic approach with applications at different levels: physiologic, metabolic, and genetic. Springer Series in Biophysics, vol 16. pp 63–114

Demongeot J, Hazgui H, Escoffier J, Arnoult C (2014) Inhibitory regulation by microRNAs and circular RNAs. In: Romero R et al (eds) Medicon'13 IFBME Proceedings 41. Springer, New York, pp 722–725

Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M (2014) Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. Mol Cell 54:766–776

Dimova DK, Dyson NJ (2005) The E2F transcriptional network: old acquaintances with new faces. Oncogene 24:2810–2826

Donsker MD, Varadhan SRS (1975) Asymptotic evaluation of certain Markov process expectations for large time. Commun Pure Appl Math 28:1–47

Elena A, Demongeot J (2008) Interaction motifs in regulatory networks and structural robustness. In: Barolli L et al (eds) Proceedings ARES-CISIS'08. IEEE Press, New York, pp 682–686

Fogelman-Soulié F, Goles E, Weisbuch G (1982) Specific roles of the different Boolean mappings in random networks. Bull Math Biol 44:715–730

Freidlin MI, Wentzell AD (1984) Random perturbations of dynamical systems. Springer, New York

Glade N, Elena A, Fanchon E, Demongeot J, Ben Amor H (2011) Determination, optimization and taxonomy of regulatory networks. The example of A thaliana flower morphogenesis. In: Barolli L et al (eds) Proceedings AINA'11. IEEE Press, New York, pp 488–494

Goldstein S (1981) Entropy increase in dynamical systems. Israel J Math 38:241–256

Goldstein S, Penrose O (1981) A nonequilibrium entropy for dynamical systems. J Stat Phys 24:325–343

Goles E (1981). Sequential iteration of threshold functions. Springer Series in Synergetics 9:64–70

Goles E (1985) Dynamics of positive automata networks. Theoretical Comp Sci 41:19–32

Goles E (1986) Antisymmetric neural networks. Discret Appl Math 13:97–100

Goles E, Olivos J (1980) Comportement itératif des fonctions à multiseuil. Inf Control 45:300–313

Goles E, Fogelman-Soulie F, Pellegrin D (1985) Decreasing energy functions as a tool for studying threshold networks. Discret Appl Math 12:261–277

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495:384–388

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402:47–52

Hastings HM (1982) The May-Wigner stability theorem for connected matrices. Bull Am Math Soc 7:387–388

He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, Chen C, Lowe SW, Cleary MA, Hannon GJ (2007) A microRNA component of the p53 tumour suppressor network. Nature 447:1130–1134

Hermeking H, Eick D (1994) Mediation of c-Myc-induced apoptosis by p53. Science 265:2091–2093

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 79:2554–2558

Ivanov VN, Hei TK (2005) Combined treatment with EGFR inhibitors and arsenite upregulated apoptosis in human EGFR-positive melanomas: a role of suppression of the PI3 K-AKT pathway. Oncogene 24:616–626

Jensen JH, Ellis DPW, Christensen MG, Jensen SH (2007) Evaluation of Distance Measures Between Gaussian Mixture Models of MFCCs. In: Dixon S et al (eds) Proceedings ISMIR'07. Austrian Computer Society, Wien, pp 107–108

Kohn KW (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol Biol Cell 10:2703–2734

Krasnoselski MA (1968) Translation along trajectories of differential equations. AMS, Providence
Lesne A (2008) Robustness: confronting lessons from physics and biology. Biol Rev Camb Philos Soc 83:509–532
Liapunov AM (1992) The general problem of the stability of motion. Taylor & Francis, London
Liu M, Wang K (2011) Global stability of a nonlinear stochastic predator–prey system with Beddington–DeAngelis functional response. Commun Nonlinear Sci Numer Simul 16:1114–1121
Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495:333–338
Meyn SP (2008) Control techniques for complex networks. Cambridge University Press, New York
Moriya H, Shimizu-Yoshida Y, Kitano H (2006) In Vivo Robustness Analysis of Cell Division Cycle Genes in Saccharomyces cerevisiae. PLoS Genet 2:1034–1045
O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D (2010) Physiological and pathological roles for microRNAs in the immune system. Nat Rev Immunol 10:111–122
Pelengaris S, Khan M, Evan G (2002) c-MYC: more than just a matter of life and death. Nat Rev Cancer 2:764–776
Reder C (1988) Metabolic control theory: a structural approach. J Theor Biol 135:175–201
Ribeiro CS, Rosseti L, Vallejos R (2009) On the use of run time distributions to evaluate and compare stochastic local search algorithms. Lect Notes Comput Sci 5752:16–30
Ribeiro CS, Rosseti L, Vallejos R (2012) Exploiting run time distributions to compare sequential and parallel stochastic local search algorithms. J Glob Optim 54:405–429
Schey HM (1992) Div, Grad, Curl, and All That. Norton, New York
Sheaff RJ, Groudine M, Gordon M, Roberts JM, Clurman BE (1997) Cyclin E-CDK2 is a regulator of p27Kip1. Genes & Dev 11:1464–1478
Shieh SF (2011) Eigenvalue estimates using the Kolmogorov-Sinaïentropy. Entropy 13:2036–2048
Suzuki HI, Miyazono K (2010) Dynamics of microRNA biogenesis: crosstalk between p53 network and microRNA processing pathway. J Mol Med 88:1085–1094
Thébault E, Fontaine C (2010) Stability of ecological communities the architecture of mutualistic & trophic networks. Science 329:853–856
Thom R (1975) Structural stability and morphogenesis. Benjamin, Reading
Tonnelier A, Meignen S, Bosch H, Demongeot J (1999) Synchronization and desynchronization of neural oscillators: comparison of two models. Neural Netw 12:1213–1228
Tuljapurkar S (1993) Entropy and convergence in dynamics and demography. J Math Biol 31:253–271
Waddington CH (1942) Canalization of development and the inheritance of acquired characters. Nature 150:563–565
Wainrib G, Touboul J (2013) Topological and dynamical complexity of random neural networks. Phys Rev Lett 110:118101
Weaver DC, Workman CT, Stormo GD (1999) Modeling regulatory networks with weight matrices. Pac Symp Biocomp 4:112–123
Wentzell AD, Freidlin MI (1970) On small random perturbations of dynamical systems. Rus Math Surv 25:1–55
Xiao C, Calado DP, Galler G, Thai TH, Patterson HC, Wang J, Rajewsky N, Bender TP, Rajewsky K (2007) MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 131:146–159

Entropy **2015**, *17*, 1-x manuscripts; doi:10.3390/e170x000x

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Entropy for genetic networks. Role of RNA effectors

Jacques Demongeot [1,*] and Hana Hazgui [1]

[1] Laboratory AGIM, Faculty of Medicine, University J. Fourier Grenoble, La Tronche, 38700, France; E-Mail: Hana.Hazgui@agim.eu

* Author to whom correspondence should be addressed; E-Mail: Jacques.Demongeot@agim.eu; Tel.: +33-476-637-153; Fax: +33-476-637-492.

Academic Editor:

Received: / Accepted: / Published:

Abstract: RNA molecules are often involved in the regulation of complex genetic networks as effectors, activators (small RNAs involved in transcription factors), inhibitors (microRNAs) or hybrids (circular RNAs). We give examples of such genetic networks and show that i) RNA-RNA or RNA-peptide interactions could be explained by the presence of RNA "relics" having played an important role during the evolution and survived in many genomes, their present probability distribution being quantified by the static entropy, and ii) the dynamical entropy related to the dynamics of these networks can be used to characterize their robustness.

Keywords: complex biological networks; small RNAs; microRNAs; circular RNAs; dynamical entropy; attractor entropy; robustness

PACS Codes: ….

---

1. Introduction

RiboNucleic Acid (RNA) is an important living molecule with the propensity to regulate many fundamental mechanisms of the cell. RNA molecules are indeed often involved in the regulation of many complex biological networks as effectors. They can be activators, like the small RNAs involved in transcription factors, inhibitors like microRNAs involved in translation, or hybrids like circular RNAs, which are inhibiting microRNAs. The present genomes are the result of a long evolution from the start of the life on the earth until the appearance of mammals

and human. We will show that RNA relics persist coming from the beginning of life. These relics are small RNA sequences having fulfilled an important role for transmitting the living information already at the first steps of the evolution and playing still now a notable function in the control of important genetic networks involved in numerous living systems. In this paper, we will give some examples of complex biological networks regulated by RNAs and show that:
i) present RNA-RNA or RNA-peptide interactions could be explained by the presence of RNA "relics" having played an important role during the evolution and having survived in many genomes, their present probability distribution being quantified by the static entropy
ii) the dynamical entropy related to the dynamics of the biological networks networks can be used to characterize the robustness of the RNA regulated systems.

We will describe the RNA relics in Section 2, then show in Section 3 their survival in present genomes and describe in Section 4 the role they play presently in the regulation of important genetic networks. Eventually, in Section 5, we will define the dynamical entropy and its role in the quantification of the network robustness.

2. RNA relics

Let us consider now the two RNA sequences called respectively AL (for Archetypical Loops) and AB (for Ancestral Bases):

5'-UGAAUGGUACUGCCAUUCAAGA-3' (AL)
5'-UGAAUGGUGCCAUUCAAGACUA-3' (AB).

Their main features have been previously described [1-7], hence we will describe only the most relevant for our argumentation about their survival until the present genomes.

The first small circular RNAs (“RNA rings”) when life emerged could be stabilized by a proto-membrane of amino acids determined by the affinities to their different codons, overlapping in a circular sequence [5]. A high diversity of codons in these rings corresponding to different amino acids, would increase the chance that the rings attract enough amino acids to support their stability. In this way, the optimal RNA structure turns out to be a ring which codes for all the different available amino acids, *i.e.*, having at least one codon in each of the synonymy classes of the genetic code.

At the same time, the ring should be as small as possible, since bigger rings will be more prone to accidental denaturation. This would lead to the determination of RNA rings of 20 nucleotides, which contain codons in each of the coding classes.

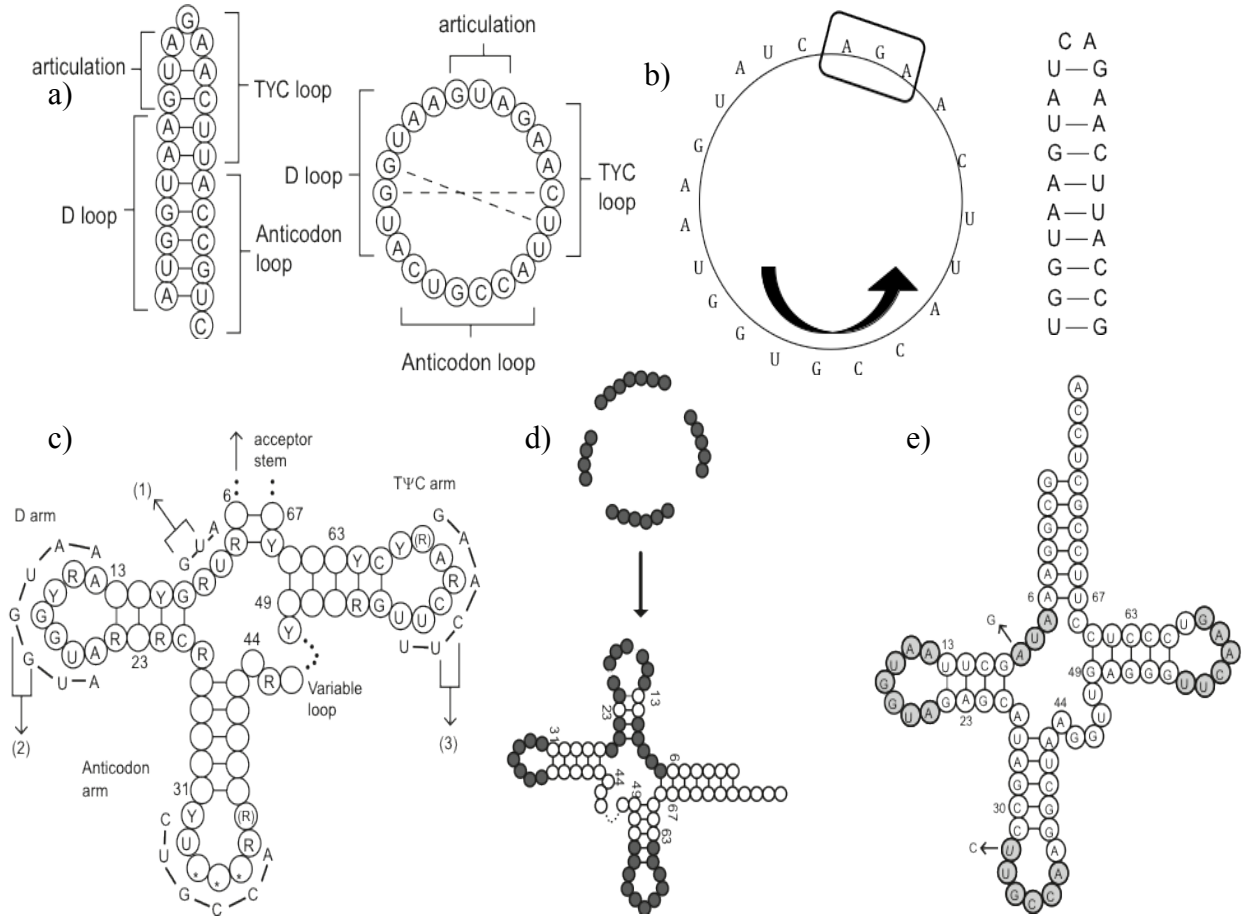


Figure 1. (**a**) The relic sequence AL in hairpin and circular forms. (**b**) The relic sequence AL in hairpin and circular form. (**c**) AL matching the loops of the archetypal tRNA. (**d**) Correspondence between the circular and the clover-leave structure. The ring is cut into segments aligned with the conserved domains of the tRNA (loops), respecting the 5'-3' orientation. (**e**) AL matching the loops of the $tRNA^{Gly}$ of *Arabidopsis thaliana*. Only positions 9 and 32 differ in AL, as shown.

However, two observations lead to study a different set of minimal rings. The first is combinatorial: exhaustive listing through backtracking algorithms shows that rings with the required property do not exist; hence, some latitude must be allowed. The second observation is based on the dynamics of the scenario: with structures being routinely denaturated, and new structures emerging from the pieces that float in the environment, there will be a selective advantage for those structures that are more likely to break in ways that facilitate a later reconstruction. Since the links inside peptides and RNA are likely to be stronger than the affinity-determined binding between them (based on weak forces like van der Waals or electromagnetic, following the stereochemical theory of the genetic code), a RNA-peptide ring structure is likely to be occasionally turned into a small strand of RNA and a peptide. The reconstruction of the same species of such a structure will be easier if it breaks preferentially in a certain point along the ring, *i.e.*, if it contains an unprotected spot, one without an amino acid/codon complex. This, according to our assumptions, is expected to happen if the ring contains a stop codon. It follows that we must look at minimal rings able to code for the 21

synonymy classes of the genetic code (including the stop class). The combinatorics shows that the variational problem satisfying two opposite criteria, *i.e.*, i) to have the minimal length and ii) to code for these 21 classes, has solutions only if the sequence begins with the start codon AUG. The set of corresponding solutions has a barycentre for the Hamming distance, called the AL ring. The AB ring is another solution whose hairpin form is more stable than the AL one (see Figure 1).

The RNA ring AL matches well with the sequence obtained from the succession, in their order in the cloverleaf structure, of the tRNA loops (D-loop, anti-codon loop and T-loop) [8-13], and that for many tRNAs (Figure 2) like the consensus tRNA loops proposed by B. Lewin [14] and could be considered as part of an ancient proto-ribosomal system, in which AL have played two roles: i) it attracted amino acids with weak bonds like the present tRNAs, and ii) favoured the creation of strong peptidic bonds between these amino acids like in the present ribosomes.

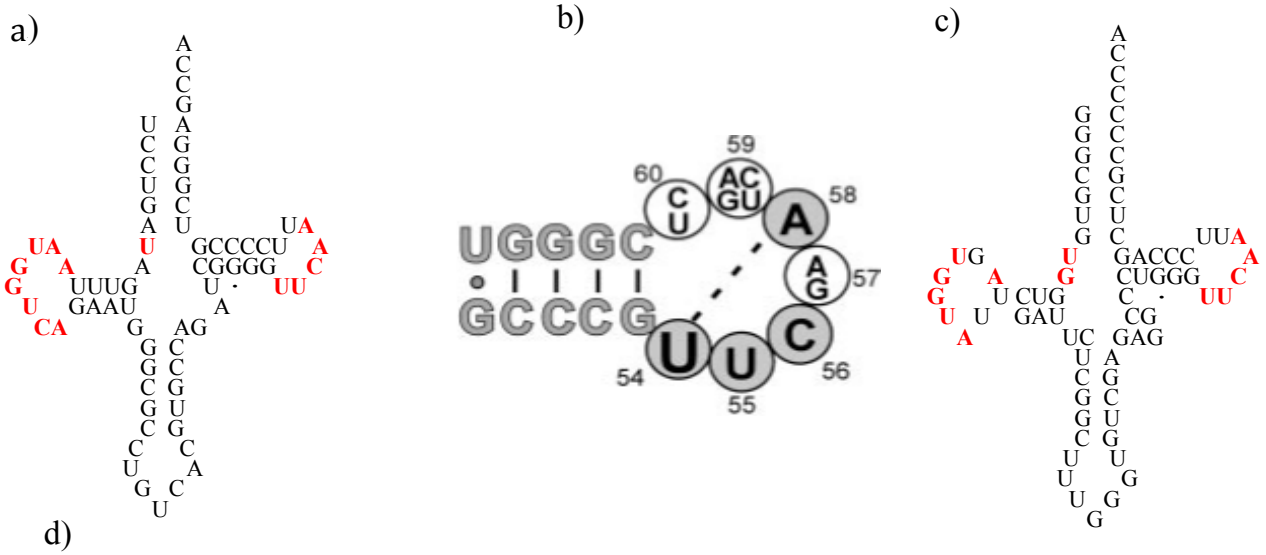


Figure 2. (**a**) *S. cerevisiae* $tRNA^{Asp}$ [10]. (**b**) T loop consensus sequence derived from sequences of 11 species of *T. thermophilus* tRNAs contained in the tRNA Compilation Data base [12,13]. (**c**) *S. cerevisiae* $tRNA^{Pr}$ [14]. (**d**) Consensus sequence proposed by B. Lewin for the tRNA loops [15]. The AL bases matching with Lewin's bases are in blue.

3. Static entropy and diversity of the nucleotides motifs

3.1. Nucleotide sequences from RNA and DNA worlds

In order to compare AL with nucleotide sequences coming from genetic databases, we computed the number W of occurrences of AL subsequences of length equal or greater than 5 (called 5-ALsq) in RNA or DNA sequences. To assess the W significance, we ran 1000 computations, in which each nucleotide sequence was randomized by shuffling its bases, but keeping base content and length. The mean number for these randomized cases was $\langle W_R \rangle$ =

91.665, with a standard deviation σ_R of 286 (cf. Table 1). By computing W for all RNA sequences, we obtain a total of 120.441 significantly greater than $<W_R>$. The number W of matches between AL and RNA sequences was indeed well above $<W_R>$ for any empirical distribution given on Table 1 [5].

Table 1. Subsequences of length 5 common between AL and RNA families. The type snRNA corresponds to the aggregation of snRNA types in Rfam database [16] (except for spliceosomal snRNAs listed apart) and contains almost only snoRNAs.

RNA type	S	W	$\langle W_R \rangle$	σ_R	$\sigma_R/\langle W_R \rangle$
tRNA conserved domains	2.77	3498	1262	32	0.025
Intron	1.46	65,119	44,599	202	0.005
snRNA (spliceosomal)	1.43	5421	3799	59	0.016
snRNA	1.39	5478	3941	59	0.015
rRNA	1.37	14,213	10,366	97	0.009
miRNA	1.33	716	540	22	0.041
tRNA	1.12	12,709	11,376	102	0.009
Ribozyme	1.11	2114	1902	41	0.022
sRNA	1.03	498	484	20	0.041
Cis-reg	0.97	9539	9809	93	0.009
Gene	0.96	4483	4678	65	0.014
Antisense	0.88	151	171	12	0.070
Whole Rfam	1.31	120,441	91,665	286	0.003

The frequencies of the histogram given on Figure 3 are consistent with the idea that AL fragments were the source material for the primitive RNAs, with higher probability for the fragments that were not disrupted in the AL hairpin form. The entropy of the distribution corresponding to the histogram given on Figure 3 is equal to $H_1 = 2.93176$.

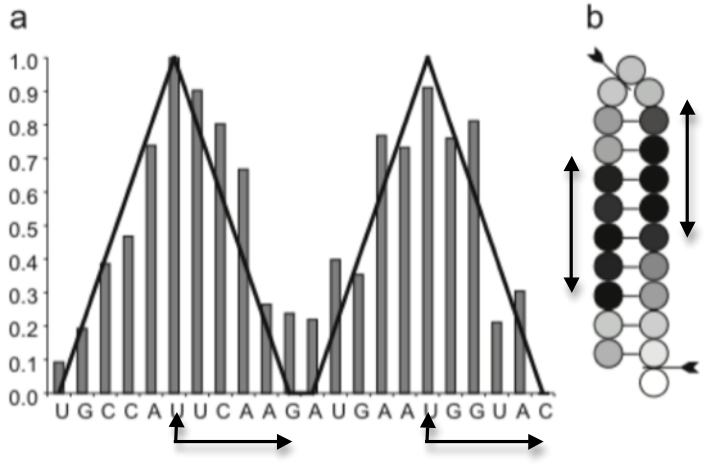


Figure 3. Frequencies in AL of the start nucleotide of the subsequences of length 5 (5-ALsq) matching with Rfam sequences. For better comparison, frequencies have been normalized and the arrows indicate the position of the most frequent 5-ALsqs. The triangle lines in (**a**) correspond to the distances of each base with respect to the two inter-base positions marked in the AL hairpin form (**b**), which shows the values graphed in (**a**) as shades of gray (with white and black representing the minimum and maximum values of this distance, respectively).

3.2. RNA sequences from human mitochondrial genome

Human mitochondria contain small RNAs including microRNA, piRNA, tRNA, rRNA, and RNA repeats. The sequencing generated 19.5 and 17.7 million reads from human cell lines HEK-293 [17] and HeLa [18] respectively, 91% and 97% of these sequences respectively corresponding to various classes of small RNAs regulating critical cellular processes.

Analyzing small RNAs in human mitochondria, we used two samples coming from HEK-293 (GSM797669 with 730293 nucleotides [17]) and HeLa (GSM797670 with 363056 nucleotides [18]). For HEK293 (resp. HeLa), the least frequent 7-ALsq AAGAUGA (the least frequent also on the histogram of the Figure 3) is observed 1760 (resp. 890) times, for an expected number of 730293/1024 = 713 ± 160 (6σ) (resp. 363056/1024 = 355 ± 113 (6σ)), showing a very significant occurrence of AL relics inside the human mitochondrial genome.

3.3. Circular RNA and DNA sequences

The circular RNA sequences correspond to possible inhibitors of the microRNAs ("microRNA sponges" [18-20]). An example is given on Figure 4, which shows the inhibitory power of human circular RNAs like CDR1as and ciRs7 respectively on miR-671 and miR-485, and on miR-671 and miR-485, as well as the inhibitory power of human microRNAs miR-671 and miR-485 on their proteic targets, respectively GPI (Glucose-6-Phosphate Isomerase involved in glycolysis) and FPN1 (Ferroportin involved in the iron regulation as shown in Section 4.3).

```
3’-ACGACCUUCGGGACCUCUACGACCUUCU-5’   CDR1as
      3’-GUUGGUGAGUUUUACUUGUUGU-5’ anti-AL 15 anti-matches vs hsa-miR-671-5p
    5’-GGAAGCCCUGGAGGGGCUGGAGG-3’ hsa-miR-671-5p 21 anti-matches vs CDR1as
  3'-CGAUCUGUACCUCCCCGACC-5' GPI 16/20 anti-matches vs hsa-miR-671

   3’-UGUUGGUGAGUUUUACUUGUUG-5’ anti-AL 14 anti-matches vs ciRs7
5’-AGAGAGGAUGGGGGAGUUGUGUAUUCUUCCAGGUUC-3’   ciRs7
                              3’-CUGGAUCAGUGGAUCUA-5’ IRE-FPN1a 12/17 anti-matches vs ciRs7
5’-AUGGGGCAACAUAUUGUAUGAA-3  FPN1a 14 anti-matches vs hsa-miR-485
3’- UCUCUCCUCUCGGCACAUACUG-5’  hsa-miR-485 15 anti-matches vs ciRs 7
5’-AGAGAGGAUGGGGGAGUUGUGUAUUCUUCCAGGUUC-3’   ciRs7
                        5’-CCUGUUGGUCUCUUCCAGGUAC-3’   IRP 14 matches with ciRs7
```

Figure 4. Correspondence between AL, anti-AL, circular RNAs and their targets, the microRNAs.

Exploring the circular RNA data base [21], we observe many circular RNAs with a significant frequency of occurrence of AL subsequences. We give a characteristic example where these occurrences are indicated in red or blue:

hsa_circ_0071327 Alignment block 2 of 2 in window, 159076767 - 159076977, 211 bps

AACAAAAAA**TCAAAGAG**T**GCCAT**CTTGGACCACTC**ATGATGATG**TAT**TTCAGTAC**
AACCCG**ATTCA**GGCTTG**GGTACT**CG**GCCATTCTGCCA**GCATT**TCTGT**TTCA**GCAA
CTGCTGATAAGTTCCCCAGGTGAGCTTAACAG**AAGAATGG**GTGT**CATTACTT**GC**TG**
AAGATA**AAGATG**CATCCCA**AAGAATGATGG**GGC**ATGG**GCG**GCCAT**

We observe on the sequence above 19 times 5-ALsqs from the sequence corresponding to D and T RNA loops: **UUCAAGATGAATGGTAC**, for 207x13/1024 = 2.6 ± 9.7 (6σ) expected. Exploring in the same way the circular RNA data base [22], we observe also many circular RNAs with a significant frequency of occurrence of AL subsequences (ALsq).

We give another characteristic example where these occurrences are indicated in red or blue, both for the circular RNA and its microRNA target:

RP11-877E17.2:hsa-miR-194-5p
miRNA 3'-aggtGTACC**TCAA**CGAC**AATG**t-5'
| | | | | | | | | | |
ncRNA 5'-t**tgcCATT**TA**GCCACTG**T**TACt**-3'

From [23], the distribution of the AL subsequences of length 5 (cf. Figure 5) in Assphage circular DNA (97065 bp) is equal to:

ttcaa 250 tcaag 154 caaga 146 aagat 163 agatg 163 gatga 122 atgaa 211 tgaat 238 gaatg 152 aatgg 145 atggt 156 tggta 120 ggtac 62 gtact 90 tactg 143 actgc 129 ctgcc 160 tgcca 155 gccat 121 ccatt 198 cattc 155 attca 206.

The total number of 5-ALsqs observed is equal to 3439, with an expected total equal to 22x97061/1024 = 2085 ± 73 (0.95-confidence interval). The entropy of the above distribution is equal to: $H_2 = 2.77723$.

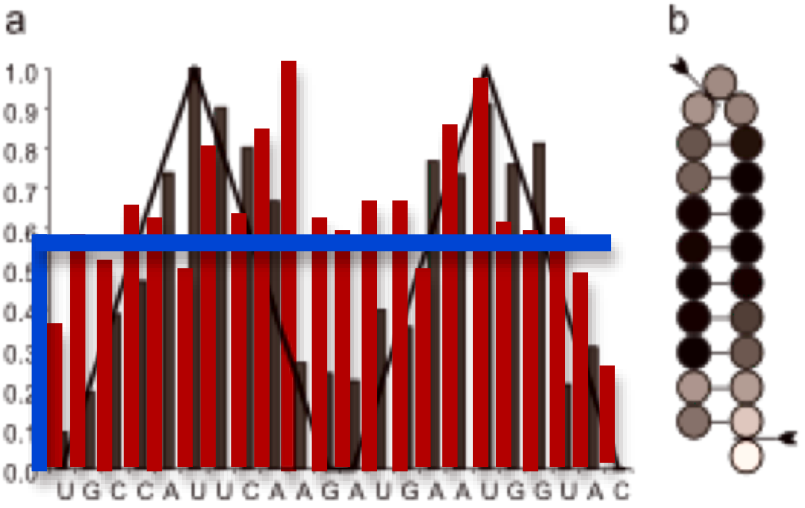


Figure 5. Histogram of the 5-ALsqs observed in Assphage circular DNA [23]. Frequencies up to the blue bar correspond to the 0.95-significativity vs the uniformity.

If we are looking more precisely to the distribution of AL subsequences of length more than 6, we have 17 observed 9-ALsq, with 8.145 ± 4.57 expected and 69 observed 8-ALsq, with 32.6 ± 9 expected (Figure 6). The entropy of the distribution corresponding to the histogram given on Figure 6 is equal to: $H_3 = 2.26074$.

9 & 10-ALsq		8-ALsq	
agatgaatg	1	agatgaat	2
gatgaatgg	0	gatgaatg	3
atgaatggt	1	atgaatgg	2
atgaatggta	1	tgaatggt	2
tgaatggta	1	gaatggta	3
gaatggtac	0	aatggtac	2
aatggtact	2	atggtact	4
atggtactg	0	tggtactg	1
tggtactgc	0	ggtactgc	0
ggtactgca	0	gtactgcc	0
tactgccat	1	tactgcca	2
tgccattca	1	actgccat	1
tgccattcaa	1	ctgccatt	1
gccattcaa	1	tgccattc	2
ccattcaat	0	gccattca	2
attcaagat	2	ccattcaa	4
ttcaagatg	0	cattcaag	1
attcaagatg	1	attcaaga	4
ttcaagatg	3	ttcaagat	10
tcaagatga	3	tcaagatg	9
aagatgaat	1	caagatga	4
aagatgaatg	1	aagatgaa	10

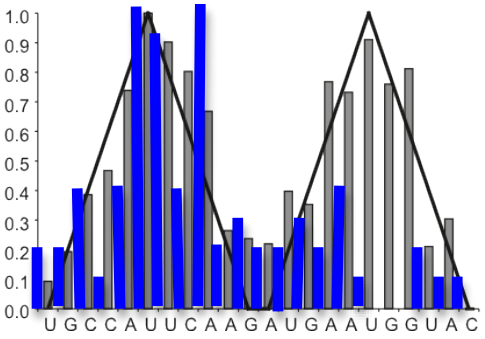


Figure 6. Histogram of the 8-AL sequences observed in Assphage circular DNA [23].

Entropy of the distribution corresponding to the histogram of Figure 6 is equal to: H_3=2.26074.

3.4. DNA sequences from genes of different species

By exploring the complete genomes of microbes [24], which contains15,358,075,464 bp for 4867 sequences, we observe the AL subsequence TTCAAGATGAATGGT (D and T loops being indicated in red) 51 times for 14 ± 7.5 expected. We discover also quasi-perfect matchings:
Desaturase gene fragment, 5218 bp [25]: CAGCCCTCCAAGATGAATGGTA 19 AL-matches
Candida Ca20chr5, 1191531 bp [26]: TGGTACTGCCATTGAAGATAGA 19 AL-matches
Glyma14g34640.1, 3709 bp [27]: TGCTATTCAAGACTATGAAATG 19 AB-matches.

By exploring the complete genomes of some animals [28], we observe frequently 12-ALsq centred on D or T loops:

102 times CCATTCAAGATG in Mus musculus (Mouse) RefSeq RNA
Number of letters: 319,391,879. Number of sequences: 106,780
Expected: 19 ± 7

38 times UGAAUGGUACUG in Oryctolagus cuniculus (Rabbit) RefSeq RNA
Number of letters: 118,511,990. Number of sequences: 42,502
Expected: 7 ± 4.2

104 times CCATTCAAGATG in Pantroglodytes (Chimpanzee) RefSeq RNA
Number of letters: 206,359,824. Number of sequences: 68,378
Expected: 12 ± 5.5

33 CCATTCAAGATG in Danio rerio (Zebrafish) RefSeq RNA
Number of letters:155,246,628. Number of sequences: 54,469
Expected: 9

We can also notice that we observe in [28] 191 times the 11-ALsq GCCATTCAAGA in the 370 M bp of plasmid genomes instead 88 ± 19 expected.

3.5. The use of the static entropy

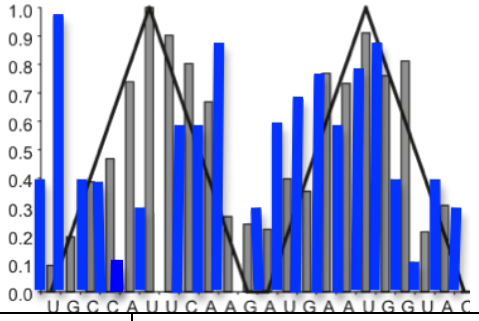
By exploiting the thermodynamic parameters given in [8,9] (cf. Table 2), we can calculate the theoretical probability of occurrence of the AL subsequences (cf. Figure 6).

Table 2. Nearest-neighbor thermodynamic parameters for DNA/DNA duplexes [8,9].

Nearest-neighbor sequence (5'-3'/3'-5')	ΔH kJ/mol	ΔS J/(mol·K)	$\Delta G_{37°C}$ kJ/mol
AA/TT	−33.1	−92.9	−4.26
AT/TA	−30.1	−85.4	−3.67
TA/AT	−30.1	−89.1	−2.50
CA/GT	−35.6	−95.0	−6.12
GT/CA	−35.1	−93.7	−6.09
CT/GA	−32.6	−87.9	−5.40
GA/CT	−34.3	−92.9	−5.51
CG/GC	−44.4	−113.8	−9.07
GC/CG	−41.0	−102.1	−9.36
GG/CC	−33.5	−83.3	−7.66
Terminal A-T base pair	9.6	17.2	4.31
Terminal G-C base pair	0.4	−11.7	4.05

The static Shannon entropies of the empirical histogram (H_4) and of the predicted distribution of the AL sequences of length 5 (H_5) occurring in 100 circular RNAs of length less than 250 bases (cf. Figure 7) chosen by chance in [21], are respectively equal to: $H_4 = 2.50232$ and $H_5 =$

3.05508. We can observe that the thermodynamic prediction H_5 is highly greater than all the other entropies H_i (i = 1,...,4), showing that the occurrence of AL relics in the present genomes is definitely not due to the chance nor to thermodynamic constrains during the evolution.



5-uple	Nb	frequency = Nb/106	-ΔG (kJ/mol)	exp(-ΔG/RT)	exp(-ΔG/RT)/Z % Z = 10794
aatgg	8	0.075	13.32	175	1.6%
atggt	9	0.085	14.92	325	3%
tggta	4	0.038	14.92	325	3%
ggtac	1	0.0095	14.27	253	2.3%
gtact	4	0.038	11.75	95	0.9%
tactg	3	0.028	11.75	95	0.9%
actgc	4	0.038	18.61	1361	12.6%
ctgcc	10	0.094	20.41	2735	25.3%
tgcca	4	0.038	20.58	2921	27%
gccat	4	0.038	18.45	1325	12.3%
ccatt	1	0.0095	12.32	119	1.1%
cattc	3	0.028	11.43	84	0.8%
attca	0	0	10.91	69	0.6%
ttcaa	6	0.057	11.53	87	0.8%
tcaag	6	0.057	12.93	150	1.4%
caaga	9	0.085	12.93	150	1.4%
aagat	0	0	10.33	55	0.5%
agatg	3	0.028	12.42	123	1.1%
gatga	6	0.057	12.45	125	1.1%
atgaa	7	0.066	10.91	69	0.6%
tgaat	8	0.075	10.91	69	0.6%
gaatg	6	0.057	11.43	84	0.8%

Figure 7. Empirical and theoretical (thermodynamically predicted) histograms corresponding to the observation of 106 5-ALseq for 69 ± 13.3 expected in a sample of 100 circular RNAs of length less than 250 b (Figure 6) chosen by chance in [21].

4. Small RNAs as regulators in genetic networks

4.1. The immunetwork responsible of the Toll Like Receptor (TLR) expression

The activation of Natural Killer (NK) cells, involved in innate immune response, is controlled by PU.1 gene (Figure 8) and by the ligands of the Toll Like Receptors (TLR). The RP105 (CD180) is a TLR-related protein identified in B cells (responsible of the humoral adaptive immune response) [29-30] and acting as regulator of B cell proliferation. B-cells lacking RP105 were shown to be severely impaired in antibody production. The protein ICAM1 is a type of intercellular adhesion molecule continuously present in low concentrations in the membranes of leucocytes involved in humoral adaptive immune response.

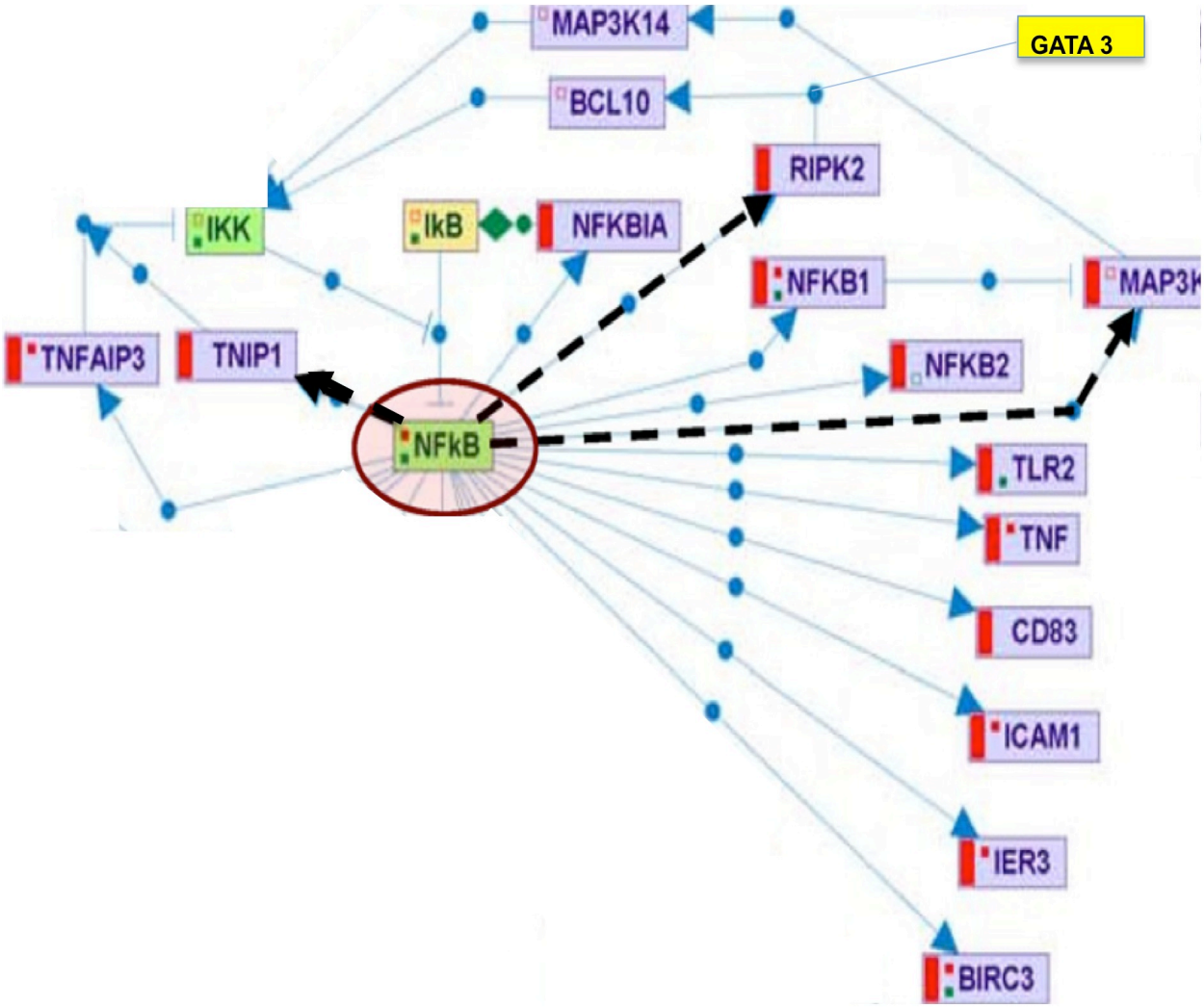


Figure 8. The network controlling the production of the Toll Like Receptors of ICAM1.

The network controlling the Toll Like Receptors (TLR) and ICAM1 expression contains a couple of circuits, one positive 4-circuit tangential to a negative 4-circuit (Figure 8), giving only one attractor, which corresponds to the activation of TLR2.

4.2. The links with the microRNAS

All the genes introduced above have links with microRNAs exerting a negative control on them and susceptible to decide what attractor will occur, by cancelling their target gene activity. For example, we have:

1) for the TLR protein pUNO-hRP105 sub-sequence (4937 bp), the hybridization is made by the microRNA miR 200, close to the reference sequence AL:

5’-CCAUUCAAGAUGAAUGGUACUG-3’ AL 14 anti-matches
3’-UGUAGCAAUGGUCUGUCACAAU-5’ hsa miR 200a 12 anti-matches
5’-UUGUGCUCAUUGAGAUGAAUGG-3’ pUNO-hRP105 sequence starting in position 531
5’-UACUGCCAUUCAAGAUGAAUGG-3’ AL 15 matches

2) for the GATA-3 gene:

5’-GCCAUUCAAGAUGAA--UGGUACU -3’ AL 13 anti-matches
3'-AGGUAGUAAUGGGC--CGUCAUAA-5' has miR 200c 16 anti-matches
5’-UCUGCAUUUUUGCAGGAGCAGUA-3' GATA-3 sequence starting in position 57

3) for the ICAM1 sub-sequence CD54 cDNA (1615 bp):

5’-GGUGCCUAUUAACAAUAUGAAU-3’ AB 15 anti-matches
3’-GUACGUGUACGUGUG--UAUGUA-5’ has miR 297 15 anti-matches
5’-CCUCCCCA------CCCAC--AUACAU-3’ ICAM1 sequence starting in position 832

We can remark on the above matching that i) gene expressing the TLR hRP105 protein contains the sequence AGATGAA frequently observed in the not coding genome as part of the CAAGATGAA sequence, belonging both to the T loop of the tRNAs and to the reference sequence AL, which signifies the affiliation to an ancestral genome, and ii) the microRNAs hybridizing the genes TLR hR, GATA-3 and ICAM1 are also close to the reference sequence AL, showing the old origin of the innate immunologic system.

4.3. The IRP/IRE Network

The regulatory network controlling the iron (Fe) metabolism is given on Figure 9. The arrows are defined by the interactions described in Section 3.3. We can calculate the gene expression behavior (called attractor) obtained for any initial condition of the expression of the genes of the network, belonging to the attraction basin of this attractor.

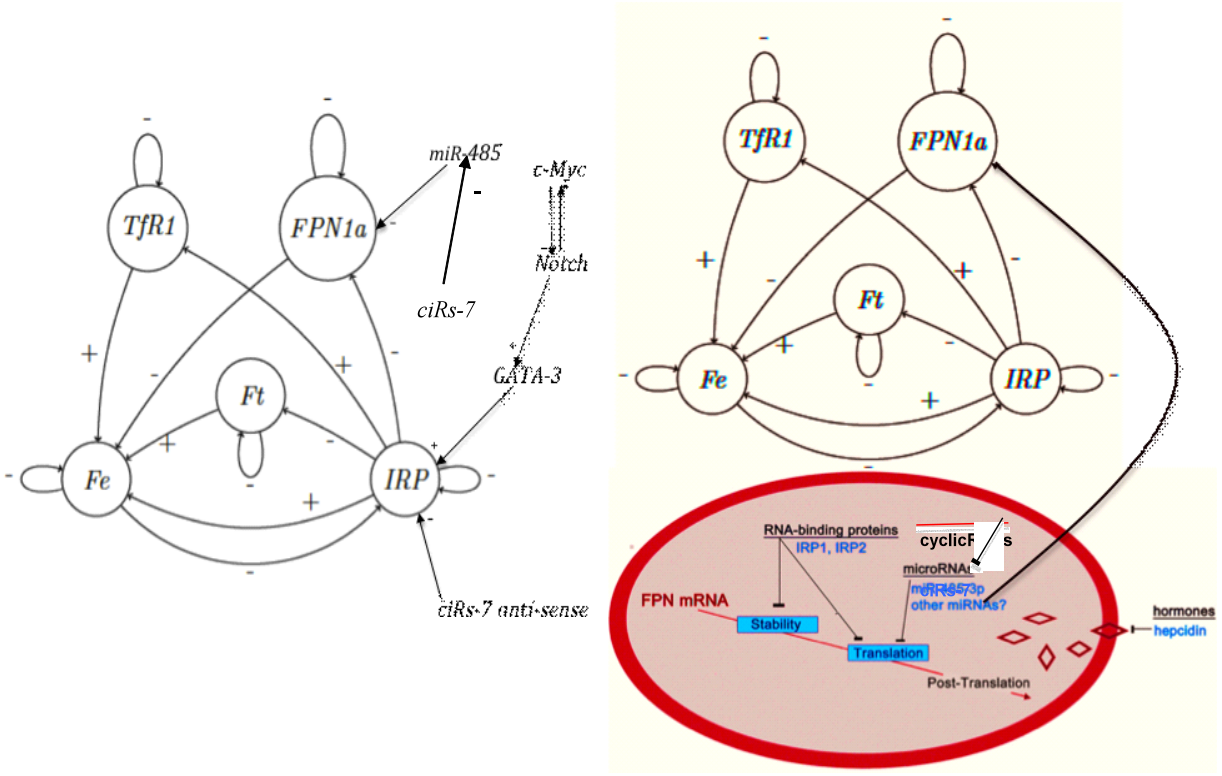


Figure 9. The interaction graph of the iron regulatory network with inhibitory and activatory interactions (left) and the role [32,33] of the small RNAs (right).

Table 3. Recapitulation of the attractors of the iron metabolic system, for the parallel updating mode, with the list of expressed (in state 0) and not expressed (in state 1) genes and their attraction basin sizes.

Position	Gene	Fixed Point 1	Fixed Point 2	Limit Cycle 1			Limit Cycle 2		
1	TfR1	0	0	0	0	1	1	0	
2	FPN1a	0	0	0	0	0	0	0	
3	C-Myc	0	1	0	0	0	0	0	
4	Notch	0	0	1	1	1	1	1	
5	GATA-3	0	0	1	1	1	1	1	
6	IRP	0	0	0	1	0	0	1	
7	Ft	0	0	0	0	0	0	0	
8	Fe	0	0	0	0	1	0	1	
9	miR-485	0	0	0	0	0	0	0	
10	ciRs-7 anti-sense	0	0	0	0	0	0	0	
	Relative Attraction Basin Size (RABS)	512/1024	256/1024	216/1024			40/1024		

5. Dynamical entropy and network robustness

We define [29,30] the energy U and frustration F of a genetic network N of size n by:

$$\forall\, x \in \Omega,\ U(x) = \Sigma_{i,j \in \{1,n\}} \alpha_{ij} x_i x_j = Q_+(N) - F(x),$$

where x is a configuration of gene expression ($x_i = 1$, if the gene i is expressed and $x_i = 0$, if not), $Q_+(N)$ is the number of positive edges in the interaction graph G of the network N and $F(x)$ the global frustration of x, *i.e.*, the number of pairs (i,j) where the values of x_i and x_j are contradictory with the sign α_{ij} of the interaction between genes i and j: $F(x)=\Sigma_{i,\, j \in \{1,n\}} F_{ij}(x)$, where F_{ij} is the local frustration of the pair (i,j) defined by:
$F_{ij}(x) = 1$, if $\alpha_{ij}=1$, $x_j=1$ and $x_i=0$, or $x_j=0$ and $x_i=1$, and if $\alpha_{ij} = -1$, $x_j=1$ and $x_i=1$, or $x_j=0$ and $x_i=0$,
$F_{ij}(x) = 0$, elsewhere.

Then the dynamical entropy E of the network is defined by the Kolmogorov-Sinaï entropy of the transition operator allowing the passage from a state of expression to the next state of expression [34]. When this transition is stochastic, E can be approached by calculating the attractor entropy:

$$E \approx \log 2^n - E_{attractor},$$

where $E_{attractor}$ can be evaluated by the quantity: $E_{attractor} = - \Sigma_{k=1,m \leq 2^n} RABS(A_k) \log RABS(A_k)$, with $RABS(A_k)$ equal to the size of the attraction basin of the attractor A_k, divided by 2^n.

E serves as robustness parameter, being related to the capacity the genetic network has to return to the equilibrium after endogenous or exogenous perturbations [35]. The maximum value of E is indeed n and for the immunologic network, there is a unique attractor, hence $E_{attractor} = 0$ and E takes this maximum value equal to the number of the genes of the network, *i.e.* $E = 19$. In the iron network, we have: $E_{attractor} = 2.59412$, hence $E = 10 - 2.59412 = 7.40588$, which corresponds to a relatively low robustness. We can prove more generally [30] that the robustness of the network decreases if the frustration F of the network increases, because of the formula: $\partial E / \partial \log w = - \mathrm{Var} F$, where w is the absolute value of the interaction (here supposed to be the same for all interactions) between the genes of the network.

6. Conclusions

The genetic networks regulated by the small RNAs show important properties necessary to control a biological function. The role of the microRNAs is to provide a partially unspecific inhibitory noise leaving only the control circuits having sufficiently strong interactions to be able to express the attractors. Circular RNAs are inhibiting the microRNAs in order to have, like in neural networks, the possibility to get a double reciprocal influence (inhibitory and anti-inhibitory) on mRNAs and genes, *i.e.*, on the genetic expression. The presence of RNA relics in a genome contributed to its robustness, essentially by reducing the number of the circuits of the interaction graph, responsible (when they are positive) for the multiplication of the attractors [4].

In any case, the use of entropy (static or dynamical) can serve to quantify the role of small RNAs in maintaining during evolution both ancestral control and robustness.

Acknowledgments

The authors acknowledge the European project VPH (Virtual Physiological Human) and the National French project VHP (VisioHome Presence) for their support.

Author Contributions

Both authors contributed equally to this work. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Demongeot, J. Sur la possibilité de considérer le code génétique comme un code à enchaînement. *Revue de Biomaths* **1978**, *62*, 61-66.
2. Demongeot, J.; Besson, J. Code génétique et codes à enchaînement I. *C.R. Acad. Sc. Série III* **1983**, *296*, 807-810.
3. Demongeot, J.; Besson, J. Genetic code and cyclic codes II. *C.R. Acad. Sc. Série III* **1996**, *319*, 520-528.
4. Demongeot, J.; Aracena, J.; Thuderoz, F.; Baum, T.P.; Cohen, O. Genetic regulation networks: circuits, regulons and attractors. *C. R. Biologies* **2003**, *326*, 171-188.
5. Demongeot, J.; Moreira, A. A circular RNA at the origin of life. *J. Theor. Biol.* **2007**, *249*, 314-324.
6. Demongeot, J.; Drouet, E.; Moreira, A.; Rechoum, Y.; Sené, S. MicroRNAs: viral genome and robustness of the genes expression in host. *Phil. Trans. Royal Soc. A* **2009**, *367*, 4941-4965.
7. Demongeot, J.; Glade, N.; Moreira, A.; Vial, L. RNA relics and origin of life. *Int. J. Molecular Sciences* **2009**, *10*, 3420-3441.
8. Klingler, T.; Brutlag, D.L. Detection of Correlations in tRNA with structural implications. *Intelligent Systems for Molecular Biology* **1993**, *1*, 225-233.
9. Rychlik, W.; Spencer, W. J.; Rhoads, R. E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **1990**, *18*, 6409–6412.
10. Choi, H.; Gabriel, K.; Schneider, J.; Otten, S.; McClain, W.H. Recognition of acceptor-stem structure of $tRNA^{Asp}$ by *Escherichia coli* aspartyl-tRNA synthetase. *RNA* **2003**, *9*, 386-393.
11. Shigi, N.; Suzuki, T.; Tamakoshi, M.; Oshima T.; Watanabe, K. Conserved Bases in the T C Loop of tRNA Are Determinants for Thermophile-specific 2-Thiouridylation at Position 54*. *J. Biol. Chem.* **2002**, *277*, 39128–39135. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **1998**, *26***,** 148–153.
12. Tanaka, T.; Kikuchi, Y. Origin of the cloverleaf shape of transfer RNA - the double-hairpin model: Implication for the role of tRNA intro and the long extra loop. *Viva Origino* **2001**, *29*, 119-142.

13. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; A. Zamir, A. Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462-1465.
14. Lewin B. *Genes IX*; Jones & Bartlett: Boston 2008.
15. Griffiths-Jones, S.; M. Marshall, M.; Khanna, A.; Eddy, S.R.; Bateman, A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **2005**, *33*, 121-124 and http://rfam.xfam.org/.
16. http://sra.dnanexus.com/experiments/SRX097598
17. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM797670
18. T.B. Hansen, E.D. Wiklund, J.B. Bramsen, S.B. Villadsen, A.L. Statham, S.J. Clark, J. Kjems, miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA, The EMBO Journal 30 (2011) 4414-4422.
19. Hansen, T.B.; Jensen, T.I.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* **2013**, *495*, 384-388.
20. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak S D, Gregersen L H, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495:333-338
21. http://www.circbase.org/cgi-bin/tablebrowser.cgi
22. http://starbase.sysu.edu.cn/
23. http://www.ncbi.nlm.nih.gov/nuccore/JQ
24. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes
25. http://www.google.bj/patents/CN100591773C?cl=en
26. http://www.candidagenome.org/download/sequence/C.../Ca20chr5.embl
27. http://www.phytozome.net/cgi-bin/gbrowse_details/soybean?class=mRNA;name=Glyma14g34640.1
28. http://blast.ncbi.nlm.nih.gov/
29. Demongeot, J.; Cohen, O.; Henrion-Caude, A. MicroRNAs and Robustness in Biological Regulatory Networks. A Generic Approach with Applications at Different Levels: Physiologic, Metabolic, and Genetic. *Springer Series in Biophysics* **2013**, *16*, 63-114.
30. Demongeot, J.; Ben Amor, H.; Hazgui, H.; Waku, J. Robustness in Neural and Genetic Regulatory Networks: Mathematical Approach and Biological Applications. *Acta Biotheoretica* **2014**, *62*, 243-284.
31. Sangokoya, C.; Doss, J.F.; Chi, J.T. Iron-Responsive miR-485-3p Regulates Cellular Iron Homeostasis by Targeting Ferroportin. *PLoS Genet.* **2013**, *9*, e1003408.
32. http://mirdb.org/miRDB/
33. http://mirnamap.mbc.nctu.edu.tw/
34. Demetrius, L. Statistical mechanics and population biology. *J. Stat. Physics* **1983**, *30*, 709–753.
35. Demetrius, L. Directionality principles in thermodynamics and evolution. *P.N.A.S.* **1997**, *94*, 3491–3498.

Chapitre 3

Dynalets

Sommaire

Introduction

Les différentes méthodes qui représentent les signaux biologiques, ont pour but ou bien d'expliquer les mécanismes ayant produit le signal ou de faciliter son utilisation dans des applications médicales... Ces signaux peuvent provenir des capteurs de signaux électrophysiologiques comme l'ECG, les capteurs de pouls, les dispositifs d'investigation au niveau moléculaire, comme la spectrométrie de résonance nucléaire ou de masse.
Dans ce chapitre, nous exposons la méthode nommée Dynalets, utilisée pour traiter ces signaux biologiques.

3.1 Présentation générale de la méthode des Dynalets

La méthode Dynalets consiste à utiliser un système d'équations différentielles de Liénard. L'identification de protéines par leur spectre permet par exemple la construction de réseaux de contrôle génétiques complexes, tels que ceux trouvés dans la régulation du système immunitaire, où les protéines clés sont des activateurs ou des inhibiteurs de l'expression génétique. Bien sûr, d'autres techniques alternatives pour estimer les spectres de protéines existent déjà, mais elles ne sont pas liées au mécanisme de la production du signal de la protéine. La transformée Dynalets donne sens à de nombreuses données métaboliques(très lourdes en terme informationnel). Ces données ne sont pas en général utilisées par les cliniciens (en particulier en cas d'urgence).

Le 17 février 1816, le docteur René Laennea a inventé le stéthoscope.Cette technique permet la conversion sonore d'une information fonctionnelle, décrite auparavant sont décrites par les médecins sur de nombreux niveaux anatomiques ou physiologiques. Nous continuons dans la même optique, en représentant les informations spectrales de masse et de résonance en signaux convertis en sons. En pratique, pour effectuer la transformée Dynalets, il faut choisir les valeurs des paramètres suivants :

— le paramètre μ, lié à la tel que la période du cycle limite de van der Pol
— des translations de l'origine des axes, afin de fixer le point de phase 0 du van der Pol
— des homothéties sur ces axes en minimisant la distance entre deux ensembles de points à la fois ceux provenant du van der Pol, et ceux venant de l'acquisition du signal original.

3.2 Description de l'outil d'application

Une interface a été développée avec le langage Java, au sein de notre équipe de recherche, afin de mettre en œuvre la méthode des Dynalets. Par la suite, nous détaillerons le fonctionnement de notre modèle avec cette interface. Notre point de départ est le signal complet (voir graphe ci-dessous)du spectre RMN de l'acide aminé (ou de la base azotée). Ensuite, on en a extrait le pic le plus intéressant pour notre étude, de façon à ne retenir qu'une information spécifique, « pathognomonique » de la molécule étudiée.

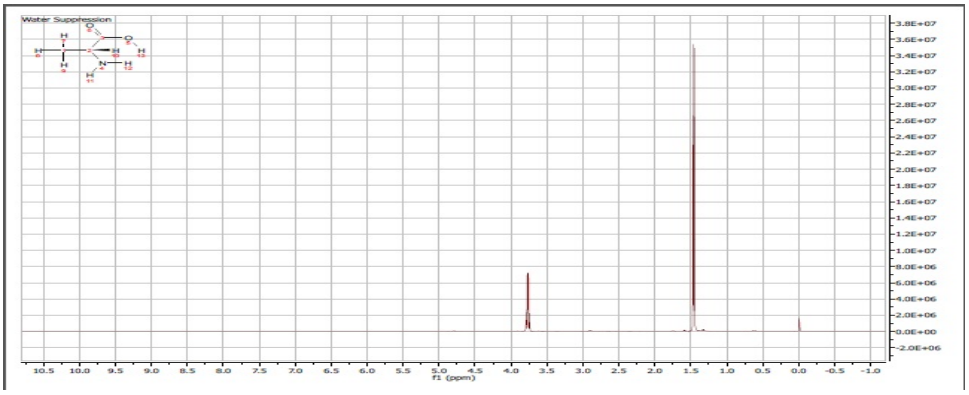


FIGURE 3.1 – Spectre RMN de l'alanine

Par la suite, on procède aux calculs de la dérivée du signal X, notée D, à chaque temps T :

$$D = \frac{dX}{dT} = \frac{X_i - X_{i-1}}{T_i - T_{i-1}}$$

Cette dérivée constitue notre troisième valeur d'entrée pour notre programme de modélisation (X,T et D). Les valeurs sont normalisées. Une fois qu'on a notre fichier d'entrée, nous devons modifier les paramètres suivants :
-μ, X_0 et Y_0 ce qui permet de donner une nouvelle forme au signal VdP.
-translation en X , ainsi qu'homothéties en X et Y, de même coefficient, qui permettent de rapprocher le signal VdP du signal expérimental.
Enfin, il ne faut pas le paramètre "pas", ainsi que le nombre de points du modèle de Van der Pol fixés dés le départ et qui ne changent pas, contrairement aux autres paramètres.
On calcule le modèle de Van der Pol à l'aide de la méthode d'analyse numérique **Runge Kutta** d'ordre 4 (les valeurs seront normalisées aussi). Le nom de cette méthode, élaborée en 1901, provient des noms de deux mathématiciens Carl **Runge** et Martin Wilhelm **Kutta**. Ensuite, on applique les transformations et on trace les trois signaux : expérimental (rouge), VdP (vert clair) et VdP après transformations (vert foncé). Le Signal VdP est généré avec beaucoup plus de points que le signal expérimental. L'entrée de l'interface est présentée par le graphique ci-dessous :

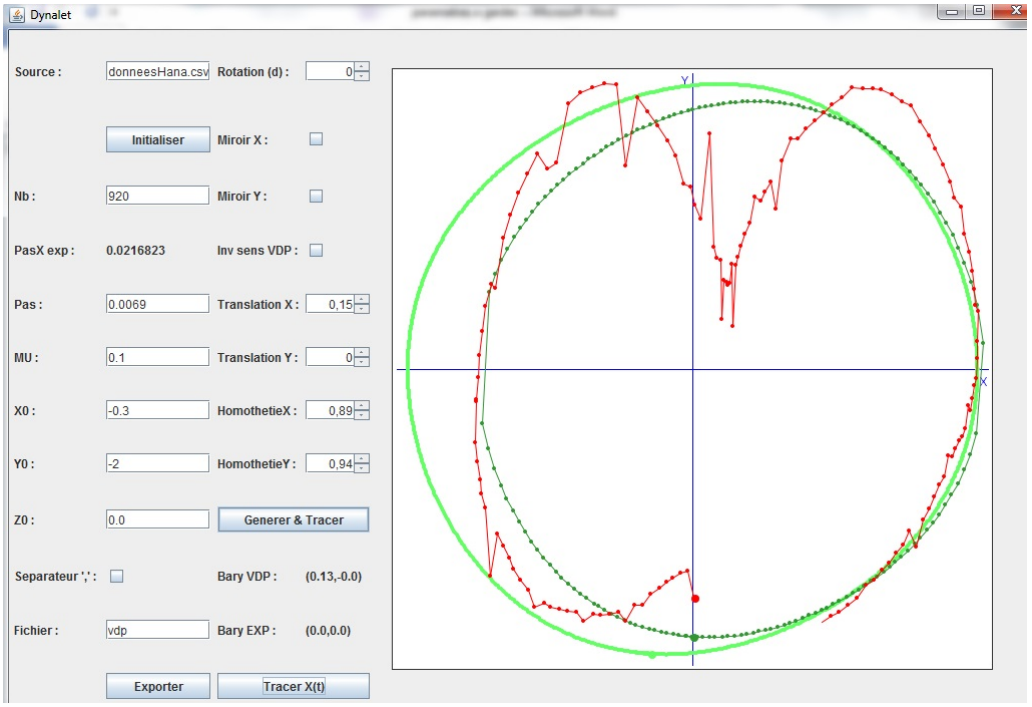


FIGURE 3.2 – Tracé des 3 signaux après la transformation Dynalet

Lors du tracé X,Y, les temps du signal VdP ont été calculés à partir de la dérivée Y. Ils sont donc utilisés pour les tracés des signaux temporels pour le VdP. Voici les tracés : expérimental (rouge), VdP (vert clair) et VdP après transformations (vert foncé).

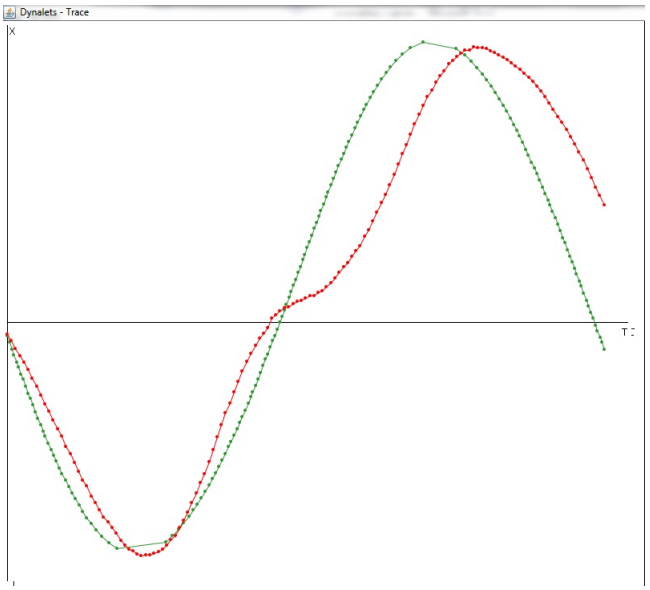


FIGURE 3.3 – Tracé temporel des 3 signaux pour le VdP.

Une fois le signal modélisé et les paramètres ajustés, il est possible de sauver les données générées dans des fichiers csv :

1. Un premier fichier ”csv” : récapitulatif des signaux expérimentaux et VdP avec le même nombre de points et le calcul des différences
2. Un deuxième fichier ”csv” : le signal expérimental tel qu'il a été entré
3. Un troisième fichier ”csv” : le signal VDP généré (tous les points).

Ensuite, nous calculons l'harmonique en repartant d'un nouveau signal, égal à la différence des deux signaux obtenus précédemment. On obtient un nouveau X normalisé, résultat de EXP-VDP. Puis il est remis en entrée d'un calcul d'approximation VdP, avec deux tours cette fois-ci. Pour conserver l'échelle du signal principal, on divise le résultat par le coefficient qui a servi à normaliser EXP-VdP. La dernière étape consiste à calculer la fondamentale ainsi que l'harmonique à partir des résultats obtenus par la méthode Dynalet.

3.3 Application

Par la suite, nous illustrons cette méthode par une application, en utilisant des signaux de spectrométrie RMN. Le but d'utiliser la méthode des Dynalets ici est de convertir les pics du spectre en sons, ce qui permet souvent de distinguer à l'oreille des tracés difficilement distinguables à la vue.

3.3.1 Domaine d'application

Nous utilisons pour notre application des signaux de spectroscopie par Résonance Magnétique Nucléaire ”RMN”. Cette méthode fut développée par deux équipes de physiciens, celle de Félix Bloch (1905-1983) et celle de Edward Mills Purcell (1912-1997). Elle exploite les propriétés magnétiques de certains noyaux atomiques et elle est basée sur le phénomène de résonance magnétique nucléaire, d'où son nom. Les signaux exploités par

notre méthode Dynalets correspondent au graphique de l'analyse d'un spectre RMN du proton (1H NMR Spectrum). La méthode a été appliquée sur 20 acides aminés ainsi que les 4 bases azotées(ci dessous la liste).

Éléments	Nom	Code
Acides	Alanine	ala
	Arginine	Arg
	Asparagine	Asn
	Aspartate	Asp
	Cystéine	Cys
	Glutamate	Glu
	Glutamine	Gln
	Glycine	Gly
	Histidine	His
	Isoleucine	Ile
	Leucine	Leu
	Lysine	Lys
	Méthionine	Met
	Phénylalanine	Phe
	Proline	Pro
	Sérine	Ser
	Thréonine	Thr
	Tryptophane	Trp
	Tyrosine	Tyr
	Valine	Val
Bases	Adénine	T
	Cytosine	C
	Guanine	G
	Thymine	T

Table 3.1 – Liste des acides aminés et des bases nucléiques

3.3.2 Résultats

Nous avons appliqué la méthode Dynalet sur les 24 éléments cités dans la section précédente. Nous avons obtenu 24 sons (un son par composante). Ces sons font l'objet d'un travail en cours afin de constituer un clavier musical.

Actimetry@home: Actimetric Tele-surveillance and Tailored to the Signal Data Compression

Jacques Demongeot(✉), Olivier Hansen, Ali Hamie, Hana Hazgui, Gilles Virone, and Nicolas Vuillerme

AGIM (Ageing, Imaging & Modelling) Laboratory, Faculty of Medicine, FRE 3405 CNRS-UJF-EPHE, 38700 La Tronche, France
Jacques.Demongeot@yahoo.fr,
{Olivier.Hansen,Ali.Hamie,Hana.Hazgui,
Gilles.Virone,Nicolas.Vuillerme}@agim.eu

Abstract. An early diagnosis of a neurodegenerative process like the Alzheimer's disease needs a tele-surveillance at home based on the recording of pathologic signals coming both from the cardiac activity (for detecting the loss of the sinus respiratory arrhythmia) and from the repetition of tasks of the daily life (signing a pathologic behavior called perseveration), whose non-invasive detection can lead to an early diagnosis, if it triggers secondly a battery of tests based on brain imaging, clinical neurology and cognitive sciences to confirm the suspicion of neuronal degeneration. For increasing the efficiency of alarms triggering these tests, we use dedicated tailored data compression methods, whose two examples will be presented, the Dynalets method for quantitative compression of the physiologic signals and the monotonic signature for qualitative compression.

Keywords: Tele-surveillance at home · Alarm triggering · Tailored to the signal data compression

1 Introduction

The worldwide population is ageing, this phenomenon being accompanied by a dramatic increase of the neurodegenerative diseases: in 2050, the number of elders aged 65+ in the world will have increased by 100 % compared to the year 1950 [1]. The prevalence of neurodegenerative diseases in developed countries has been found to double with every five-year increase in age, from 3 percent at age 70 to 25 percent at age 85, the increase per year ranging from 0.84 to 3.50 percent [2]. Hence, new pervasive technologies have been developed for smart homes or smart residences for seniors [3, 4] with assistive technologies [5]. Experimental platforms and living labs are a prototype support for conducting experiments taking normally place in these healthcare facilities. This offers advantages such as the ability to pre-validate prototypes before their use in the real-world. In case of infrastructure absence, computer simulations [5–12] can play a key role, for simulating different activity trends based on heterogeneous parameters (*e.g.,* age, education, seasons, etc.) [13–16], testing uncommon scenarios of everyday life on demand (instead of waiting for unpredictable real-life apparitions), and assessing

C. Bodine et al. (Eds.): ICOST 2014, LNCS 8456, pp. 59–70, 2015.
DOI: 10.1007/978-3-319-14424-5_7

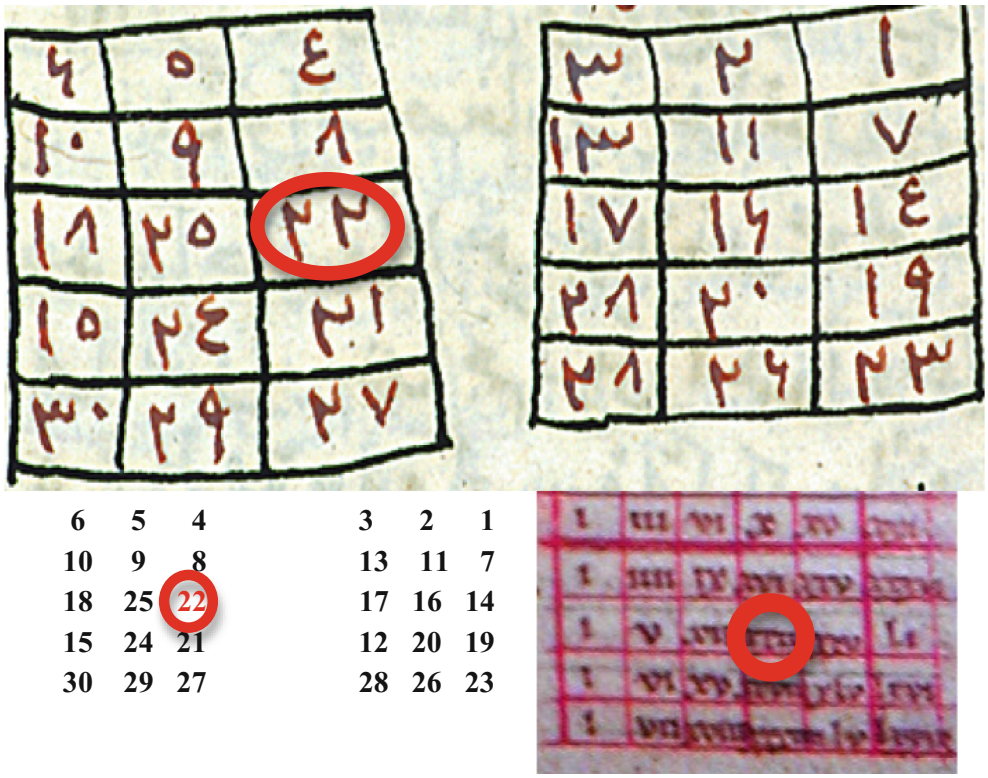
specific algorithms invented and used in the area of activity or cognition [17–20]. We have for example modeled a persistence in certain tasks of daily life to infer a more generalized decline induced by a neurodegenerative disease. This persistence, also known as perseveration, in performing common tasks already completed successfully, but repeated pathologically was already the subject of comment in Romans, who willingly quoted the proverb: "errare humanum est, perseverare diabolicum"… The basis of such scenarios of daily living implemented in our computer simulations started from a previous research described in [14]. The present article deals with two methodologies of compression tailored to the physiologic or pathologic mechanisms generating the signal used to follow the person at home: (i) first concerns ECG signal, whose loss of the sinus respiratory arrhythmia is a good marker of the entrance in a neurodegenerative disease and (ii) second is a qualitative method for interpreting a sequence of increase/decrease of an activity, by focusing only on its successive intervals of monotony. Note that samples used for validating these methods are essentially simulated using random numbers generator, bootstrapping data from already observed empirical distributions [3, 4].

2 History: Magical Medicine, Hermetic and Mysterious Numerology

Greek and Romanesque medicine have used commonly tables for predicting the fatal (resp. happy) evolution of a patient, if a mysterious calculation involving his age and pulse frequency belonged to a certain "lethal" part of an arithmetic table (Fig. 1 top left). Heir of Aristotle, hermetic pre-Roman and early Middle Age physicians gave rise to tables such as the "secret of secrets" (from an apocryphal letter by Aristotle to Alexander, attributed to Rhazes (circa 900), but probably from Hermes Trismegistus (circa 100)) showing on Fig. 1 in its center the number 22 (number of vertebrae in humans, nucleic bases in a microRNA,…). The table by G. d'Auberive (1180) [21], however, remains of more obscure origin usefulness in medicine, but shows also the intention to condensate health predictions in hermetic numerology.

3 Data Compression: From Tailored to Signal Physiology Towards Alarm Adapted Schematized Information

As the ancient Greeks, today's physicians, who seek to follow the evolution of a patient at home, need fast calculation of physiological parameters derived from clinical, biological or imaging records. For this, methods of rapid real-time processing must be developed to quickly compare patient evolutions with those from databases containing comparable patients. We give here two examples of the compression needed to quickly extract relevant parameters, in order to trigger an early and reliable alarm: one concerns an ECG signal decomposition closer to the genesis of the physiological signal that the Fourier transform, and the other deals with actimetric data reduced to their signature monotony (*i.e.*, to the sequence of the signs of their derivative).



6	5	4	3	2	1
10	9	8	13	11	7
18	25	22	17	16	14
15	24	21	12	20	19
30	29	27	28	26	23

Fig. 1. Top: Table of secret of secrets and its translation into modern Arabic digits in the bottom left. A patient was declared fatal prognosis, if a formula with its age and the numerical value of the name belonged to the left side of the table. Bottom right: G. d'Auberive table, with the sum of the first integers in first row and the cumulative sum of multiples of the previous amount in column.

3.1 The Compression of the ECG Signal: The Dynalets Method

We propose in the following a new method of compression of the QRS wave of the ECG based on its mechanism of generation inside the sinus node of the heart.

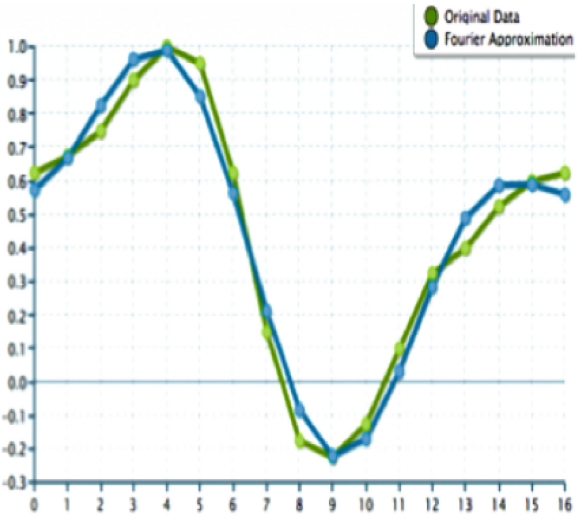
The method, called Dynalets decomposition, consists of three main steps:

- we have represented together in the phase plane (amplitude, speed), the physiological signal (Fig. 2) and the approximation function belonging to a family of approximating selected functions (adapted to the mechanism of genesis of the physiological signal), and their proximity is maximized by minimizing a given distance in the phase plane (distance "delta", which minimizes the area between the curves, quadratic distance, which minimizes the sum of the distances between points chosen on the curves, or Hausdorff distance between the inner subsets of the curves) (Fig. 3)
- we subtract the function thus obtained, called fundamental signal, to the experimental signal, so as to obtain a second function, which we seek to approximate in the selected functional family
- iterations are stopped when the quadratic relative error (QRE) is below 0.1 and the SNR (signal to noise ratio) is over 20 dB, where:

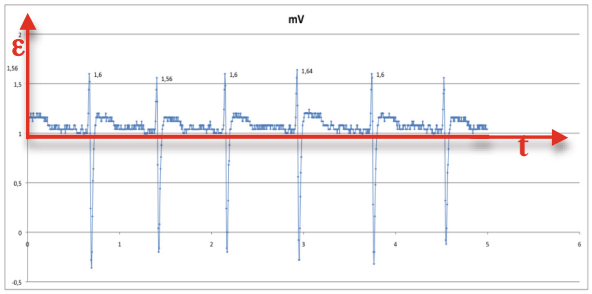
a)

X Value	Original Data	Fourier Approximation	Cosine Coefficients		Sine Coefficients		Frequencies
0	0.625	.42084	.45441	☑			0
1	0.675	.45153	.37475	☑	.26192	☑	(1*2π)/17
2	0.75	.5922	-.25517	☑	-.05949	☑	(2*2π)/17
3	0.9	.94409	.06679	☐	-.04856	☐	(3*2π)/17
4	1	1.25021	-.00298	☐	.00972	☐	(4*2π)/17
5	0.95	1.29622	-.01453	☐	-.00177	☐	(5*2π)/17
6	0.625	.81241	.00719	☐	.00989	☐	(6*2π)/17
7	0.15	.02632	.00158	☐	-.00333	☐	(7*2π)/17
8	-0.175	-.48417	-.00705	☐	-.00098	☐	(8*2π)/17
9	-0.225	-.48792	-.00705	☐	.00098	☐	(9*2π)/17
10	-0.125	-.24792	.00158	☐	.00333	☐	(10*2π)/17

b)



c)



d)

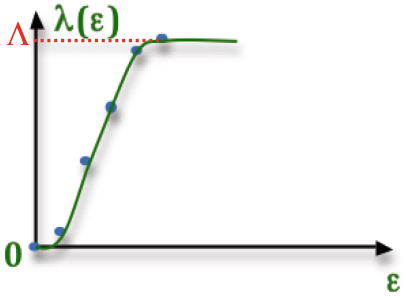


Fig. 2. Top left a): first coefficients of the Fourier transform of the QRS complex: the fundamental and two first harmonics are labeled. Top right b): original QRS complex of the ECG (green) and Fourier (blue) reconstruction with two harmonics matching. Bottom left c): experimental ECG. Bottom right d): evolution of the Lévy time λ(ε) corresponding to the duration time the signal has passed between 0 and ε during 7 cycles (Colour figure online).

$$\text{QRE} = \left(\sum_{i=1,K} (X_i - Y_i)^2 \Big/ \sum_{i=1,K} X_i^2\right)^{1/2} \text{and SNR} = -20\ \text{Log}_{10}\text{QRE} \quad (1)$$

For ECG or the signal of the pulse, the chosen family is the solutions of the Liénard equations, specifically one called van der Pol equation. In the example of the Figs. 2 and 3, the QRS part of the ECG is approximated, after identification and extraction of the values less than $\inf\left(\lambda^{-1}(\{\Lambda\})\right)$ from the inter-beats base line, by using successively Fourier transform and then the family of the limit cycles (having a polynomial approximation [22]) of the van der Pol equation (called Dynalets transform). In the example of Fig. 3, QRE (resp. SNR) is equal to 8 % with Fourier transform and 9 % with Dynalets decomposition (resp. 22 % and 21 %). We can notice that Fourier transform used in Fig. 3 needs 6 parameters (including the value of the period), while the Dynalets transform requires only 5 parameters.

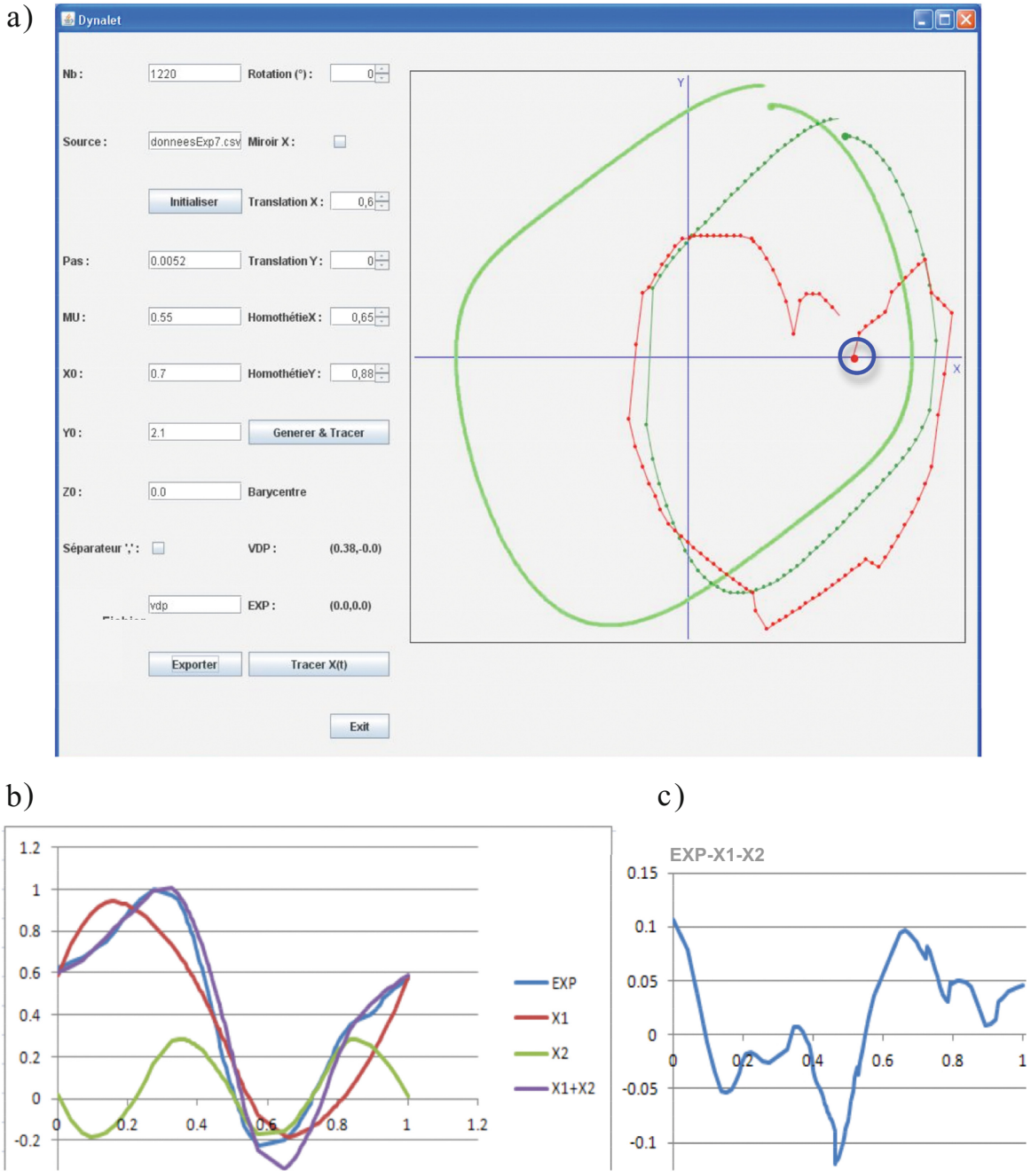


Fig. 3. Top (a): original ECG (red) and van der Pol (green) signal matching, after translation of the origin of the (xOy) referential and localization of the points of the base line (blue circle). Bottom left (b): superposition of the QRS complex of the experimental ECG signal without the base line (EXP in blue), the van der Pol fundamental signal (X1 in red), the first harmonic signal (X2 in green) and the reconstructed fundamental + first harmonic signal (X1 + X2 in violet). Bottom right (c): residual signal equal to EXP-(X1 + X2) (Colour figure online).

3.2 The Detection of the Loss of the Sinus Respiratory Arrhythmia

The cardiac instantaneous period (*i.e.*, the lapse of time between two cardiac beats) is anti-correlated with the time in the inspiration the actual cardiac cycle occurs (Fig. 4), due to a coupling (in the bulb) between the two (respiratory and cardiac) pacemakers.

Fig. 4. Left: the sinus physiologic respiratory arrhythmia characterized by a decrease of the cardiac period during the inspiration. Right: the recording device giving the respiratory and cardiac rhythm from a smart tee-shirt containing a conductor whose impedance varies with chest expansion.

We consider that these pacemakers can be modelled by two van der Pol oscillators, classical examples of regulons (the simplest regulation systems showing one positive and one negative loop inside the graph having their Jacobian matrix as incidence matrix). Indeed, for the sake of simplicity, by neglecting the peripheral Aschow-Tawara node, the cardiac control system is made of two groups of excitable cells, one located in the bulb, composed of neurons and called the cardio-moderator centre CM with electric activity x, which inhibits the sinus node S located in the heart septum, whose activity is denoted by the variable y and results from an autocatalytic loop moderated by CM.

Reciprocally, S activates CM. Then, the van der Pol system representing the rhythmic cardiac activity reads:

$$dx/dt = y,\ dy/dt = -\omega x + \mu(1 - x^2/b^2)y, \tag{2}$$

where μ represents the anharmonic parameter of the oscillator. Because the generator of the heart rate is well modeled by the van der Pol equation, the QRS ECG waveform is well approximated by a sum of solutions of this equation, having as period sub-multiple of the fundamental period. Early detection of neurodegenerative diseases such as Alzheimer's or Parkinson's disease, is then made possible by looking at the gradual disappearance of the inspiratory cardiac acceleration, which serves as an alarm, triggering the consultation in a hospital department of neurology.

4 Actigram Emulation

To emulate persistent behaviors linked to the Alzheimer's disease in residential care settings, we first start from stable behaviors deriving from basic scenarios encountered in everyday life in the home environment [14]. Based on the fact that we all possess

basic needs and patterns (*e.g.*, circadian or nycthemeral activity rhythms [23]), no matter the living conditions and the geographical location, we assumed that these scenarios would remain essentially identical in centers for Alzheimer's. In other words, we basically model the Alzheimer's conditions of life in a healthcare setting putting forward the hypothesis that patients still tend to follow a regular life rhythm based on 24 h (circadian or nycthemeral), such as in familiar environment at-home. Then, we slightly modify these scenarios to reveal persistent behaviors at the spatial-temporal level setting abnormal prolonged periods staying in a room or performing a daily routine.

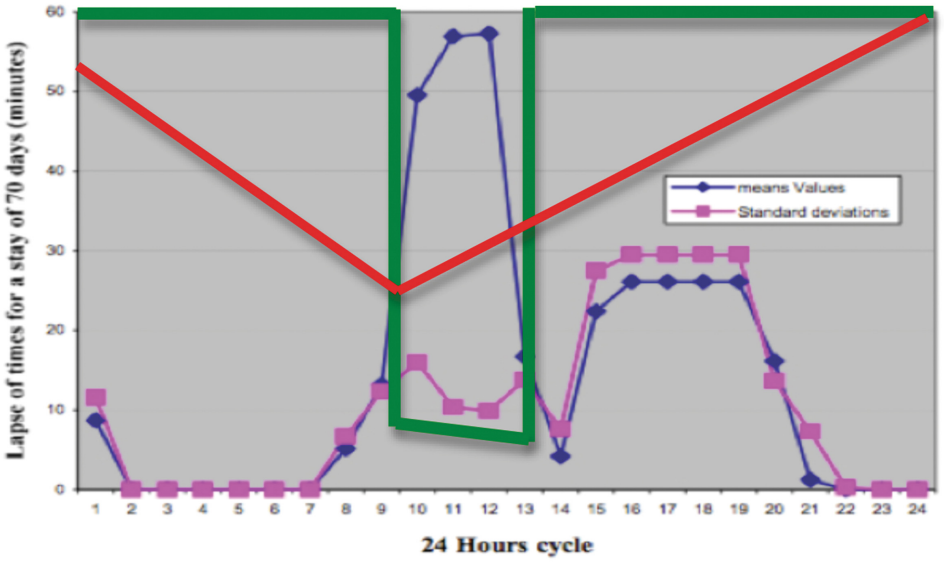


Fig. 5. Expectation (in blue), standard deviation (in pink) and variation coefficient (physiologic in green and pathologic in red) of the lapse of times passed in the kitchen at different hours of the nychthemeron, calculated after 70 days of observation (Colour figure online).

4.1 Data Modeling

To model usual scenarios of daily living progressing toward persistent behaviors, we have used homogeneous Markov chain model, which is a sequential method quite adapted to describe resident's successive room occupancies (or activities) in a home or by extension in a healthcare facility. After 70 days of observation in such a residency for elderly people, we get the statistics given on Fig. 5 for staying in the different rooms of the observed smart flat in which different sensors record the activity of the dependent early person. These statistics allow to calculate different temporal or histogram profiles assigning the observed person in different clusters corresponding to a normal or pathologic behavior [23, 24]. Alarms are triggered when passing from a normal type of nycthemeral activity to a pathologic one. For example, we model the phenomenon of persistence in an activity by setting atypical extended occupancy periods in a room.

4.2 Modeling Persistent Behaviors with Memory Parameters

Modeling persistence with memory effects consists in taking into account the time elapsed in a state before to determine the outcome of the next trial. In other words, trials

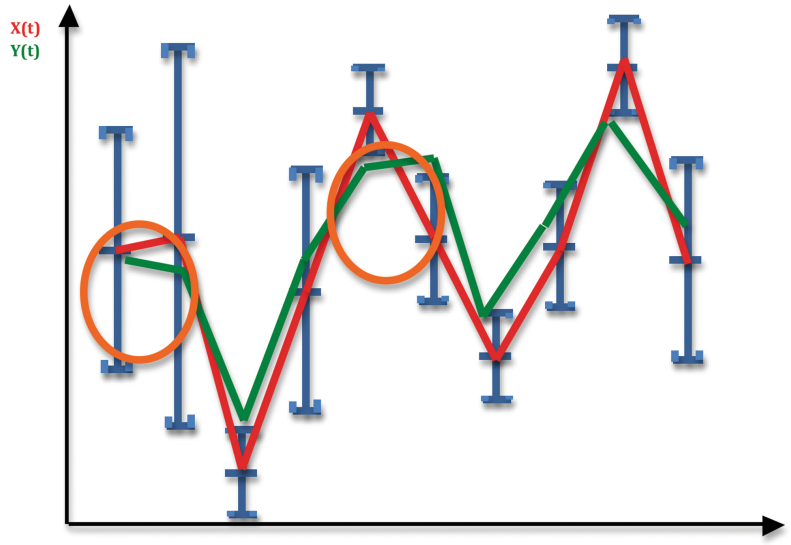


Fig. 6. Temporal profiles of signals X and Y (red and green) over time t, indicating monotonic segments between successive averages, each average being the center of an empirical 95 % confidence interval (Colour figure online).

are not independent. Let's consider a Polya's urn model using first a negative law of parameter $\alpha = 1$. This provokes an increasing waiting time Y (to stay in a state) as the trials go along without any success.

If W and B denote respectively a white and a black ball numbers, and k the number of successes (*i.e.*, we have drawn consecutively $k + 1$ black balls before drawing a white ball), then (3) gives the probability to stay in the same state (a room or an activity):

$$\begin{aligned} P(Y = k+1) &= \left(\frac{B}{B+W}\right) \times \left(\frac{B+1}{B+1+W}\right) \cdots \left(\frac{B+k}{B+k+W}\right)\left(\frac{W}{B+k+1+W}\right) \\ &= \left(\frac{(B+k)!(B)}{(B-1)!} \times \frac{W}{(B+k+1+W)!}\right) \end{aligned} \tag{3}$$

The reverse effect (the more we stay in the room, the more we have some chance to leave it) can be thus obtained using a negative law of parameter $\alpha < 0$ (2) whose expectancy $E(Y)$ is given by (4) and by (6) using Stirling approximation (5) for large factorials:

$$E(Y) = \sum_{k=0}^{\infty} k\left(\frac{B}{B+W}\right) \cdots \left(\frac{B+k\alpha}{B+k\alpha+W}\right)\left(\frac{W}{B+(k+1)\alpha+W}\right) \tag{4}$$

$$If\ K = -[-k\alpha]\ and\ \mathrm{M}! \approx \frac{\mathrm{M}^{\mathrm{M}}\sqrt{2\pi\mathrm{M}}}{\mathrm{e}^{\mathrm{M}}} \tag{5}$$

$$E(Y) \approx \sum_{k=0,\infty} \frac{k(B+W-1)^{B+W-1}\sqrt{2\pi(B+W-1)}e^{B+K+W}W(B+K)!}{e^{B+W-1}(B+K+W)^{B+K+W}\sqrt{2\pi(B+K+W)}(B+K+\alpha+W)(B-1)!} \tag{6}$$

Considering B >> W, permits to simplify (6) in (7):

$$E(Y) \approx \sum_{k=0,\infty} k(1-(k+1)\alpha/B)W/B \tag{7}$$

4.3 Alarm Triggering of Persistence or Perseveration

We focus here on persistent behaviors causing an excessive occupancy periods in the same state (i.e., a room or possibly a well identified daily activity), or corresponding to the persevering repetition such as often opening and closing the doors or checking the contents of the refrigerator without necessity. These abnormal behaviors can be detected when room occupancy curves (Fig. 5) show a profile significantly different from classical temporal profiles of activity corresponding to the profiles cluster to which the patient belongs. A way to compare different temporal evolutions of the room occupancies is to test if the signature of monotony of a temporal profile (*i.e.*, the succession of signs + for an increasing interval and – for a decreasing one) is significantly different from a reference profile (Fig. 6). We can compare the signatures of two signals X (observed) and Y (reference) and test the similarity of their signatures, against a random choice of signs, using the probability P_ to decrease from x to y (Fig. 7):

$$\text{if } D_2 \geq \delta,\ P\_ = \int_{\sup(0,D1)}^{\inf(\delta,D2)} (x - D_1)dx/\delta D = \int_{\sup(0,D1)}^{\inf(\delta,D2)} f(x)F(x)dx, \tag{8}$$

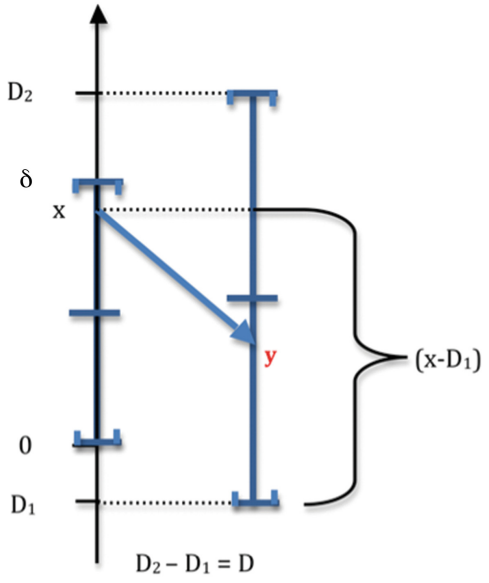


Fig. 7. Calculation of the probability P_ of negative monotony.

If $D_2 < \delta$, $P\_ = 1 - P_+ = 1 - \int_{\sup(0,D1)}^{\inf(\delta,D2)} dx/\delta D$, where f (resp. F) is the uniform density function (resp. cumulative distribution function).

If the law of errors is Gaussian, (8) becomes:

$$P\_ = \int_{-\infty}^{+\infty} f_1(x)F_2(ax+b)dx \tag{9}$$

where f_1 (resp. F_2) is the density function (resp. cumulative distribution function) of the first (resp. second) error on x, *i.e.*, of the Gaussian law $N(\mu_1,\sigma_1)$ (resp. $N(\mu_2,\sigma_2)$), where $a = \sigma_1/\sigma_2$ and $b = (\mu_1-\mu_2)/\sigma_2$. Because the cumulative distribution functions of uniform law on $[-2\sigma, 2\sigma]$ and Gaussian law $N(0,\sigma)$ are very close, results concerning calculation of $P\_$ are similar (Fig. 8).

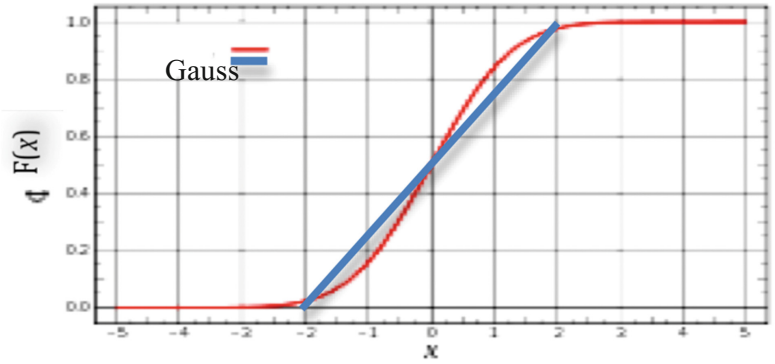


Fig. 8. Cumulative distribution function of the uniform (blue) and Gaussian (red) laws (Colour figure online)

There are 6 different cases:

There are 6 different cases:

1) $D_1<0\leq\delta\leq D_2$	$P\_=[(\delta-D_1)^2-D_1^2]/2\delta D = \delta/2D-D_1/D$	I
2) $D_1<0\leq D_2<\delta$	$P\_=1-D_2^2/2\delta D$	II
3) $D_2\geq\delta\geq D_1\geq 0$	$P\_=(\delta-D_1)^2/2\delta D$	III
4) $0\leq D_1\leq D_2<\delta$	$P\_=1-(D_2^2-D_1^2)/2\delta D = (2\delta-(D_2+D_1))/2\delta$	IV
5) $D_1>\delta$	$P\_=0$	V
6) $D_2<0$	$P\_=1$	VI

The first confidence interval for the observed signal equals $[0, \delta]$. The second to be compared to the first for deciding if there will be a decrease equals $[D_1,D_2]$. P_ is just the probability to go from x in $[0, \delta]$ to $y \leq x$ in $[D_1,D_2]$. On Fig. 6, the probability of decay of the observed signal X(t) is equal to 0.4 for the first interval of monotony, 1 for the second and the fifth (circled in orange in Fig. 6), the sixth and the ninth, and the probability of growth of X(t) is equal to 0 for the third, fourth, seventh and eighth. The values of X(t) at different time points are assumed to be stochastically independent. Let denote by **P** the probability of having at most 1.4 difference into the signs of monotony between the

observed signal X(t) and the reference signal Y(t) like in Fig. 6. Suppose that the event (+, −, +, +, −, −, +, +, −) is obtained by pure chance (0.5) of having the allowed sign at any monotony interval. Then the probability **P** in this hypothesis H0 is equal to:

$$0.4\ C_9^2/2^9 + \ C_9^1/2^9 + \ C_9^0/2^9 = 24.4/2^9 \approx 4.8\,\% \tag{10}$$

We can therefore consider that the probability of rejecting falsely the hypothesis of similarity of monotony due to chance, except one interval at most, is less than 5 %, and say that the similarity of monotony observed between X and Y is not due to a uniform random tossing of signs of monotony, and this with an error less than 5 %. This test is less powerful than a correlation test, but is sufficient in the case of a low number of longitudinal observations in which the amplitude of the signal is not pertinent compared to the signal monotony, for which the variance of the empirical correlation is important. The test requires that reference Y(t) and observed X(t) signals are known at same instants and that the values of X(t) at various time are stochastically independent.

5 Conclusion and Perspectives

Perspectives consider two complementary aspects of the actimetric and physiologic supervision of the person at-home: (i) the ability to record noninvasively different physiological parameters should allow, in the future, documenting automatically after a suitable filter (because the data volume is very large) the personalized medical record of the person followed at-home, which would greatly facilitate its updating, so enhance its pertinence, (ii) the rehabilitation programs in nutrition, locomotion and perception using the serious games allow, thanks to the bio-feedback inherent in these games, to integrate many individual data sources of customization and alarms. Eventually, through their ability of patient empowerment and the virtuous circle that can be installed between the person and the e-educator, these e-learning programs might be the best informational framework to improve the “health trail” during the lifetime of a person.

References

1. World Population Ageing 2009. Department of economical and social affairs report, United Nations publication, New York (2010)
2. Jamison, D.T., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D.B., Jha, P., Mills, A., Musgrove, P.: Disease Control Priorities in Developing Countries. World Bank, Washington (2006)
3. Demongeot, J., Virone, G., Duchêne, F., Benchetrit, G., Hervé, T., Noury, N., Rialle, V.: Multi-sensors acquisition, data fusion, knowledge mining and triggering in health smart homes for elderly people. C.R. Biol. **325**, 673–682 (2002)
4. Virone, G., Noury, N., Demongeot, J.: A system for automatic measurement of circadian activity deviation in telemedicine. IEEE Trans. Biomed. Eng. **49**, 1463–1469 (2002)
5. Abdulrazak, B., Mokhtari, M., Feki, M.A., Ghorbel, M.: Integration of home networking in a smart environment dedicated to people with disabilities. In: ICTTA’04, pp. 125–126. IEEE, Piscataway (2004)

6. Benneyan, J.C.: An introduction to using computer simulation in healthcare: patient wait case study. J. Soc. Health Syst. **5**, 1–15 (1997)
7. Lowery, J.C.: Introduction to simulation in healthcare. In: WSC'96 28th Conference on Winter simulation, pp. 78–84. IEEE Press, Piscataway (1996)
8. O'Connor, C.M., Smith, R., Nott, M.T., Lorang, C., Mathews, R.M.: Using video simulated presence to reduce resistance to care and increase participation of adults with dementia. Am. J. Alzeimer's Dis. Other Dementias **26**, 317–325 (2011)
9. Lowery, J.C., Martin, J.B.: Design and validation of a critical care simulation model. J. Soc. Health Syst. **3**, 15–36 (1992)
10. Gibson, B., Weir, C.: Development and preliminary evaluation of a simulation-based diabetes education module. AMIA Annu. Symp. Proc. **2010**, 246–250 (2010)
11. Berg, D.R., Carlson, A., Durfee, W.K., Sweet, R.M., Reihsen, T.: Low-cost, take-home, beating heart simulator for health-care education. Stud. Health Technol. Inf. **163**, 57–59 (2011)
12. Wong, P., Graves, M.J., Lomas, D.J.: Integrated physiological flow simulator and pulse sequence monitoring system for MRI. Med. Biol. Eng. Comput. **46**, 399–406 (2008)
13. Mahmoud, S.M., Akhlaghinia, M.J., Lotfi, A., Langensiepen, C.: Trend modelling of elderly lifestyle within an occupancy simulator. In: UKSim'11 International Conference on Computer Modelling and Simulation, Cambridge, pp. 156–161. IEEE Press, Piscataway (2011)
14. Virone, G., Lefebvre, B., Noury, N., Demongeot, J.: Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system. In: HealthCom'03, pp. 118–127. IEEE Press, Piscataway (2003)
15. Nabih, K., Gomaa, M.M., Osman, H.S., Aly, G.M.: Modeling, simulation, and control of smart homes using petri nets. Int. J. Smart Home **5**, 1 (2011)
16. Cardinaux, F., Brownsell, S., Hawley, M.S., Bradley, D.: A home daily activity simulation model for the evaluation of lifestyle monitoring systems. Comput. Biol. Med. **43**, 1428–1436 (2013)
17. Lazovik, A., Kaldeli, E., Lazovik, E., Aiello, M.: Planning in a smart home: visualization and simulation. In: ICAPS'09, pp. 13–16. AAAI Press, Menlo Park (2009)
18. Poland, M.P., Nugent, C.D., Wang, H., Chen, L.: Development of a smart home simulator for use as a heuristic tool for management of sensor distribution. Techn. Health Care **17**, 171–182 (2009)
19. Virone, G., Istrate, D.: Integration of an environmental sound module to an existing in-home activity simulator. In: 29th IEEE-EMBS Engineering in Medicine and Biology Society) Microtechnologies in Medicine & Biology, pp. 3810–3813. IEEE, Piscataway (2007)
20. Istrate, D., Castelli, E.: Information extraction from sound for medical telemonitoring. IEEE Trans. Inf Technol. Biomed. **10**, 264–274 (2006)
21. Albæripæ, W.: Tractatus Numerorum a ternario usque ad duodenarium, ad Thomam monachum. Manu- scrit incunable, Médiathèque de Troyes, manuscript 969, l° 195 (vers 1180)
22. Demongeot, J., Françoise, J.P.: Approximation for limit cycles. C. R. Biol. **329**, 967–970 (2006)
23. Virone, G., Vuillerme, N., Mokhtari, M., Demongeot, J.: Persistent behaviour in healthcare facilities: from actimetric tele-surveillance to therapy education. In: Mellouk, A., Fowler, S., Hoceini, S., Daachi, B. (eds.) WWIC 2014. LNCS, vol. 8458, pp. 297–311. Springer, Heidelberg (2014)
24. Franco, C., Fleury, A., Guméry, P.Y., Diot, B., Demongeot, J., Vuillerme, N.: iBalance-ABF: a smartphone-based audio-biofeedback balance system. IEEE Tr. Biomed. Eng **60**, 211–215 (2013)

Conclusions et perspectives

Les travaux de cette thèse concernent la modélisation de l'information protéique, ainsi que la modélisation des recombinaisons de segments de gènes.

Dans le premier chapitre, nous avons vu que le modèle stochastique proposé montre que l'accessibilité des gènes à des réarrangements est fonction des vitesses d'ouverture de la double chaîne d'ADN. Cette modélisation met donc en exergue la dynamique du répertoire humain thymique.

Dans le deuxième chapitre, nous montrons que les architectures de réseaux de régulation biologiques impliqués dans le fonctionnement du système immunitaire (les "immunetworks") contiennent souvent des circuits qui interagissent par leurs intersections. Delbrück (1949) considère que les attracteurs de réseaux théoriques représentent les différents types cellulaires contrôlés par les réseaux de régulation génétique. Dans la même optique, nous mettons l'accent sur le nombre d'attracteurs des immunetworks, qui représentent la diversité des récepteurs du répertoire des lymphocytes T. Plus précisément, nous soulignons l'importance du rôle des intersections de circuits sur la dynamique de ces réseaux. De plus, nous avons déterminé les interactions critiques, ainsi que les configurations des états de gènes permettant de comprendre la défectuosité de la morphogénèse des voies biliaires, dans le cas de l'atrésie biliaire.

Dans le dernier chapitre, nous proposons une autre alternative pour étudier l'information protéique. Il s'agit de l'approche dite "Dynalet". Elle est adaptée à des signaux biologiques non-symétriques, car ils correspondent à des pics de relaxation de l'état protéine, après mise en excitation dans un champ magnétique. Nous avons utilisé cette méthode pour représenter, de manière comprimée, des pics de spectroscopie RMN et d'enregistrements ECG. A l'issue de ce travail, nous proposons des sons caractéristiques des 20 acides aminés et des 4 bases nucléiques. Nous proposons de nommer cette démarche "stéthoscope protéique".

Dans la continuité de ces travaux, nous suggérons d'étudier les autres réseaux impliqués dans la réponse immunitaire, que ce soit la réponse adaptative ou la réponse innée. Ainsi, l'étude des recombinaisons peut être poursuivie par l'intégration de tous les segments V et J. Le modèle déjà établi chez l'homme pourra être ainsi généralisé et le répertoire thymique du nouveau-né comparé au répertoire périphérique de l'adulte.

De façon plus générale, l'analyse des réseaux de régulation génétique, en comparant les individus sains et malades, permet de mettre l'accent sur des profils spécifiques des patients en termes d'expression de gènes. Cette étude pourra intervenir dans la prise en charge des personnes atteintes de diverses pathologies (cancer, maladies neuro-dégénératives,...), tant au niveau diagnostique, qu'au niveau du suivi thérapeutique.

La méthode de traitement et de restitution sous forme sonore du signal spectroscopique RMN des protéines appelée Dynalet, devra être généralisée, pour obtenir la signature sonore des protéines courantes, sous la forme de la succession des sons correspondant à leurs acides aminés. De même, on peut envisager d'obtenir la signature sonore des acides nucléiques (ARN ou ADN), en construisant la "mélodie" faite de la succession des sons correspondant à leurs bases nucléiques.

Bibliographie

[1] C. Aude-Garcia, M. Gallagher, P. N. Marche, and E. Jouvin-Marche, "Preferential adv-aj association during recombination in the mouse t-cell receptor alpha/delta locus," *Immunogenetics*, vol. 52, no. 3-4, pp. 224–230, 2001.

[2] S. F. Altschul, E. M. Gertz, R. Agarwala, A. A. Schäffer, and Y.-K. Yu, "Psi-blast pseudocounts and the minimum description length principle," *Nucleic acids research*, vol. 37, no. 3, pp. 815–824, 2009.

[3] T. P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky, "A direct estimate of the human $\alpha\beta$ t cell receptor diversity," *Science*, vol. 286, no. 5441, pp. 958–961, 1999.

[4] T. Azuma, V. Igras, E. B. Reilly, and H. N. Eisen, "Diversity at the variable-joining region boundary of lambda light chains has a pronounced effect on immunoglobulin ligand-binding activity," *Proceedings of the National Academy of Sciences*, vol. 81, no. 19, pp. 6139–6143, 1984.

[5] T.-P. Baum, N. Pasqual, F. Thuderoz, V. Hierle, D. Chaume, M.-P. Lefranc, E. Jouvin-Marche, P.-N. Marche, and J. Demongeot, "Imgt/geneinfo: enhancing v (d) j recombination database accessibility," *Nucleic acids research*, vol. 32, no. suppl 1, pp. D51–D54, 2004.

[6] C. H. Bassing, W. Swat, and F. W. Alt, "The mechanism and regulation of chromosomal v (d) j recombination," *Cell*, vol. 109, no. 2, pp. S45–S55, 2002.

[7] K. Bleakley, M.-P. Lefranc, and G. Biau, "Recovering probabilities for nucleotide trimming processes for t cell receptor tra and trg vj junctions analyzed with imgt tools," *BMC bioinformatics*, vol. 9, no. 1, p. 408, 2008.

[8] C. Coleclough, "Chance, necessity and antibody gene dynamics.," *Nature*, vol. 303, no. 5912, pp. 23–26, 1982.

[9] J.-L. Casanova, P. Romero, C. Widmann, P. Kourilsky, and J. Maryanski, "T cell receptor genes in a series of class i major histocompatibility complex-restricted cytotoxic t lymphocyte clones specific for a plasmodium berghei nonapeptide: implications for t cell allelic exclusion and antigen-specific repertoire.," *The Journal of experimental medicine*, vol. 174, no. 6, pp. 1371–1383, 1991.

[10] F. Davodeau, M. Difilippantonio, E. Roldan, M. Malissen, J.-L. Casanova, C. Couedel, J.-F. Morcet, M. Merkenschlager, A. Nussenzweig, M. Bonneville, *et al.*, "The tight interallelic positional coincidence that distinguishes t-cell receptor jα usage does not result from homologous chromosomal pairing during vαjα rearrangement," *The EMBO journal*, vol. 20, no. 17, pp. 4717–4729, 2001.

[11] P. Fuschiotti, N. Pasqual, V. Hierle, E. Borel, J. London, P. N. Marche, and E. Jouvin-Marche, "Analysis of the tcr α-chain rearrangement profile in human t lymphocytes," *Molecular immunology*, vol. 44, no. 13, pp. 3380–3388, 2007.

[12] H. Gahéry-Ségard, E. Jouvin-Marche, A. Six, C. Gris-Liebe, M. Malissen, B. Malissen, P.-A. Cazenave, and P. N. Marche, "Germline genomic structure of the b10. a mousetcra-v2 gene subfamily," *Immunogenetics*, vol. 44, no. 4, pp. 298–305, 1996.

[13] P. N. el Marche, "Differential chronology of tcradv2 gene use by> and p chains of the mouse tcr," *Eur. J. Immunol*, vol. 28, pp. 818–827, 1998.

[14] A. Hawwari, C. Bock, and M. S. Krangel, "Regulation of t cell receptor α gene assembly by a complex hierarchy of germline jα promoters," *Nature immunology*, vol. 6, no. 5, pp. 481–489, 2005.

[15] S. Gilfillan, A. Dierich, M. Lemeur, C. Benoist, and D. Mathis, "Mice lacking tdt: mature animals with an immature lymphocyte repertoire," *Science*, vol. 261, no. 5125, pp. 1175–1178, 1993.

[16] M. S. Krangel, "Mechanics of t cell receptor gene rearrangement," *Current opinion in immunology*, vol. 21, no. 2, pp. 133–139, 2009.

[17] M. S. Krangel, J. Carabana, I. Abbarategui, R. Schlimgen, and A. Hawwari, "Enforcing order within a complex locus: current perspectives on the control of v (d) j recombination at the murine t-cell receptor α/δ locus," *Immunological reviews*, vol. 200, no. 1, pp. 224–232, 2004.

[18] M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, J. Lane, *et al.*, "Imgt®, the international immunogenetics information system®," *Nucleic acids research*, vol. 37, no. suppl 1, pp. D1006–D1012, 2009.

[19] E. Jouvin-Marche, P. Fuschiotti, and P. N. Marche, "Dynamic aspects of tcrα gene recombination: qualitative and quantitative assessments of the tcrα chain repertoire in man and mouse," in *V (D) J Recombination*, pp. 82–92, Springer, 2009.

[20] M. Malissen, J. Trucy, E. Jouvin-Marche, P.-A. Cazenave, R. Scollay, and B. Malissen, "Regulation of tcr α and β gene allelic exclusion during t-cell development," *Immunology today*, vol. 13, no. 8, pp. 315–322, 1992.

[21] M. A. Oettinger, D. G. Schatz, C. Gorka, and D. Baltimore, "Rag-1 and rag-2, adjacent genes that synergistically activate v (d) j recombination," *Science*, vol. 248, no. 4962, pp. 1517–1523, 1990.

[22] N. Pasqual, M. Gallagher, C. Aude-Garcia, M. Loiodice, F. Thuderoz, J. Demongeot, R. Ceredig, P. N. Marche, and E. Jouvin-Marche, "Quantitative and qualitative changes in vj α rearrangements during mouse thymocytes differentiation implication for a limited t cell receptor α chain repertoire," *The Journal of experimental medicine*, vol. 196, no. 9, pp. 1163–1174, 2002.

[23] H. Petrie, F. Livak, D. Schatz, A. Strasser, I. N. Crispe, and K. Shortman, "Multiple rearrangements in t cell receptor alpha chain genes maximize the production of useful thymocytes.," *The Journal of experimental medicine*, vol. 178, no. 2, pp. 615–622, 1993.

[24] J. P. Rast, M. K. Anderson, S. J. Strong, C. Luer, R. T. Litman, and G. W. Litman, "α, β, γ, and δ t cell antigen receptor genes arose early in vertebrate phylogeny," *Immunity*, vol. 6, no. 1, pp. 1–11, 1997.

[25] S. Takeshita, M. Toda, and H. Yamagishi, "Excision products of the t cell receptor gene support a progressive rearrangement model of the alpha/delta locus.," *The EMBO journal*, vol. 8, no. 11, p. 3261, 1989.

[26] S. D. Thompson, J. Pelkonen, M. Rytkönen, J. Samaridis, and J. L. Hurwitz, "Nonrandom rearrangement of t cell receptor j alpha genes in bone marrow t cell differentiation cultures.," *The Journal of Immunology*, vol. 144, no. 7, pp. 2829–2834, 1990.

[27] F. Thuderoz, M.-A. Simonet, O. Hansen, N. Pasqual, A. Dariz, T. P. Baum, V. Hierle, J. Demongeot, P. N. Marche, and E. Jouvin-Marche, "Numerical modelling of the vj combinations of the t cell receptor tra/trd locus," *PLoS Comput Biol*, vol. 6, no. 2, pp. e1000682–e1000682, 2010.

[28] H. Uenishi, H. Hiraiwa, R. Yamamoto, H. Yasue, Y. Takagaki, T. Shiina, E. Kikkawa, H. Inoko, and T. Awata, "Genomic structure around joining segments and constant regions of swine t-cell receptor α/δ (tra/trd) locus," *Immunology*, vol. 109, no. 4, pp. 515–526, 2003.

[29] I. Villey, D. Caillol, F. Selz, P. Ferrier, and J.-P. de Villartay, "Defect in rearrangement of the most 5? tcr–jα following targeted deletion of t early α (tea): Implications for tcr α locus accessibility," *Immunity*, vol. 5, no. 4, pp. 331–342, 1996.

[30] A. Warmflash and A. R. Dinner, "A model for tcr gene segment use," *The Journal of Immunology*, vol. 177, no. 6, pp. 3857–3864, 2006.

[31] A. Wilson, H. R. MacDonald, and F. Radtke, "Notch 1–deficient common lymphoid precursors adopt a b cell fate in the thymus," *The Journal of experimental medicine*, vol. 194, no. 7, pp. 1003–1012, 2001.

[32] C. Antonopoulos, V. Basios, J. Demongeot, P. Nardone, and R. Thomas, "Linear and nonlinear arabesques: a study of closed chains of negative 2-element circuits," *International Journal of Bifurcation and Chaos*, vol. 23, no. 09, p. 1330033, 2013.

[33] A. Aubert, R. Costalat, P. J. Magistretti, and L. Pellerin, "Brain lactate kinetics: modeling evidence for neuronal lactate uptake upon activation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 45, pp. 16448–16453, 2005.

[34] A. Aubert, L. Pellerin, P. J. Magistretti, and R. Costalat, "A coherent neurobiological framework for functional neuroimaging provided by a model integrating compartmentalized energy metabolism," *Proceedings of the National Academy of Sciences*, vol. 104, no. 10, pp. 4188–4193, 2007.

[35] S. Bandiera, S. Ruberg, M. Girard, N. Cagnard, S. Hanein, D. Chrétien, A. Munnich, S. Lyonnet, and A. Henrion-Caude, "Nuclear outsourcing of rna interference components to human mitochondria," *PloS one*, vol. 6, no. 6, p. e20746, 2011.

[36] S. Bandiera, R. Matégot, M. Girard, J. Demongeot, and A. Henrion-Caude, "Mitomirs delineating the intracellular localization of micrornas at mitochondria," *Free Radical Biology and Medicine*, vol. 64, pp. 12–19, 2013.

[37] E. Barrey, G. Saint-Auret, B. Bonnamy, D. Damas, O. Boyer, and X. Gidrol, "Pre-microrna and mature microrna in human mitochondria," *PloS one*, vol. 6, no. 5, p. e20220, 2011.

[38] M. Bier, B. Teusink, B. N. Kholodenko, and H. V. Westerhoff, "Control analysis of glycolytic oscillations," *Biophysical chemistry*, vol. 62, no. 1, pp. 15–24, 1996.

[39] A. BoITEUX, A. Goldbeter, and B. Hess, "Control of oscillating glycolysis of yeast by stochastic, periodic, and steady source of substrate: a model and experimental study," *Proceedings of the National Academy of Sciences*, vol. 72, no. 10, pp. 3829–3833, 1975.

[40] P. Cui, R. Ji, F. Ding, D. Qi, H. Gao, H. Meng, J. Yu, S. Hu, and H. Zhang, "A complete mitochondrial genome sequence of the wild two-humped camel (camelus bactrianus ferus): an evolutionary history of camelidae," *BMC genomics*, vol. 8, no. 1, p. 241, 2007.

[41] J. Demongeot and A. Doncescu, "Modeling the glycolysis: an inverse problem approach," in *Advanced Information Networking and Applications Workshops, 2009. WAINA'09. International Conference on*, pp. 930–935, IEEE, 2009.

[42] J. Demongeot and S. Sené, "Asymptotic behavior and phase transition in regulatory networks. ii simulations," *Neural Networks*, vol. 21, pp. 971–979, 2008.

[43] J. Demongeot and F. Seydoux, "Oscillations glycolytiques: Modélisation d'un système minimum à partir des données physiologiques et moléculaires," *Elaboration et Justification de Modèles*, pp. 519–536, 1979.

[44] J. Demongeot and J. Waku, "Robustness in biological regulatory networks i: Mathematical approach," *Comptes Rendus Mathématique*, vol. 350, no. 3, pp. 221–224, 2012.

[45] J. Demongeot, N. Glade, A. Moreira, and L. Vial, "Rna relics and origin of life," *International journal of molecular sciences*, vol. 10, no. 8, pp. 3420–3441, 2009.

[46] J. Demongeot, M. Noual, and S. Sené, "Combinatorics of boolean automata circuits dynamics," *Discrete Applied Mathematics*, vol. 160, no. 4, pp. 398–415, 2012.

[47] J. Demongeot, H. Hazgui, and N. Villerme, "Micrornas: unspecific inhibitory regulation in immunologic control and in mitochondrial respiration," in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, pp. 1509–1516, IEEE, 2013.

[48] G. Farquhar, S. v. von Caemmerer, and J. Berry, "A biochemical model of photosynthetic co2 assimilation in leaves of c3 species," *Planta*, vol. 149, no. 1, pp. 78–90, 1980.

[49] S. Genc, I. A. Kurnaz, and M. Ozilgen, "Astrocyte-neuron lactate shuttle may boost more atp supply to the neuron under hypoxic conditions-in silico study supported by in vitro expression data," *BMC systems biology*, vol. 5, no. 1, p. 162, 2011.

[50] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman, "Rfam: annotating non-coding rnas in complete genomes," *Nucleic acids research*, vol. 33, no. suppl 1, pp. D121–D124, 2005.

[51] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, pp. C47–C52, 1999.

[52] J. Hervagault, M. Duban, J. Kernevez, and D. Thomas, "Multiple steady states and oscillatory behavior of a compartmentalized phosphofructokinase system," *Proceedings of the National Academy of Sciences*, vol. 80, no. 18, pp. 5455–5459, 1983.

[53] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[54] H. Kacser, , and J. Burns, "The control of flux," in *Symp. Soc. Exp. Biol.*, vol. 27, pp. 65–104, 1973.

[55] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of theoretical biology*, vol. 22, no. 3, pp. 437–467, 1969.

[56] B. Liu, N. Zhang, Z. Liu, Y. Fu, S. Feng, S. Wang, Y. Cao, D. Li, D. Liang, F. Li, *et al.*, "Rp105 involved in activation of mouse macrophages via tlr2 and tlr4 signaling," *Molecular and cellular biochemistry*, vol. 378, no. 1-2, pp. 183–193, 2013.

[57] K. Miyake, H. Ogata, Y. Nagai, S. Akashi, and M. Kimoto, "Innate recognition of lipopolysaccharide by toll-like receptor 4/md-2 and rp105/md-1," *Journal of endotoxin research*, vol. 6, no. 5, pp. 389–391, 2000.

[58] J. Ovádi, "Old pathway—new concept: control of glycolysis by metabolite-modulated dynamic enzyme associations," *Trends in biochemical sciences*, vol. 13, no. 12, pp. 486–490, 1988.

[59] C. Reder, "Metabolic control theory: a structural approach," *Journal of Theoretical Biology*, vol. 135, no. 2, pp. 175–201, 1988.

[60] J. B. Ritter, Y. Genzel, and U. Reichl, "Simultaneous extraction of several metabolites of energy metabolism and related substances in mammalian cells: optimization using experimental design," *Analytical biochemistry*, vol. 373, no. 2, pp. 349–369, 2008.

[61] P. Ruoff, M. K. Christensen, J. Wolf, and R. Heinrich, "Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations," *Biophysical chemistry*, vol. 106, no. 2, pp. 179–192, 2003.

[62] E. Sbisà, F. Tanzariello, A. Reyes, G. Pesole, and C. Saccone, "Mammalian mitochondrial d-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications," *Gene*, vol. 205, no. 1, pp. 125–140, 1997.

[63] A. Surono, Y. Takeshima, T. Wibawa, M. Ikezawa, I. Nonaka, and M. Matsuo, "Circular dystrophin rnas consisting of exons that were skipped by alternative splicing," *Human molecular genetics*, vol. 8, no. 3, pp. 493–500, 1999.

[64] R. Thomas, "Boolean formalisation of genetic control circuits," *Journal of theoretical biology*, vol. 42, pp. 565–583, 1973.

[65] D. C. Weaver, C. T. Workman, G. D. Stormo, *et al.*, "Modeling regulatory networks with weight matrices.," in *Pacific symposium on biocomputing*, vol. 4, pp. 112–123, World Scientific, 1999.

[66] E. R. Weibel, C. R. Taylor, and H. Hoppeler, "The concept of symmorphosis: a testable hypothesis of structure-function relationship," *Proceedings of the National Academy of Sciences*, vol. 88, no. 22, pp. 10357–10361, 1991.

[67] J. Wolf and R. Heinrich, "Effect of cellular interaction on glycolytic oscillations in yeast: a theoretical investigation," *Biochem. J*, vol. 345, pp. 321–334, 2000.

[68] D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, and P. Musgrove, *Disease control priorities in developing countries*. World Bank Publications, 2006.

[69] J. Demongeot, G. Virone, F. Duchêne, G. Benchetrit, T. Hervé, N. Noury, and V. Rialle, "Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people," *Comptes Rendus Biologies*, vol. 325, no. 6, pp. 673–682, 2002.

[70] G. Virone, N. Noury, and J. Demongeot, "A system for automatic measurement of circadian activity deviations in telemedicine," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 12, pp. 1463–1469, 2002.

[71] B. Abdulrazak, M. Mokhtari, M. A. Feki, and M. Ghorbel, "Integration of home networking in a smart environment dedicated to people with disabilities," in *Information and Communication Technologies: From Theory to Applications, 2004. Proceedings. 2004 International Conference on*, pp. 125–126, IEEE, 2004.

[72] J. C. Benneyan, "An introduction to using computer simulation in healthcare: patient wait case study," *Journal of the Society for Health Systems*, vol. 5, no. 3, pp. 1–15, 1997.

[73] J. C. Lowery, "Introduction to simulation in health care," in *Proceedings of the 28th conference on Winter simulation*, pp. 78–84, IEEE Computer Society, 1996.

[74] C. O'Connor, R. Smith, M. Nott, C. Lorang, and R. Mathews, "Using video simulated presence to reduce resistance to care and increase participation of adults with dementia," *American journal of Alzheimer's disease and other dementias*, vol. 26, no. 4, pp. 317–325, 2011.

[75] J. C. Lowery and J. B. Martin, "Design and validation of a critical care simulation model.," *Journal of the Society for Health Systems*, vol. 3, no. 3, pp. 15–36, 1991.

[76] B. Gibson and C. Weir, "Development and preliminary evaluation of a simulation-based diabetes education module," in *AMIA Annual Symposium Proceedings*, vol. 2010, p. 246, American Medical Informatics Association, 2010.

[77] D. R. Berg, A. Carlson, W. K. Durfee, R. M. Sweet, and T. Reihsen, "Low-cost, take-home, beating heart simulator for health-care education.," *Studies in health technology and informatics*, vol. 163, pp. 57–59, 2010.

[78] P. Wong, M. J. Graves, and D. J. Lomas, "Integrated physiological flow simulator and pulse sequence monitoring system for mri," *Medical & biological engineering & computing*, vol. 46, no. 4, pp. 399–406, 2008.

[79] S. M. Mahmoud, M. J. Akhlaghinia, A. Lotfi, and C. Langensiepen, "Trend modelling of elderly lifestyle within an occupancy simulator," in *Computer Modelling and Simulation (UKSim), 2011 UkSim 13th International Conference on*, pp. 156–161, IEEE, 2011.

[80] G. Virone, B. Lefebvre, N. Noury, and J. Demongeot, "Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system," in *Enterprise Networking and Computing in Healthcare Industry, 2003. Healthcom 2003. Proceedings. 5th International Workshop on*, pp. 111–117, IEEE, 2003.

[81] A. K. Nabih, M. M. Gomaa, H. S. Osman, and G. M. Aly, "Modeling, simulation, and control of smart homes using petri nets," *International Journal of Smart Home*, vol. 5, no. 3, pp. 1–14, 2011.

[82] F. Cardinaux, S. Brownsell, D. Bradley, and M. S. Hawley, "A home daily activity simulation model for the evaluation of lifestyle monitoring systems," *Computers in biology and medicine*, vol. 43, no. 10, pp. 1428–1436, 2013.

[83] A. Lazovik, E. Kaldeli, E. Lazovik, and M. Aiello, "Planning in a smart home: visualization and simulation," in *Application Showcase Proceedings of the 19th (ICAPS) Int. Conf. Automated Planning and Scheduling*, 2009.

[84] M. P. Poland, C. D. Nugent, H. Wang, and L. Chen, "Development of a smart home simulator for use as a heuristic tool for management of sensor distribution," *Technology and Health Care*, vol. 17, no. 3, p. 171, 2009.

[85] G. Virone and D. Istrate, "Integration of an environmental sound module to an existing in-home activity simulator," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 3810–3813, IEEE, 2007.

[86] D. Istrate, E. Castelli, M. Vacher, L. Besacier, and J.-F. Serignat, "Information extraction from sound for medical telemonitoring," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, no. 2, pp. 264–274, 2006.

[87] J. Demongeot, O. Hansen, A. Hamie, H. Hazgui, G. Virone, and N. Vuillerme, "Actimetry@ home: actimetric tele-surveillance and tailored to the signal data compression," in *Smart Homes and Health Telematics*, pp. 59–70, Springer, 2015.

[88] J. Demongeot and J.-P. Françoise, "Approximation for limit cycles and their isochrons," *Comptes rendus biologies*, vol. 329, no. 12, pp. 967–970, 2006.

[89] G. Virone, N. Vuillerme, M. Mokhtari, and J. Demongeot, "Persistent behaviour in healthcare facilities: from actimetric tele-surveillance to therapy education," in *Wired/Wireless Internet Communications*, pp. 297–311, Springer, 2014.

[90] C. Franco, A. Fleury, P.-Y. Guméry, B. Diot, J. Demongeot, and N. Vuillerme, "ibalance-abf: a smartphone-based audio-biofeedback balance system," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 1, pp. 211–215, 2013.

[91] J. Demongeot, A. Henrion-Caude, A. Lontos, and E. Promayon, "General architecture of a genetic regulatory network. applications to embryologic control," in *ECAL 2011, Eleventh European Conference on the Synthesis and Simulation of Living Systems*, pp. 1–8, MIT Press, 2011.

[92] J. Demongeot, E. Drouet, A. Elena, A. Moreira, Y. Rechoum, and S. Sené, "Micrornas: viral genome and robustness of gene expression in the host," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1908, pp. 4941–4965, 2009.

[93] J. Demongeot, A. Elena, M. Noual, S. Sené, and F. Thuderoz, ""immunetworks", intersecting circuits and dynamics," *Journal of theoretical biology*, vol. 280, no. 1, pp. 19–33, 2011.

[94] R. Elkon, C. Linhart, Y. Halperin, Y. Shiloh, and R. Shamir, "Functional genomic delineation of tlr-induced transcriptional networks," *BMC genomics*, vol. 8, no. 1, p. 394, 2007.

[95] P. Bulet, C. Hetru, J.-L. Dimarcq, and D. Hoffmann, "Antimicrobial peptides in insects; structure and function," *Developmental & Comparative Immunology*, vol. 23, no. 4, pp. 329–344, 1999.

[96] G. T. Byrd, R. F. Sage, and R. H. Brown, "A comparison of dark respiration between c3 and c4 plants," *Plant Physiology*, vol. 100, no. 1, pp. 191–198, 1992.

[97] H. Tang, M. Lee, O. Sharpe, L. Salamone, E. J. Noonan, C. D. Hoang, S. Levine, W. H. Robinson, and J. B. Shrager, "Oxidative stress-responsive microrna-320 regulates glycolysis in diverse biological systems," *The FASEB Journal*, vol. 26, no. 11, pp. 4710–4721, 2012.

[98] P. Hainaut and M. Hollstein, "p53 and human cancer: the first ten thousand mutations," *Advances in cancer research*, vol. 2000, no. 77, pp. 81–138, 1999.

[99] L. A. Demetrius and D. K. Simon, "An inverse-warburg effect and the origin of alzheimer's disease," *Biogerontology*, vol. 13, no. 6, pp. 583–594, 2012.

[100] P. C. Davies, L. Demetrius, and J. A. Tuszynski, "Cancer as a dynamical phase transition," *Theor Biol Med Model*, vol. 8, p. 30, 2011.

[101] M. Maurel, S. Jalvy, Y. Ladeiro, C. Combe, L. Vachet, F. Sagliocco, P. Bioulac-Sage, V. Pitard, H. Jacquemin-Sablon, J. Zucman-Rossi, *et al.*, "A functional screening identifies five micrornas controlling glypican-3: role of mir-1271 down-regulation in hepatocellular carcinoma," *Hepatology*, vol. 57, no. 1, pp. 195–204, 2013.

[102] L. Boominathan, "The tumor suppressors p53, p63, and p73 are regulators of microrna processing complex," *PloS one*, vol. 5, no. 5, p. e10615, 2010.

[103] M. Segura, H. Greenwald, D. Hanniford, I. Osman, and E. Hernando, "Microrna and cutaneous melanoma: from discovery to prognosis and therapy.," *Carcinogenesis*, vol. 33, no. 10, pp. 1823–1832, 2012.

[104] D. A. Tennant, R. V. Durán, and E. Gottlieb, "Targeting metabolic transformation for cancer therapy," *Nature Reviews Cancer*, vol. 10, no. 4, pp. 267–277, 2010.

[105] C. Liu, K. Kelnar, B. Liu, X. Chen, T. Calhoun-Davis, H. Li, L. Patrawala, H. Yan, C. Jeter, S. Honorio, *et al.*, "The microrna mir-34a inhibits prostate cancer stem cells and metastasis by directly repressing cd44," *Nature medicine*, vol. 17, no. 2, pp. 211–215, 2011.

[106] F. Blanchini and E. Franco, "Structurally robust biological networks," *BMC systems biology*, vol. 5, no. 1, p. 74, 2011.

[107] J. Demongeot and S. Sené, "The singular power of the environment on nonlinear hopfield networks in: Cmsb11," *ACM Proceedings, New York,(2011b)*, pp. 55–64, 2011.

[108] L. Demetrius, "Statistical mechanics and population biology," *Journal of Statistical Physics*, vol. 30, no. 3, pp. 709–753, 1983.

[109] L. Demetrius, "Directionality principles in thermodynamics and evolution," *Proceedings of the National Academy of Sciences*, vol. 94, no. 8, pp. 3491–3498, 1997.

[110] R. Kühn, "Equilibrium analysis of complex systems," *Lecture Notes 7CCMCS03, King's College, London*, 2010.

[111] A. Lesne, "Robustness: confronting lessons from physics and biology," *Biological Reviews*, vol. 83, no. 4, pp. 509–532, 2008.

[112] J. Gunawardena, "Biological systems theory," *Science*, vol. 328, no. 5978, pp. 581–582, 2010.

[113] C. H. Waddington *et al.*, "Organisers and genes.," *Organisers and Genes.*, 1940.

[114] J. Demongeot, M. Noual, and S. Sené, "On the number of attractors of positive and negative boolean automata circuits," in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pp. 782–789, IEEE, 2010.

[115] J. Demongeot, O. Cohen, A. Doncescu, and A. Henrion-Caude, "Mitomirs and energetic regulation," in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, pp. 1501–1508, IEEE, 2013.

[116] J. Demongeot, "Sur la possibilité de considérer filecode génétique comme un code à enchaînement," *Revue de Biomaths*, vol. 62, pp. 61–66, 1978.

[117] J. Demongeot and J. Besson, "Code génétique et codes à enchaînement i," *CR Acad. Sci. Paris*, vol. 296, pp. 807–810, 1983.

[118] J. Demongeot and J. Besson, "The genetic code and cyclic codes.," *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*, vol. 319, no. 6, pp. 443–451, 1996.

[119] J. Demongeot, J. Aracena, F. Thuderoz, T.-P. Baum, and O. Cohen, "Genetic regulation networks: circuits, regulons and attractors," *Comptes Rendus Biologies*, vol. 326, no. 2, pp. 171–188, 2003.

[120] J. Demongeot and A. Moreira, "A possible circular rna at the origin of life," *Journal of theoretical biology*, vol. 249, no. 2, pp. 314–324, 2007.

[121] T. M. Klingler and D. L. Brutlag, "Detection of correlations in trna sequences with structural implications.," in *ISMB*, pp. 225–233, 1993.

[122] W. Rychlik, W. Spencer, and R. Rhoads, "Optimization of the annealing temperature for dna amplification in vitro," *Nucleic acids research*, vol. 18, no. 21, pp. 6409–6412, 1990.

[123] H. CHOI, K. GABRIEL, J. SCHNEIDER, S. OTTEN, and W. H. McCLAIN, "Recognition of acceptor-stem structure of trnaasp by escherichia coli aspartyl-trna synthetase," *Rna*, vol. 9, no. 4, pp. 386–393, 2003.

[124] N. Shigi, T. Suzuki, M. Tamakoshi, T. Oshima, and K. Watanabe, "Conserved bases in the tψc loop of trna are determinants for thermophile-specific 2-thiouridylation at position 54," *Journal of Biological Chemistry*, vol. 277, no. 42, pp. 39128–39135, 2002.

[125] T. Tanaka and Y. Kikuchi, "Origin of the cloverleaf shape of transfer rna-the double-hairpin model: implication for the role of trna intron and the long extra loop," *Viva Origino*, vol. 29, no. 134, pp. 134–142, 2001.

[126] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir, "Structure of a ribonucleic acid," *Science*, vol. 147, no. 3664, pp. 1462–1465, 1965.

[127] T. B. Hansen, E. D. Wiklund, J. B. Bramsen, S. B. Villadsen, A. L. Statham, S. J. Clark, and J. Kjems, "mirna-dependent gene silencing involving ago2-mediated cleavage of a circular antisense rna," *The EMBO journal*, vol. 30, no. 21, pp. 4414–4422, 2011.

[128] J. Demongeot, O. Cohen, and A. Henrion-Caude, "Micrornas and robustness in biological regulatory networks. a generic approach with applications at different levels: physiologic, metabolic, and genetic," in *Systems Biology of Metabolic and Signaling Networks*, pp. 63–114, Springer, 2014.

[129] J. Demongeot, H. Hazgui, H. B. Amor, and J. Waku, "Stability, complexity and robustness in population dynamics," *Acta biotheoretica*, vol. 62, no. 3, pp. 243–284, 2014.

[130] C. Sangokoya, J. F. Doss, and J.-T. Chi, "Iron-responsive mir-485-3p regulates cellular iron homeostasis by targeting ferroportin," *PLoS Genet*, vol. 9, no. 4, p. e1003408, 2013.

[131] J. Demongeot, E. Goles, M. Morvan, M. Noual, and S. Sené, "Attraction basins as gauges of robustness against boundary conditions in biological complex systems," *PloS one*, vol. 5, no. 8, p. e11793, 2010.

[132] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, *et al.*, "Circular rnas are a large class of animal rnas with regulatory potency," *Nature*, vol. 495, no. 7441, pp. 333–338, 2013.

[133] T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems, "Natural rna circles function as efficient microrna sponges," *Nature*, vol. 495, no. 7441, pp. 384–388, 2013.

[134] X. Li, J. J. Cassidy, C. A. Reinke, S. Fischboeck, and R. W. Carthew, "A microrna imparts robustness against environmental fluctuation during development," *Cell*, vol. 137, no. 2, pp. 273–282, 2009.

[135] G. Weil, K. Heus, T. Faraut, and J. Demongeot, "An archetypal basic code for the primitive genome," *Theoret. Comp. Sc*, vol. 322, pp. 313–334, 2004.

[136] A. Kumar, A. K.-L. Wong, M. L. Tizard, R. J. Moore, and C. Lefèvre, "mirna_targets: A database for mirna target predictions in coding and non-coding regions of mrnas," *Genomics*, vol. 100, no. 6, pp. 352–356, 2012.

[137] V. Pierre, G. Martinez, C. Coutton, J. Delaroche, S. Yassine, C. Novella, K. Pernet-Gallay, S. Hennebicq, P. F. Ray, and C. Arnoult, "Absence of dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus," *Development*, vol. 139, no. 16, pp. 2955–2965, 2012.

[138] J. Escoffier, S. Boisseau, C. Serres, C.-C. Chen, D. Kim, S. Stamboulian, H.-S. Shin, K. P. Campbell, M. De Waard, and C. Arnoult, "Expression, localization and functions in acrosome reaction and sperm motility of cav3. 1 and cav3. 2 channels in sperm cells: an evaluation from cav3. 1 and cav3. 2 deficient mice," *Journal of cellular physiology*, vol. 212, no. 3, pp. 753–763, 2007.

[139] J. Demongeot, H. Pempelfort, J. Martinez, R. Vallejos, M. Barría, and C. Taramasco, "Information design of biological networks: application to genetic, immunologic, metabolic and social networks," in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, pp. 1533–1540, IEEE, 2013.

[140] N. Huang, J. Lin, J. Ruan, N. Su, R. Qing, F. Liu, B. He, C. Lv, D. Zheng, and R. Luo, "Mir-219-5p inhibits hepatocellular carcinoma cell proliferation by targeting glypican-3," *FEBS letters*, vol. 586, no. 6, pp. 884–891, 2012.

[141] L. Boominathan, "The guardians of the genome (p53, ta-p73, and ta-p63) are regulators of tumor suppressor mirnas network," *Cancer and Metastasis Reviews*, vol. 29, no. 4, pp. 613–639, 2010.

[142] E. Hernando, M. Segura, H. Greenwald, D. Hanniford, and I. Osman, "Microrna & cutaneous melanoma: from discovery to prognosis and therapy," *Carcinogenesis*, p. bgs205, 2012.

[143] C. Georgescu, W. J. Longabaugh, D. D. Scripture-Adams, E.-S. David-Fung, M. A. Yui, M. A. Zarnegar, H. Bolouri, and E. V. Rothenberg, "A gene regulatory network armature for t lymphocyte specification," *Proceedings of the National Academy of Sciences*, vol. 105, no. 51, pp. 20100–20105, 2008.

[144] L. Jiang, X. Liu, Z. Chen, Y. Jin, C. Heidbreder, A. Kolokythas, A. Wang, Y. Dai, and X. Zhou, "Microrna-7 targets igf1r (insulin-like growth factor 1 receptor) in tongue squamous cell carcinoma cells," *Biochem. J*, vol. 432, pp. 199–205, 2010.

[145] L. Abbas, J. Demongeot, and N. Glade, "Synchrony in reaction–diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1908, pp. 4829–4862, 2009.

[146] S. Allesina and S. Tang, "Stability criteria for complex ecosystems," *Nature*, vol. 483, no. 7388, pp. 205–208, 2012.

[147] G. Almanza, A. Fernandez, S. Volinia, X. Cortez-Gonzalez, C. M. Croce, and M. Zanetti, "Selected microrna define cell fate determination of murine central memory cd8 t cells," *PLoS One*, vol. 5, no. 6, p. e11243, 2010.

[148] J. Aracena, J. Demongeot, E. Fanchon, and M. Montalva, "On the number of different dynamics in boolean networks with deterministic update schedules," *Mathematical biosciences*, vol. 242, no. 2, pp. 188–194, 2013.

[149] P. Baconnier, P. Pachot, and J. Demongeot, "An attempt to generalize the control coefficient concept," *Journal of Biological Systems*, vol. 1, no. 03, pp. 335–347, 1993.

[150] J. Demongeot, "Imgt/geneinfo: enhancing v (d) j recombination database accessibility," *Nucleic Acids Res*, vol. 32, p. 5154Beaudoin, 2004.

[151] H. B. Amor, J. Demongeot, and S. Sené, "Structural sensitivity of neural and genetic networks," in *MICAI 2008: Advances in Artificial Intelligence*, pp. 973–986, Springer, 2008.

[152] R. Bowen, "On axiom a diffeomorphisms:[expository lectures from the cbms regional conference held at north dakota state university, june 20-24, 1977]," American Math. Soc., 1978.

[153] J. A. Broderick and P. D. Zamore, "Competitive endogenous rnas cannot alter microrna function in vivo," *Molecular cell*, vol. 54, no. 5, pp. 711–713, 2014.

[154] O. Cinquin and J. Demongeot, "Positive and negative feedback: striking a balance between necessary antagonists," *Journal of Theoretical Biology*, vol. 216, no. 2, pp. 229–241, 2002.

[155] J. E. Cohen and C. M. Newman, "When will a large complex system be stable?," *Journal of theoretical Biology*, vol. 113, no. 1, pp. 153–156, 1985.

[156] M. Cosnard and J. Demongeot, "On the definitions of attractors," in *Iteration theory and its functional equations*, pp. 23–31, Springer, 1985.

[157] M. Cosnard and E. Goles, "Discrete state neural networks and energies," *Neural Networks*, vol. 10, no. 2, pp. 327–334, 1997.

[158] M. Cosnard, E. G. Chacc, and D. Moumida, "Bifurcation structure of a discrete neuronal equation," *Discrete applied mathematics*, vol. 21, no. 1, pp. 21–34, 1988.

[159] L. Demetrius, "Adaptive value, entropy and survivorship curves," 1978.

[160] L. Demetrius and M. Ziehe, "Darwinian fitness," *Theoretical population biology*, vol. 72, no. 3, pp. 323–345, 2007.

[161] E. Goles and J. Olivos, "Comportement périodique des fonctions à seuil binaires et applications," *Discrete Applied Mathematics*, vol. 3, no. 2, pp. 93–105, 1981.

[162] E. G. Ch *et al.*, "The convergence of symmetric threshold automata," *Information and Control*, vol. 51, no. 2, pp. 98–104, 1981.

[163] M. Cosnard and J. Demongeot, "Attracteurs: une approche déterministe," *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, vol. 300, no. 15, pp. 551–556, 1985.

[164] M. Noual, *Updating automata networks*. PhD thesis, Ecole normale supérieure de lyon-ENS LYON, 2012.