



HAL
open science

Méthodologie semi-formelle pour l'étude de systèmes biologiques : application à l'homéostasie du fer

Nicolas Mabilia

► **To cite this version:**

Nicolas Mabilia. Méthodologie semi-formelle pour l'étude de systèmes biologiques : application à l'homéostasie du fer. Hématologie. Université Grenoble Alpes, 2015. Français. <NNT : 2015GREAS042>. <tel-01561060>

HAL Id: tel-01561060

<https://theses.hal.science/tel-01561060v1>

Submitted on 12 Jul 2017

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

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



HAL Authorization

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Modèles, Méthodes et Algorithmes en Biologie**

Arrêté ministériel : 7 août 2006

Présentée par

Nicolas Mobilia

Thèse dirigée par **Jacques Demongeot**
et coencadrée par **Éric Fanchon**

préparée au sein de

Univ. Grenoble Alpes, TIMC-IMAG, F-38000 Grenoble, France

CNRS, TIMC-IMAG, F-38000 Grenoble, France

CHU de Grenoble, 38700 La Tronche

et de l'**EDISCE**

Méthodologie semi-formelle pour l'étude de systèmes biologiques

Application à l'homéostasie du fer

Thèse soutenue publiquement le **29/09/2015**,
devant le jury composé de :

Mme Anne Siegel

DR, IRISA - Rennes, Présidente

M. Sylvain Sené

Prof., LIF - Marseille, Rapporteur

M. Vitaly Volpert

DR, Institut Camille Jordan - Lyon, Rapporteur

M. Jean-Marc Moulis

DR, LBFA/CEA - Grenoble, Examineur

M. Jacques Demongeot

PU-PH, AGIM - Grenoble, Directeur de thèse

M. Éric Fanchon

CR, TIMC-IMAG - Grenoble, Co-Encadrant de thèse

M. Oded Maler

DR, Verimag - Grenoble, Invité



Remerciements

En premier lieu, je tiens à remercier mes chefs Éric et Jacques, ainsi que toutes les personnes avec qui j'ai eu la chance de collaborer, à savoir Laurent Trilling, Alexandre Donzé, Fabien Corblin, Alexandre Rocca. Je tiens également à remercier nos collaborateurs biologistes : Jean-Marc Moulis, Fiona Louis et Emmanuel Pourcelot qui ont survécu à nos salves d'équations différentielles[1]. Je remercie également Youssef Hamadi et Microsoft Research pour leur financement.

Je tiens aussi à remercier le service informatique et le secrétariat du laboratoire pour leur présence et leur réactivité.

J'ai une pensée pour ma mère et mon frère, ainsi que pour mes amis (Olivier, Marina, Vincent, Mélanie, Adrien, Aurélia, Steven, Anne-Laure, Romain, Adrien et tous les autres) qui m'ont soutenu tout au long de ce travail.

En dernier lieu, je voudrais remercier Planti, ma plante verte, qui n'a pas dit un mot malgré toutes les fois où j'ai oublié de l'arroser.

Mémoire

Cette page est destinée à être découpée pour accompagner le lecteur dans sa compréhension des acronymes tout au long de sa lecture.

- ACO** : *Aconitase* : enzyme facilitant la transformation du citrate en iso-citrate.
- ALAS** : enzyme nécessaire pour la synthèse de PPIX.
- ARNm** : *ARN messenger* : intermédiaire dans la synthèse de protéine à partir d'ADN.
- FBXL5** : protéine provoquant la dégradation de IRP2.
- FPN** : *Ferroportine* : protéine permettant de faire sortir du fer des cellules.
- Ft** : *Ferritine* : protéine stockant du fer au sein des cellules.
- IRE** : *Iron Responsive Element* : structure de l'ARN messenger sur laquelle peut venir se fixer un IRP.
- IRP** : *Iron Regulatory Protein* : protéine régulant la traduction ou la dégradation de certains ARNm. C'est un acteur principal dans la régulation de l'homéostasie du fer.
- mRNA** : *messenger RNA* : acronyme anglais de l'ARNm.
- PPIX** : *Protoporphyrine IX* : molécule servant d'intermédiaire dans la synthèse de l'hème.
- Tf** : *Transferrine* : molécule permettant le transport du fer jusqu'aux cellules.
- TfR** : *Transferrin Receptor* : récepteur à la surface des cellules permettant l'entrée de fer au sein des cellules.
- UTR** : *UnTranslated Region* : deux régions de l'ARNm (nommées 5'-UTR (au début de l'ARNm) et 3'-UTR (à la fin de l'ARNm)) qui ne sont pas traduites en protéine. Les IRE sont situés dans ces parties.
- ASP** : *Answer Set Programming* : langage de programmation déclaratif non monotone, ici utilisé pour implémenter les réseaux de Thomas.
- EDO** : *Équation Différentielle Ordinaire* : relation mathématique entre des variables et leurs dérivées. Les EDO sont utilisées ici pour définir la dynamique de systèmes biologiques.
- ODE** : *Ordinary Differential Equation* : acronyme anglais de EDO.
- STL** : *Signal Temporal Logic* : logique temporelle utilisée ici pour traduire les données comportementales du système biologique.

Résumé

Les travaux de cette thèse portent principalement sur le développement d'une méthodologie pour la modélisation de réseaux biologiques. Cette méthodologie, basée sur une modélisation en équations différentielles, intègre aussi bien des méthodes formelles (solveur sur intervalles, solveur de formules STL) qu'analytiques (calcul de stabilité d'état stationnaire) ou numériques (algorithmes d'optimisation, analyses statistiques). Elle fournit un cadre cohérent permettant la prise en compte de différents types de données, telles la réponse comportementale à une perturbation ou des données quantitatives (demi-vie, concentrations) et permet d'obtenir des déductions qualitatives comme quantitatives sur le système étudié. En collaboration avec une équipe de biologistes, cette méthodologie est appliquée, avec succès, au système de l'homéostasie du fer : nous étudions la réponse intracellulaire du système, via des protéines régulatrices spécifiques (protéines IRP), face à une situation de carence en fer. Un résultat majeur de cette étude concerne l'affinement des connaissances sur la concentration de fer intracellulaire nécessaire à la prolifération des cellules : l'ordre de grandeur étonnamment faible de cette concentration a d'abord été souligné par l'étude du modèle, puis a été étudié et validé expérimentalement par nos collaborateurs.

Mots-clés: Bioinformatique, Modélisation, Homéostasie du fer

Abstract

The major part of this PhD consists in the creation of a methodology to model biological networks. This methodology considers models based on differential equations, and uses formal methods (interval solver, verification of STL formula), analytical methods (study of stability) and numerical methods (optimization algorithm, statistical analysis). It provides a consistent framework where many kinds of data, like behavioral response to perturbation, or quantitative data (metabolite half-life and concentration) can be integrated. Qualitative and quantitative deductions on the system can be found. In collaboration with a biologist team, this methodology is successfully applied to the iron homeostasis network: we study the response of the system to an iron depletion, at the intracellular level, based on specific regulatory proteins (IRP proteins). A major output of this study is an insight into the level of iron required for cells proliferation: the unexpectedly low order of magnitude of this concentration is pointed out by the study of the model, and has been experimentally studied and validated by our collaborators.

Keywords: Bio-informatics, Modeling, Iron homeostasis

Table des matières

1	Introduction	7
1.1	Cadre	7
1.1.1	Cadre méthodologique	7
1.1.2	Cadre biologique	8
1.2	Sujet	8
1.3	État de l'art : approches de modélisation	9
1.3.1	Approches discrètes	10
1.3.2	Approches continues	13
1.3.3	Choix du formalisme du modèle	15
1.4	État de l'art : formalisation des données	15
1.4.1	Relations et intervalles de paramètres	15
1.4.2	Logiques temporelles	16
1.4.3	Au sein d'outils d'analyse	17
1.4.4	Choix du formalisme des données	19
1.5	Plan	19
2	Description biologique	21
2.1	Introduction	21
2.2	Fonctionnement du mécanisme IRP-IRE	22
2.3	Entrée et sortie du fer au sein de la cellule	22
2.4	Stockage du fer	24
2.5	Synthèse de l'hème	25
2.6	Autres acteurs	25
2.7	Contexte d'étude et considérations biologiques	25
2.7.1	Contexte d'étude	25
2.7.2	Considérations biologiques	26
3	Modèle préliminaire	27
3.1	Obtention d'un modèle mathématique	27

3.1.1	Hypothèses de modélisation	27
3.1.2	Conventions de notation	28
3.1.3	Régulation par les IRP	28
3.1.4	Régulation des IRP	30
3.1.5	Production et consommation des agrégats fer-soufre	30
3.1.6	Synthèse de l'hème	31
3.1.7	Entrée, utilisation et stockage du fer	32
3.1.8	Entrées du modèle	32
3.2	Pistes d'étude du modèle	32
3.2.1	Modélisation du scénario d'étude	33
3.2.2	Adimensionnement	34
3.2.3	Système lent-rapide	35
3.2.4	Simplification	35
4	Modèle spécifique de la régulation par les IRP	39
4.1	Introduction	39
4.1.1	Définition du modèle	39
4.1.2	Méthodologie d'étude	40
4.1.3	Implémentation	41
4.2	Présentation des travaux	41
4.3	Détails de la méthodologie	79
4.3.1	À l'état stationnaire	79
4.3.2	Étude de la dynamique	81
4.3.3	Impact de l'étude du modèle	84
4.4	Comparaison de modèles	84
4.4.1	Comparaison avec le modèle d'Omholt <i>et al.</i>	85
4.4.2	Comparaison avec le modèle de Mitchell <i>et al.</i>	86
4.4.3	Comparaison avec le modèle de Chifman <i>et al.</i>	86
5	Modèle intégrant les ARNm	91
5.1	Introduction	91
5.2	Présentation du modèle	92
5.2.1	Hypothèses de modélisation	92
5.2.2	Conventions de notation	93
5.2.3	Système d'équations différentielles	94
5.3	Modélisation des données	98
5.3.1	Intervalles	99

	5
5.3.2	Contraintes sur la dynamique 99
5.3.3	Réécriture 100
5.4	Étude de l'état stationnaire 101
5.4.1	Réduction de modèle 101
5.4.2	Calcul analytique des concentrations 101
5.4.3	Stabilité de l'état stationnaire 103
5.4.4	Réduction des intervalles 104
5.5	Recherche d'une région valide 104
5.5.1	Recherche de jeux de paramètres valides à l'état stationnaire 105
5.5.2	Recherche d'un jeu de paramètres valide 108
5.6	Résultats 108
6	Conclusion 113
Annexes	115
A	Données et intervalles pour le modèle intégrant les ARNm 117
A.1	Paramètres 117
A.2	Concentrations à l'état stationnaire 121
A.3	Contraintes à l'état stationnaire 122
A.4	Ré-écriture de contraintes 124
B	Modélisation en ASP des réseaux de Thomas 127
B.1	Introduction 127
B.2	Présentation des travaux 128
Bibliographie	157

Chapitre 1

Introduction

Ce chapitre commence par présenter le cadre général, tant méthodologique que biologique, dans lequel s'inscrit ce doctorat (section 1.1). Ensuite, la section 1.2 décrit le sujet lui-même en présentant les différents objectifs de ces travaux. Puis, nous présentons un état de l'art des techniques de modélisation : la section 1.3 présente les différents formalismes utilisés pour la modélisation de systèmes biologiques, et la section 1.4, présente les différentes manières de représenter des données biologiques et de prendre en compte les grandes incertitudes de ces données. Nous indiquons, au cours de ces sections, le choix des formalismes utilisés pour modéliser le système biologique et les données disponibles (sous-sections 1.3.3 et 1.4.4). Enfin, l'organisation générale du document est décrite dans la dernière section de ce chapitre.

1.1 Cadre

Ces travaux s'inscrivent dans le champ de la bio-informatique, discipline à l'intersection, entre autres, de l'informatique, des mathématiques et de la biologie. Nous présentons ici le contexte méthodologique, ainsi que le contexte biologique, dans lesquels s'inscrivent ces travaux.

1.1.1 Cadre méthodologique

Notre connaissance des réseaux métaboliques et génétiques nécessaires au fonctionnement du corps humain ne cesse de croître. En effet, l'évolution technologique et le perfectionnement des méthodologies expérimentales permettent la réalisation de nouvelles expériences, et l'amélioration de la précision des résultats obtenus, générant ainsi continuellement de nouvelles données et approfondissant nos connaissances sur la complexité de ces réseaux.

La biologie systémique, qui étudie les processus biologiques en les considérant comme des systèmes complexes dynamiques, tente de répondre à cette problématique. Elle consiste à intégrer, au sein d'un modèle, les effets combinés d'un ensemble d'entités moléculaires et non plus à les explorer entité par entité comme usuellement en biologie moléculaire [2]. Un des objectifs majeurs de ce domaine d'étude est donc de fournir des modèles permettant une meilleure compréhension des phénomènes cellulaires à un niveau moléculaire.

Différents formalismes mathématiques ou informatiques existent pour la modélisation de réseaux biologiques. Le choix du formalisme dépend du système biologique modélisé, du niveau de description choisi, du type et de la quantité des données disponibles, ainsi que des objectifs de l'étude. En effet, l'objectif peut aller de la visualisation de la structure du réseau et de l'analyse de sa topologie, jusqu'à l'analyse détaillée de sa dynamique dans différentes conditions,

la prédiction de son comportement dans des conditions nouvelles, ou même le contrôle de sa dynamique pour réaliser une fonction spécifique.

1.1.2 Cadre biologique

Le fer est un élément essentiel pour la plupart des organismes vivants. Chez l'homme, ce métal intervient dans la composition de nombreuses protéines, principalement sous la forme d'agrégat fer-soufre ou d'hème. La dérégulation de ce métal au sein de l'organisme est impliquée dans plusieurs pathologies telles l'ataxie de Friedreich ou les anémies [3]. L'organisme humain tente alors de maintenir la concentration en fer dans un intervalle viable. Ce processus est nommé l'homéostasie du fer.

La régulation du fer au niveau de l'organisme implique plusieurs organes, dont principalement les intestins (assimilant le fer de l'alimentation), la moelle osseuse (produisant des globules rouges), la rate (permettant le recyclage des globules rouges sénescents) et le foie (stockant le fer). Nous nous concentrons, dans ce travail, sur les cellules érythroïdes car ce sont les plus grosses consommatrices de fer [4]. Ces cellules, présentes dans la moelle osseuse, sont les précurseurs des globules rouges et requièrent une importante quantité de fer afin de synthétiser de l'hémoglobine, principal composant des futurs globules rouges. Afin d'éviter l'apoptose¹ en situation de carence en fer, ces cellules possèdent un mécanisme de régulation, basé sur les protéines IRP (*Iron Regulatory Protein*).

Les données disponibles sur ce mécanisme de régulation sont partielles et hétérogènes, c'est-à-dire obtenues dans des conditions et des types cellulaires divers. Si l'on peut dire que le comportement qualitatif du réseau en réaction à une situation de carence en fer est connu dans les grandes lignes, le fonctionnement réel de ce système reste à découvrir. Pour cela, différentes constantes et concentrations, dans différentes conditions (présence ou carence en fer), ont été mesurées par nos collaborateurs, fournissant des données quantitatives correspondant directement au phénomène que nous voulons modéliser. Il est cependant encore nécessaire d'utiliser des données extraites de la littérature. Ces caractéristiques (connaissances partielles, grande incertitude sur les paramètres, hétérogénéité des données utilisées) sont typiques des problèmes de modélisation en biologie systémique et créent un besoin d'outils spécifiques.

1.2 Sujet

L'objectif de nos travaux est double : il s'agit en premier lieu de développer une méthodologie permettant la modélisation de systèmes biologiques. En second lieu, nous souhaitons améliorer la compréhension du système de l'homéostasie du fer chez les précurseurs des globules rouges. Nous considérons ce problème de modélisation comme typique et représentatif de toute une classe de problèmes en biologie systémique. La méthodologie que nous développons doit remplir plusieurs critères, listés ci-dessous.

Intégration de toutes les données (y compris très incertaines) Les données disponibles dans la bibliographie se regroupent en deux catégories principales. D'un côté, nous disposons de données quantitatives, telles la concentration d'une espèce lorsque le fer est en abondance dans la cellule ou encore la demi-vie d'une espèce. Ces données, bien que quantitatives, sont souvent entachées d'une grande imprécision, due, par exemple, à la variabilité des échantillons biologiques. De l'autre côté, nous disposons de données qualitatives relatives au comportement du système biologique, typiquement en réponse à une mise en

1. L'apoptose est un processus où la cellule provoque sa propre mort.

carence en fer. Les données disponibles sur ce système biologique étant peu nombreuses par rapport à la complexité du système, il est nécessaire d'exploiter toutes les données disponibles pour obtenir une famille de modèles la mieux définie possible.

Obtention de déductions Ce travail étant motivé par la collaboration entre modélisateurs et biologistes, il est également nécessaire que cette méthodologie permette de faire des déductions sur le système étudié, à partir des données fournies, permettant de guider le biologiste dans le processus de décision des expérimentations à réaliser. Ceci permet ainsi de montrer l'intérêt d'outils et méthodes appropriés pour l'approfondissement des connaissances sur les réseaux biologiques.

Considération de l'ensemble des modèles L'approche traditionnelle de modélisation consiste à instancier complètement un modèle et à confronter les comportements de ce modèle avec les résultats expérimentaux. En cas d'incompatibilité, une autre instantiation est essayée. Ce processus est répété jusqu'à obtenir un modèle \mathcal{M}_V compatible avec les données expérimentales [5]. Cette approche, bien que fournissant un modèle valide, ne répond pas au problème posé. En effet, d'autres modèles compatibles avec les données expérimentales peuvent exister et les déductions faites à partir du modèle \mathcal{M}_V peuvent ne pas être vraies pour un autre modèle valide. Pour éviter ce problème, et ainsi ne pas guider le biologiste vers des expérimentations inutiles, la méthodologie développée doit permettre de générer *l'ensemble* des modèles valides. Dans ce contexte, nous définissons une prédiction comme étant une propriété qui est vraie pour l'ensemble des modèles valides. Il est clair que plus la quantité de données que l'on peut intégrer sera grande, plus l'ensemble de modèles valides sera réduit et plus les prédictions seront précises.

Automatisation de la méthodologie Avec le temps, de nouvelles données quantitatives et d'interaction seront disponibles et les modèles seront alors amenés à évoluer. Pour faciliter l'utilisation de cette méthodologie pour l'étude de ces futurs modèles, il est nécessaire qu'elle soit générique et simple d'utilisation. Nous désirons que la méthodologie créée soit automatisable et automatisée.

Il va donc falloir, dans un premier temps, choisir le formalisme le plus adapté à la modélisation du système et à la représentation des données. Ensuite, il faudra utiliser ou créer des outils permettant l'étude du modèle ainsi créé, cette étude devant permettre d'extraire de nouvelles connaissances sur le système.

1.3 État de l'art : approches de modélisation

La compréhension des mécanismes moléculaires à la base du comportement des cellules nécessite la construction de modèles mathématiques. Parmi l'ensemble des formalismes de modélisation, les modèles cinétiques, contenant de nombreux paramètres, permettent de calculer, par simulation, le comportement du modèle [6], comme par exemple la réponse à une perturbation. Les paramètres de ces modèles étant en général mal connus, voire complètement inconnus pour certains, des formalismes et des méthodes prenant en compte ce manque de connaissance doivent être utilisés (ou développés).

Une première possibilité, largement utilisée dans le domaine, est constituée par les réseaux dits *de Thomas*. Il s'agit d'un formalisme discret décrivant la dynamique qualitative d'un système d'acteurs en interaction [7, 8, 9]. Il a été appliqué dans de multiples contextes dont la croissance des plantes [10], la réponse immunologique [11] et les réseaux de gènes [12, 13], ces derniers représentant la majorité des applications de ce formalisme. Les réseaux *booléens*, dans

lesquels chaque acteur peut être dans deux états (par exemple actif ou inactif) constituent une grande famille de formalismes utilisés en biologie systémique. Un formalisme de réseau booléen probabiliste a été utilisé pour modéliser l'homéostasie du fer chez la levure [14]. Par ailleurs, les réseaux de Petri, initialement introduits en 1962 pour l'analyse de systèmes concurrents, constituent un autre formalisme discret permettant la modélisation de systèmes biologiques. Ils ont été, entre autres, utilisés pour modéliser l'homéostasie du fer [15]

Les équations différentielles ordinaires (EDO) restent néanmoins le formalisme le plus utilisé. La simulation par intégration numérique d'un système différentiel nécessite de connaître la valeur numérique de tous les paramètres du modèle. Cela est un gros inconvénient par rapport aux réseaux de Thomas qui gèrent des niveaux qualitatifs et non pas des grandeurs quantitatives.

Nous présentons, dans la suite de cette section, les avantages et les inconvénients de ces différents formalismes, puis nous indiquons, dans la section 1.3.3, lequel est le plus adapté à notre problématique.

1.3.1 Approches discrètes

Ce type d'approche propose une représentation qualitative du système biologique, bien adaptée lorsque peu de données quantitatives sont disponibles. Ces formalismes, d'une nature plus abstraite que les formalismes continus, ne nécessitent pas une description détaillée de la dynamique des interactions du système biologique.

Réseaux de Thomas

Nous décrivons brièvement les réseaux de Thomas en utilisant la terminologie des réseaux de gènes. Chaque gène i ($i = 1, \dots, n$) produit une protéine donnée, et une protéine peut activer ou réprimer l'expression d'un ou plusieurs gènes (le gène source peut lui-même être une cible, on parle alors d'auto-régulation). L'énoncé « i active j » (resp. inhibe) signifie que lorsque la concentration de la protéine I (produite par le gène i) est au-dessus d'un certain seuil, alors la protéine I se lie au promoteur du gène j et active (resp. réprime) la transcription de ce dernier. Les gènes peuvent exister dans de multiples états d'expression auxquels correspondent des taux d'expression distincts. Dans le formalisme des réseaux de Thomas, les concentrations des protéines sont modélisées par des variables discrètes x_i . Cette discrétisation se fait selon un ou plusieurs seuils : si x_i est la concentration discrétisée d'une protéine, $x_i = 0$ signifie que la concentration est en dessous du premier seuil, $x_i = 1$ signifie que la concentration est entre le premier et le deuxième seuil, etc. En d'autres termes, chaque valeur discrète représente un intervalle entre deux seuils. Les concentrations (grandeurs réelles) sont modélisées par des niveaux discrets en nombre fini. Un gène qui n'a qu'une seule cible est donc associé à un seul seuil, et sa variable discrète associée est booléenne. L'état discret du modèle est représenté par une liste d'entiers $S = (x_1, \dots, x_n)$ et correspond à un schéma d'expression du modèle.

À chaque état discret est associé un *point focal*. Il s'agit de la liste des taux d'expression des gènes, qui sont, eux aussi, des grandeurs discrètes. Ceci traduit le fait que le taux d'expression de chaque gène dépend d'une ou plusieurs protéines (protéines régulatrices). Ces relations entre niveaux de protéines régulatrices et taux d'expression sont appelées équations discrètes. À partir de ces équations, des *règles logiques* permettent de calculer l'évolution du modèle, c'est-à-dire les successeurs possibles d'un état discret donné. Par définition, il existe une transition $S_1 \rightarrow S_2$ entre deux états S_1 et S_2 si S_2 est un successeur possible de S_1 . Le graphe de transition est un graphe dirigé dont les nœuds sont les états et les arcs sont les transitions ainsi définies. Ce graphe représente, en intension, l'ensemble de tous les comportements qualitatifs possibles du

modèle d'un réseau génétique donné.

Thomas et Snoussi [16, 9] ont montré que ce formalisme peut être vu comme une abstraction de systèmes différentiels linéaires par morceaux. L'espace des concentrations est découpé en domaines rectangulaires par des seuils. Chacun de ces domaines (produit cartésien d'intervalles) correspond à un état discret. Une transition $S_1 \rightarrow S_2$ existe entre deux états discrets si les domaines correspondants D_1 et D_2 sont adjacents (i.e. ayant une face en commun) et s'il existe une trajectoire continue allant de D_1 à D_2 . En fait, un état discret représente un domaine, donc une infinité d'états réels, et une transition représente une infinité de trajectoires entre deux domaines adjacents. Le non-déterminisme des réseaux de Thomas s'interprète donc naturellement : à partir d'un domaine donné, il peut arriver que plusieurs domaines soient atteignables, selon la condition initiale précise. Le graphe de transition représente toutes les possibilités, mais un système donné dans des conditions fixées – ces conditions n'étant pas forcément connues de l'expérimentateur – suit un seul chemin du graphe de transition. Le non-déterminisme reflète simplement la perte de connaissance due à l'abstraction.

L'interprétation en tant qu'abstraction de systèmes différentiels linéaires par morceaux (SDLM) permet en outre de bien comprendre le domaine de validité de ce formalisme. Les fonctions qui interviennent dans la définition d'un SDLM sont des fonctions (discontinues) de type *marche d'escalier* (croissante en fonction de la concentration de la protéine régulatrice pour une activation, et décroissante pour une inhibition). Ces marches d'escalier proviennent de fonctions de Hill, dont le graphe est sigmoïdal. Une marche d'escalier s'obtient par passage à la limite d'une sigmoïde, lorsqu'on fait tendre la pente de cette dernière au voisinage du seuil vers l'infini. Ceci montre que le formalisme de Thomas s'applique bien à des systèmes formés d'éléments dont la cinétique est de type Hill, avec une sigmoïde assez raide. On ne doit, en principe, pas l'appliquer si certaines cinétiques sont, par exemple, de type loi d'action de masse, c'est-à-dire polynomiales ; par contre, les cinétiques de type Michaelis-Menten (Hill d'ordre 1) ou allostériques, étant des fractions rationnelles constantes à l'infini des concentrations, peuvent relever du formalisme de Thomas, avec éventuellement plusieurs seuils, si les pentes à l'origine ou aux points d'inflexion sont très grandes. Il faut ajouter que le formalisme n'est bien fondé que si l'on prend en compte les modes glissants, c'est-à-dire les trajectoires qui « glissent » le long d'un seuil. Le formalisme étendu proposé par de Jong *et al.* [17, 18, 19] est bien fondé, dans le sens où toute trajectoire continue du système concret est associée à un chemin dans le graphe de transitions (l'inverse n'étant pas vrai).

Par ailleurs, la notion de durée (temps métrique) est éliminée dans le processus d'abstraction. Ceci vient du fait que les tailles de domaines, les valeurs numériques des taux d'expression, et les valeurs numériques des conditions initiales ne sont pas connues. Il est donc impossible de déterminer le temps mis, par exemple, pour traverser tel domaine. Lorsque des informations sur la durée des processus sont disponibles (c'est souvent le cas), on ne peut donc pas les exploiter avec ce formalisme. Différents auteurs (dont R. Thomas lui-même) ont proposé des extensions pour cela [20, 21, 22].

Nous avons à l'origine l'intention d'utiliser ce formalisme pour analyser qualitativement l'homéostasie du fer, car il permet d'utiliser des techniques de résolution par contraintes, ce qui donne accès à des fonctionnalités assez sophistiquées : inférence de paramètres, révision automatique de modèles, inférences de propriétés. Nous y avons renoncé car, comme nous le verrons, d'une part, la cinétique de certains processus n'est pas de type Hill, et, d'autre part, nous disposons d'informations de durée.

Réseaux booléens

Les réseaux booléens sont un autre formalisme permettant une modélisation discrète des systèmes biologiques [23, 24]. Dans ces réseaux, chaque acteur ne peut prendre que deux valeurs : 0 ou 1. Une fonction booléenne est associée à chaque nœud du réseau. Cette fonction donne la valeur du nœud au pas suivant, si ce nœud est mis à jour. En plus de ces fonctions booléennes, il faut spécifier des règles pour choisir les nœuds qui doivent être mis à jour lors d'un pas de temps donné. La dynamique, discrète, de ces modèles se fait par une succession de transitions entre états du modèle. Il existe trois types de dynamiques. D'un côté, nous trouvons les transitions dites synchrones où, à chaque pas de temps, *tous* les acteurs du réseau sont mis à jour selon les fonctions booléennes. Ce type de dynamique est simple et implique que le modèle est déterministe, un état ne pouvant avoir qu'un seul successeur. De l'autre côté, nous trouvons les transitions dites asynchrones où, à chaque pas de temps, un seul acteur est mis à jour selon les règles logiques. Des règles de mise à jour spécifient l'acteur qui doit être mis à jour à un pas de temps donné. Un ordre de mise à jour peut simplement être spécifié, ou bien le choix de l'acteur peut être aléatoire. Le troisième type de dynamique, à mi-chemin entre les deux précédentes, est constituée des transitions bloc-séquentielles où, à chaque pas de temps, un sous-ensemble des acteurs est mis à jour. Le choix du type de règles de mise à jour impacte la dynamique du modèle [25] : par exemple, les transitions synchrones génèrent des cycles artéfactuels.

Les réseaux booléens probabilistes [14] sont une variante des réseaux booléens avec transitions asynchrones. Cette variante consiste à assigner une probabilité à chaque transition possible et à générer un ensemble de simulations se conformant aux probabilités choisies. Elle permet l'expression de différences qualitatives des probabilités d'occurrence entre différents événements. L'extension probabiliste la plus naturelle des réseaux booléens à seuil est constituée par les réseaux de Hopfield, très utilisés en physique statistique et en modélisation neuronale [26, 27, 28, 29, 30]. Les réseaux booléens à mise à jour séquentielle probabiliste et également à transition probabiliste constituent la classe la plus générale des réseaux booléens doublement stochastiques [31].

L'avantage majeur du formalisme booléen est sa grande simplicité, ce qui permet son utilisation pour l'étude de grands réseaux. Toutefois, comme les réseaux de Thomas, il n'est pas adapté lorsque des données quantitatives sur le système biologique sont disponibles puisqu'il ne permet pas de les prendre en compte. De plus, il ne permet pas une représentation fidèle de l'ensemble des cinétiques existantes.

Réseaux de Petri

Les réseaux de Petri sont représentés par des graphes bipartis orientés contenant deux ensembles disjoints de nœuds, qui sont les *places* et les *transitions*. Dans les réseaux de Petri modélisant un système biologique, les places, généralement représentées par des cercles, modélisent les acteurs du réseau et les transitions, représentées par des rectangles, modélisent les réactions. De plus, chaque place contient un nombre (positif ou nul) de jetons qui représente, de manière quantitative ou qualitative, la quantité de chaque acteur. Une transition peut être déclenchée si les places en entrée de celle-ci contiennent suffisamment de jetons. Si elle est déclenchée, elle peut déplacer, créer et supprimer des jetons au sein des places qui lui sont reliées.

Les réseaux de Petri classiques ne permettent pas de prendre en compte les durées des réactions modélisées par les transitions. Néanmoins, selon le type de réaction (ou méta-réaction, comme par exemple, la phagocytose d'une cellule), le temps nécessaire peut varier de plusieurs ordres de grandeur. Pour pallier cela, les réseaux de Petri temporisés [32] ont été introduits. Dans ces réseaux, un intervalle de durée $[\tau_i, \tau'_i]$ est associé à chaque place i , indiquant la durée minimale

(τ_i) et maximale (τ_i') durant laquelle un jeton peut rester dans cette place. Pour la modélisation de réseaux biologiques, la représentation du temps nécessaire (noté d_i) pour que les conditions permettant le démarrage des réactions biochimiques soient réunies, se fait en considérant $\tau_i = d_i$ et $\tau_i' = \infty$. Dans ce formalisme, une transition peut être déclenchée si les places en entrée de celle-ci contiennent suffisamment de jetons présents depuis une durée supérieure à τ_i .

Si une distribution initiale des jetons au sein des places a été déterminée, ces formalismes permettent la simulation de la dynamique du réseau. À partir d'une distribution des jetons, l'ensemble des transitions déclenchables est calculé. Puis, une règle choisit, parmi cet ensemble, la transition qui est déclenchée. Ce déclenchement provoque une modification de la distribution des jetons et éventuellement du nombre de jetons. L'itération de ce processus permet d'obtenir la dynamique du réseau [33].

Ce formalisme, bien que permettant de modéliser plus de cinétiques que les réseaux booléens ou les réseaux de Thomas, ne permet pas d'intégrer des données quantitatives. De plus, le calcul de la dynamique du réseau nécessite de définir une règle déterminant quelle transition est déclenchée si plusieurs sont déclenchables, et le choix de cette règle semble souvent arbitraire.

1.3.2 Approches continues

Les formalismes continus, et plus particulièrement les équations différentielles, sont les approches les plus utilisées pour modéliser les systèmes biologiques. Dans ces approches, chaque terme du modèle représente un processus spécifique, permettant une représentation quantitative des acteurs mis en jeu et une interprétation biologique aisée des paramètres de ces équations.

Équations différentielles ordinaires

La modélisation de systèmes biologiques à base d'équations différentielles ordinaires (EDO) est couramment utilisée en biologie systémique. Elle est naturelle car les lois cinétiques classiques de la chimie et de la biochimie s'expriment en ces termes. Usuellement, chaque variable du système d'EDO représente une entité chimique du modèle. L'utilisation de ce formalisme est valide si deux hypothèses sont vérifiées. La première est que la concentration des entités considérées soit assez grande pour pouvoir négliger les fluctuations et considérer les concentrations comme des quantités continues. La deuxième est que la concentration soit homogène dans le volume du système biologique modélisé. S'il existe une organisation spatiale (par exemple des gradients), il faut recourir aux équations aux dérivées partielles. En revanche, l'existence de compartiments, eux-mêmes homogènes, mais connectés par des flux, peut être décrite en termes d'EDO.

Différentes familles d'équations sont utilisées pour modéliser les systèmes biologiques : la première famille est celle des systèmes d'EDO non-linéaires. Ces systèmes sont les plus généraux car ils permettent de modéliser tous les types d'interactions. Classiquement, un système d'équations différentielles non-linéaires est étudié en faisant des simulations. Lorsque le modèle contient des paramètres, se pose la question de la valeur de ces paramètres. Nous avons déjà signalé qu'il s'agit là d'un grand problème en biologie systémique. Pour cette raison, certains auteurs considèrent les EDO comme inadaptées à la modélisation en biologie systémique et préfèrent des formalismes de type réseaux de Thomas. La deuxième famille est celle des systèmes d'EDO linéaires. L'étude de la dynamique de ces systèmes est aisée puisque les outils d'algèbre linéaire sont applicables. Toutefois, ces systèmes ne permettent pas forcément une modélisation fidèle des processus biologiques considérés. Une des utilisations de ces systèmes vient de la linéarisation de systèmes non-linéaires. Cette linéarisation permet, par exemple, de faciliter la recherche des états stationnaires du système [34]. Ils peuvent aussi être utilisés comme approximation locale

de systèmes non-linéaires pour le calcul de la stabilité d'état stationnaire ou encore le calcul d'atteignabilité.

Bien que le formalisme déterministe des systèmes d'EDO non-linéaires nécessite de déterminer les valeurs des paramètres du système, il est adapté pour modéliser des systèmes biologiques complexes et intégrer des données quantitatives relatives aux réactions physico-chimiques modélisées. Ces données sont, par exemple, des mesures de paramètres cinétiques ou des concentrations d'acteurs dans différentes situations. Lorsque les données sont partielles ou de nature qualitative, des méthodes spécifiques doivent être développées et utilisées, ce qui constitue un des objectifs principaux de cette thèse.

Systèmes hybrides

Les systèmes hybrides sont une combinaison de systèmes continus et de systèmes discrets. Un système est dit hybride s'il est caractérisé par un ensemble de variables continues (telles les variables des systèmes d'équations différentielles) auxquelles s'ajoutent une ou plusieurs variables discrètes caractérisant le *mode* dans lequel se trouve le système. La dynamique des variables continues est définie spécifiquement pour chacun des modes du système. Le passage d'un mode à un autre se fait par le franchissement de *gardes*, qui sont des inégalités sur des fonctions des variables du système et du temps. Une simulation d'un système hybride est alors composée de la trajectoire des variables continues, ainsi que de la séquence des modes.

Ce formalisme est adapté à la modélisation de systèmes constitués d'entités ponctuelles (par exemple, des cellules) et d'acteurs continus (par exemple, des concentrations au sein de cellules). Il permet également de modéliser simplement des changements de conditions expérimentales au cours du temps. Un système hybride peut aussi être obtenu par un processus appelé, en anglais, *hybridization* [35]. L'idée est de partitionner l'espace des phases et d'approximer la dynamique du système non-linéaire d'intérêt par un ensemble de dynamiques continues plus simples, et donc plus faciles à analyser. Un système continu non-linéaire peut par exemple être approximé par un système affine par morceaux, qui est un cas particulier de système hybride. La simulation de systèmes hybrides est toutefois un peu plus complexe que celle des systèmes d'EDO puisqu'il est nécessaire de gérer le passage de gardes.

Autres approches

D'autres approches permettent de s'affranchir du problème du choix de valeurs précises pour les paramètres du modèle ou encore permettent de s'adapter au manque de connaissances sur la structure du système biologique.

Par exemple, les travaux de Sass *et al.* [36] utilisent des équations différentielles pour modéliser qualitativement un système biologique. Pour cela, les auteurs n'utilisent pas de valeurs de paramètres issues de mesures directes, mais ne considèrent que des rapports de concentrations et en déduisent des valeurs de paramètres qui n'ont de sens que d'un point de vue relatif. Ces travaux permettent de s'affranchir du problème classique du choix des valeurs des paramètres, mais ne permettent pas de prendre en compte des données disponibles fournissant la valeur d'un paramètre cinétique ou la concentration d'un acteur dans une situation donnée.

Nous pouvons également citer la *Flux Balance Analysis* et ses extensions [37] qui consistent à modéliser le système biologique par un ensemble de flux entre les acteurs du modèle et à n'étudier que les états stationnaires de ce modèle. Le système étant à l'état stationnaire, il peut être écrit selon l'équation (1.1) où v est la matrice de taille $m \times 1$ représentant les m flux, $[A]$ la matrice des concentration des n acteurs du modèle et S est la matrice contenant les coefficients

stœchiométriques, de taille $n \times m$.

$$\frac{d[A]}{dt} = 0 = S \cdot v \quad (1.1)$$

L'objectif est de calculer la matrice des flux v . Néanmoins, si le nombre n de métabolites du modèle est plus faible que le nombre m de flux, l'équation (1.1) est sous-déterminée et plusieurs solutions sont possibles. D'autres contraintes sont alors introduites, telles des contraintes d'intervalle sur les flux. De plus, la définition d'une fonction objectif permet d'obtenir une unique solution. La matrice stœchiométrique fournit des informations sur la topologie du réseau, telles le degré moyen des nœuds, ou la détection des nœuds les plus connectés. La *Flux Balance Analysis* ne concerne que les flux à l'état stationnaire et de ce fait n'a pas besoin des lois cinétiques et des valeurs des constantes de vitesse. Les modèles ainsi automatiquement construits ne permettent pas de faire de simulations puisque les lois cinétiques ne sont pas décrites.

1.3.3 Choix du formalisme du modèle

Dans ce travail sur l'homéostasie du fer, nous choisissons de modéliser notre système biologique avec des équations différentielles non-linéaires. Ce formalisme présente plusieurs avantages. D'une part, il laisse toute liberté pour représenter les lois cinétiques des réactions (ou méta-réactions) modélisées. D'autre part, ce formalisme permet l'intégration de différents types de données. Dernier avantage, la simulation de systèmes d'EDO est un domaine bien connu et des outils de simulation efficaces ont déjà été développés.

La limitation de ce formalisme est qu'il considère que, pour chaque acteur, la quantité présente est élevée. Nous supposons que c'est le cas dans nos modèles. S'il s'avère que les concentrations sont trop faibles, il faudra alors utiliser un formalisme plus adapté. Les équations stochastiques permettraient de nous affranchir de cette potentielle limitation, mais leur simulation requiert une puissance de calcul importante. Nous décidons donc de ne pas les utiliser en première approche.

Le choix des EDO pour représenter la cinétique des réactions ouvre le volet méthodologique de ce travail, qui consiste à développer une approche pratique fondée sur le formalisme des EDO face à un manque de connaissances et/ou de grandes incertitudes sur la valeur des paramètres.

1.4 État de l'art : formalisation des données

Pour obtenir les modèles qui sont les plus proches possible de la réalité, les données étant rares, il est nécessaire d'intégrer chaque parcelle de données disponible. Nous avons donc besoin de pouvoir exprimer tous les types de données. Nous modélisons le système biologique par un système d'équations différentielles. Du fait de la faible quantité de données sur les paramètres de ce système d'EDO et du nombre de paramètres, nous devons travailler dans un espace de grande dimension où les intervalles pour chaque dimension sont larges.

Les informations que nous pouvons tirer des données expérimentales sont de différents types. Nous présentons ci-dessous différentes manières de prendre en compte les données disponibles. Certaines manières permettent de ne prendre en compte que certains types de données, alors que d'autres permettent de considérer les données dans leur ensemble.

1.4.1 Relations et intervalles de paramètres

Certaines expériences ont pour but de déterminer directement un paramètre cinétique, ou la valeur du seuil et de l'exposant d'une sigmoïde. Les résultats de ces expériences peuvent être

représentés directement par un intervalle de valeur pour le (ou les) paramètre, cet intervalle étant pris plus ou moins large, selon la fiabilité et l'erreur des mesures. En effet, il faut prendre en compte l'erreur due à la méthode expérimentale, ainsi qu'à la variabilité des échantillons. D'autre part, du fait du manque de données, on est souvent amené à utiliser des données obtenues sur le même type cellulaire que celui étudié, mais d'un organisme différent, plus ou moins éloigné, ou même obtenues sur des types cellulaires différents. Pour les mesures *in vitro*, la pertinence elle-même des mesures peut parfois être mise en question : il peut arriver que, dans le contexte de la cellule, une protéine interagisse avec un autre composant et que son mode d'action ou sa cinétique soit changés.

Les mesures effectuées ne fournissent pas toujours une valeur absolue pour un paramètre, mais peuvent donner des valeurs relatives. C'est, par exemple, le cas pour la plupart des mesures de l'expression génique d'ARNm par RT-qPCR². Dans cette situation, les données extraites des expériences sont représentées sous forme de relations (égalités ou inégalités) liant plusieurs expressions contenant des variables et des paramètres du modèle.

1.4.2 Logiques temporelles

Des mesures donnant des informations sur le comportement du système biologique sont souvent disponibles. Il s'agit en général d'expériences dans lesquelles le système est perturbé à un instant initial t_0 et plusieurs grandeurs sont mesurées à $t_0 + \Delta t$. On parle de séries temporelles, lorsque les mesures sont effectuées en plusieurs points. Dans ces expériences, les observations peuvent être de nature qualitative (par exemple, telle concentration est plus faible dans telle condition expérimentale que dans telle autre) ou quantitative (par exemple, mesure absolue ou relative d'une concentration). Le formalisme des logiques temporelles est bien adapté pour représenter formellement les informations de type comportemental.

Les logiques temporelles ont été initialement créées pour vérifier formellement que des programmes informatiques respectent certaines propriétés. Pour cela, des formules logiques, appliquées à un état (que nous notons ici s_0), permettent de tester si une séquence d'états discrets s_0, s_1, \dots, s_n respecte des contraintes temporelles. Diverses évolutions des logiques temporelles ont été développées, permettant leur utilisation sur différents types de séquences. Initialement, la logique LTL (pour *Linear Temporal Logic*) permet l'évaluation booléenne d'une séquence discrète de valeurs booléennes. Cette logique comporte plusieurs opérateurs temporels dont les deux suivants :

- la formule *ev* (φ) (où φ est une formule LTL) qui, appliquée à un état s_0 , est vraie s'il existe un état s_i avec $i \geq 0$ où la formule φ est vérifiée. Cet opérateur est aussi noté « \square » ou « G » dans la littérature ;
- la formule *alw* (φ) (où φ est une formule LTL) qui, appliquée à un état s_0 , est vraie si la formule φ est vraie dans tous les états s_i où $i \geq 0$. Cet opérateur est aussi noté « \diamond » ou « F » dans la littérature.

La logique MITL permet l'évaluation booléenne d'une séquence continue de valeurs booléennes et l'utilisation d'intervalles de temps (borné ou non) aux opérateurs *ev* et *alw*. Dans cette logique, les formules temporelles ne sont plus appliquées à un état, mais à un instant de la séquence, celle-ci n'étant plus discrète. Par conséquent, la formule *ev*_[1, 2] (φ), appliquée à un instant t_0 , est vraie s'il existe un instant $t_1 \in [t_0 + 1, t_0 + 2]$ où la formule φ est vrai. La logique STL (pour

2. La RT-qPCR est une technique évaluant les quantités d'ARNm de manière relative. Elle est donc utilisée pour comparer la concentration dans différentes conditions expérimentales (par exemple : présence ou absence de fer). Pour cela, elle utilise un ARNm de référence dont la concentration est constante dans les différentes conditions.

Signal Temporal Logic [38] est une évolution de la logique MITL. Elle propose une évaluation booléenne d'une séquence continue de variables réelles. Dans cette logique, les atomes sont des inégalités. Cela permet d'exprimer facilement des comportements, typiquement transitoires, qui ont été observés. Par exemple, la formule φ (donnée par l'équation (1.2)), appliquée à un instant t_0 , est vérifiée s'il existe un instant $t_1 \geq t_0$ tel que la variable x soit inférieure à 1 et que, durant l'intervalle de temps $[t_1 + 2, t_1 + 5]$, la valeur de x soit toujours supérieure à 3.

$$\varphi = ev \left((x[t] < 1) \wedge alw_ [2, 5] (x[t] > 3) \right) \quad (1.2)$$

Les formules STL permettent également d'exprimer des relations entre paramètres. Plus récemment, Donzé *et al.* [39, 40] ont présenté une nouvelle sémantique quantitative de STL. Cette logique possède donc deux sémantiques : une première, booléenne, indiquant si une formule est vérifiée ou non, et une deuxième, quantitative, fournissant une mesure du degré de robustesse ou de violation de la formule.

Dans le cadre de la modélisation de systèmes biologiques, la dynamique du système biologique est modélisée par les trajectoires des simulations des systèmes d'équations différentielles ou des systèmes hybrides modélisant ce système biologique. L'expression de connaissances expérimentales sur le comportement du système biologique se traduit alors en propriétés sur les séquences constituées par les simulations du modèle. L'utilisation de la logique STL et de sa sémantique quantitative est donc bien appropriée.

1.4.3 Au sein d'outils d'analyse

Différents outils d'analyse intègrent nativement une représentation des données, cette représentation des données n'étant pas, en soi, la finalité de ces outils. Par exemple, certains outils statistiques représentent les données par un tableau de concentrations (éventuellement accompagnées d'écarts-type) à différents instants.

Outils analytiques

Un ensemble d'outils analytiques est classiquement utilisé pour étudier les systèmes dynamiques. Nous pouvons par exemple citer l'étude des bifurcations ou l'analyse de sensibilité.

Un système d'EDO peut avoir différents comportements asymptotiques en fonction des valeurs de ses paramètres. On dit qu'une bifurcation intervient lorsqu'un faible changement de la valeur d'un paramètre produit un changement qualitatif de la dynamique du système. Par exemple, selon la valeur d'un paramètre, les simulations d'un système d'équations différentielles peuvent converger, soit vers un état stationnaire, soit vers un cycle limite [41]. En biologie systémique, ce paramètre peut être une entrée du modèle décrivant l'environnement du système biologique, la bifurcation modélisant un changement qualitatif du comportement du système biologique en réponse au changement de l'environnement. La théorie des bifurcations requiert une étude analytique des équations qui peut être complexe à mettre en œuvre dans le cas de gros systèmes d'équations différentielles non-linéaires couplées.

L'analyse de sensibilité indique la sensibilité des simulations à une variation d'un paramètre, c'est-à-dire qu'elle indique de combien une simulation change si la valeur d'un paramètre du système change d'une valeur donnée. En biologie systémique, la robustesse (propriété permettant à un système de maintenir ses fonctions malgré une perturbation [42]) est généralement mesurée par un calcul de sensibilité. Il existe deux types de sensibilité : la sensibilité locale et la sensibilité globale. La sensibilité locale est calculée pour un système autour d'une valeur pour les paramètres et les conditions initiales. La sensibilité globale calcule la sensibilité des simulations, moyennée

sur un ensemble de valeurs pour les paramètres et les conditions initiales. L'analyse de sensibilité est une étude relativement locale et est bien adaptée lorsque les incertitudes sur les paramètres sont de petite amplitude, ce qui n'est pas notre cas.

Outils statistiques

Approximate Bayesian computation (ABC) [43] est une méthode probabiliste basée sur l'utilisation de statistiques bayésiennes. Contrairement aux méthodes d'inférence statistique classiques, elle ne s'appuie pas sur une fonction de vraisemblance, mais se base sur l'utilisation intensive de simulations. En effet, pour trouver les paramètres d'un modèle, chaque paramètre est échantillonné selon une loi de distribution. Puis, pour chaque valeur des paramètres, le comportement du modèle instancié est évalué en utilisant une statistique descriptive qui compare les données réelles au comportement du modèle. L'ensemble des valeurs de paramètres pour lesquelles le comportement du modèle est proche des données permet d'inférer la distribution des paramètres du modèle. Par ailleurs, si différents modèles sont considérés, ABC propose un critère permettant de déterminer quel modèle est le plus conforme aux données.

Cette méthode, comme d'autres méthodes statistiques utilisant des fonctions de vraisemblance [44], est utilisable pour rechercher les valeurs des paramètres d'un système d'EDO modélisant un système biologique [45]. Pour cela, il est nécessaire de définir une statistique descriptive indiquant quantitativement si le jeu de paramètres vérifie les données. Cette statistique peut, par exemple, être basée sur l'utilisation de l'évaluation quantitative des formules STL fournie par l'extension de la logique STL décrite dans la sous-section 1.4.2.

Ces méthodes statistiques permettent d'exprimer autant de données que les formules STL et sont adaptées lorsque les données sont bruitées et permettent de discriminer entre différents modèles. Toutefois, ces méthodes présentent l'inconvénient de nécessiter un nombre particulièrement important de simulations, ce qui nous empêche de les appliquer à notre problème.

Programmation par contraintes

Hickey et Wittenberg [46] proposent une approche formelle pour l'étude des systèmes hybrides dont la dynamique des variables continues est décrite par des équations différentielles non-linéaires. L'objectif de ces travaux est de proposer un cadre permettant de prouver des propriétés de sécurité. Pour cela, les auteurs utilisent un solveur d'équations différentielles, couplé à un solveur sur intervalles, qui approxime les équations différentielles par un développement en séries de Taylor. Ce solveur d'EDO est intégré au langage de programmation par contraintes CLP(F)³, ce qui permet l'expression, en tant que contraintes, de données sur le système modélisé, telles des inégalités entre paramètres, ou encore d'imposer la valeur d'une variable à un instant donné. Du fait de la cohérence du langage CLP(F), il est garanti que l'ensemble des solutions aux requêtes demandées est renvoyé, permettant alors d'inférer des propriétés du système modélisé.

Cette approche n'est, néanmoins, pas applicable à l'exploration de larges espaces ni à la simulation sur de longues périodes, du fait de la puissance de calcul requise. De plus, sa description des solutions en tant qu'union d'hyper-rectangles n'est pas adaptée à la description de solution en grande dimension.

3. Un programme CLP(F) est un programme Prolog auquel sont ajoutées des contraintes sur les fonctions et leurs dérivées.

1.4.4 Choix du formalisme des données

Nous décidons que les données quantitatives sur les paramètres du modèle sont modélisées en associant, à chaque paramètre, un intervalle décrivant les valeurs possibles pour ce paramètre. Cette représentation permet de facilement prendre en compte les incertitudes sur ces données. De la même manière, nous associons un intervalle à chaque condition initiale afin de couvrir l'ensemble des possibilités.

Certaines expériences ne fournissent pas de résultats quantitatifs, mais au contraire, des résultats qualitatifs. Nous utilisons des relations (égalités comme inégalités) liant plusieurs variables et paramètres du modèle pour modéliser ces données.

L'utilisation de formules temporelles portant sur des séquences continues de valeurs réelles permet d'exprimer aisément des résultats d'expériences biologiques. Nous décidons donc de traduire l'ensemble des données disponibles en formules STL. L'évaluation de ces formules peut se faire en utilisant l'outil *Breach*, développé par Alexandre Donzé [47]. Cet outil intègre nativement la simulation d'équations différentielles, ainsi que l'évaluation booléenne et quantitative de formules STL. L'évaluation quantitative des formules permet alors de comparer différentes simulations.

Le solveur d'équations différentielles utilisé dans *Breach* est CVODES [48]. Ce solveur, écrit en C, permet de simuler efficacement des équations non-linéaires raides en utilisant une méthode d'intégration à pas multiple (linéaire) et à ordre variable, nommée Backward Differentiation Formula.

1.5 Plan

Après ce premier chapitre d'introduction, la suite de ce document traite de la modélisation du réseau du fer. Dans le chapitre 2, nous présentons le système biologique que nous modélisons. La sous-section 2.7.1 de ce chapitre introduit le scénario d'étude utilisé pour l'analyse des modèles créés. Puis, nous présentons les trois modèles créés. Le premier, présenté dans le chapitre 3, est un modèle préliminaire ayant pour objectif de proposer une représentation mathématique de tous les processus biologiques considérés. De plus, ce modèle est utilisé pour tester certains outils pour l'étude de modèles. Toutefois, ce modèle ne fait pas l'objet d'une étude approfondie. La sous-section 3.2.1 de ce chapitre décrit la modélisation du scénario d'étude. Le chapitre 4 présente un modèle se concentrant sur la régulation par les IRP. La méthodologie créée est décrite de manière détaillée, puis est appliquée au modèle. Une présentation des différents modèles du réseau du fer présents dans la littérature, ainsi qu'une comparaison avec le modèle étudié est faite dans la section 4.4 de ce chapitre. Le chapitre 5 introduit le troisième modèle décrivant explicitement les concentrations d'ARN messenger et représentant plus finement la régulation par le mécanisme de régulation. Ce modèle permet d'aborder une question relative au mécanisme de régulation et d'intégrer des données expérimentales obtenues par nos collaborateurs. Lors de l'étude de ce troisième modèle, nous découvrons une limitation de la méthodologie que nous développons. La méthodologie est alors adaptée afin de dépasser cette limitation. Ce processus d'amélioration est présenté dans la section 5.5.

Ce document contient également deux annexes. L'annexe A présente l'ensemble des données biologiques utilisée lors de l'étude du troisième modèle. La deuxième annexe (annexe B) présente des travaux portant sur l'implémentation des réseaux de Thomas en utilisant le paradigme logique ASP (*Answer Set Programming*).

Chapitre 2

Description biologique

De nombreuses molécules sont impliquées dans les processus liés à l'utilisation et la régulation du fer au sein des cellules. Dans ce chapitre, nous décrivons de manière synthétique les protéines, ARNs et autres acteurs centraux que nous avons considérés dans nos modèles ainsi que les processus biologiques dans lesquels ces acteurs sont impliqués. Nous présentons également le scénario que nous utilisons pour l'étude de ce système, ainsi que les hypothèses simplificatrices considérées. Cette description complète du système biologique sert de base pour la création des différents modèles.

2.1 Introduction

Le fer est un métal indispensable au fonctionnement du corps humain. Il intervient dans plusieurs processus tels que le transport de l'oxygène, la synthèse de l'ADN ou encore le métabolisme énergétique [49]. Une carence en fer au niveau cellulaire provoque l'arrêt de la croissance et mène à la mort de la cellule [50]. D'autre part, la présence non contrôlée de fer dans une cellule catalyse la formation d'espèces réactives de l'oxygène (aussi nommées ROS pour *Reactive Oxygen Species*) provoquant des dégâts cellulaires [51] pouvant affecter irréversiblement l'organisme entier [52]. La carence comme la surcharge en fer provoquant des dégâts, voire la mort cellulaire, un mécanisme de régulation fin de la concentration du fer au sein des cellules s'avère être nécessaire.

Selon les différents types cellulaires, la demande en fer ainsi que son utilisation ne sont pas les mêmes. Les cellules de la lignée érythrocytaire sont les cellules de l'organisme qui consomment le plus de fer. En effet, la production d'hémoglobine nécessite environ 20 mg de fer chaque jour, ce qui constitue 80 % de la demande en fer de l'organisme complet [4]. Nous nous intéressons aux cellules progénitrices pluripotentes pour lesquelles la disponibilité en fer peut potentiellement jouer un rôle dans l'engagement vers la lignée érythrocytaire.

Plusieurs mécanismes de régulation du fer existent. Selon les mécanismes considérés, la portée de la régulation varie de l'organisme entier à la cellule. Le mécanisme de régulation IRP-IRE, agissant au niveau cellulaire, est un mécanisme fondamental [53]. Ce mécanisme se base sur la régulation de la traduction d'ARN messagers, dont ceux codant des protéines impliquées dans les processus d'entrée, de sortie, d'utilisation et de stockage du fer [54].

Dans ce chapitre, nous commencerons par l'introduction, dans la section 2.2, du fonctionnement du mécanisme de régulation IRP-IRE. Suivra ensuite la présentation des mécanismes d'entrée et d'export du fer au sein et hors de la cellule (section 2.3). Nous continuerons par la description du mécanisme de stockage du fer dans la cellule (section 2.4) et nous finirons par la

présentation du processus de synthèse de l'hème, qui est le consommateur principal du fer chez les cellules érythroïdes (section 2.5). L'ensemble de ces processus est représenté graphiquement par la figure 2.1. La fin du chapitre s'attache à présenter les considérations biologiques générales sur le système en vue de la création de modèles ainsi que le scénario d'étude du système (section 2.7).

2.2 Fonctionnement du mécanisme IRP-IRE

La protéine IRP (signifiant *Iron Regulatory Protein*) est une protéine clef dans la régulation du fer, située dans le cytosol. Certains ARN messagers (ou ARNm) possèdent une structure particulière sur laquelle un IRP peut se fixer, régulant alors la traduction de la protéine codée par cet ARNm. Cette structure, en forme d'épingle à cheveux, se nomme *Iron Responsive Element* (aussi abrégée en IRE). Elle est située dans une partie non traduite de l'ARNm, soit en amont de la partie codante (zone nommée 5'-UTR, où UTR vient de l'anglais *UnTranslated Region*), soit en aval (zone nommée 3'-UTR).

Le terme IRP recouvre deux protéines homologues. La première, IRP1, existe sous deux formes. En présence de fer dans la cellule, la protéine intègre un agrégat contenant quatre atomes de fer et quatre de soufre (ces agrégats sont notés [4Fe-4S]) lui conférant une activité enzymatique. Sous cette forme, elle catalyse la conversion réversible du citrate en iso-citrate dans le cytoplasme, et elle est appelée aconitase (notée ACO1). En l'absence de fer, la protéine ne possède pas d'agrégat fer-soufre et n'a pas d'activité enzymatique, mais présente une activité de reconnaissance de motifs IRE qui lui donne un rôle de régulateur post-transcriptionnel [4]. La deuxième protéine, nommée IRP2 [56], n'est présente que dans certaines conditions, comme en cas de carence en fer, et ne possède pas de bi-fonctionnalité. Les deux protéines ont une affinité similaire pour les IRE [57].

Lorsqu'un ARNm contient un IRE dans sa partie 5'-UTR, la fixation d'un IRP sur cet IRE empêche la formation du complexe d'initiation de la traduction [58], inhibant donc sa traduction en protéine. Au contraire, lorsqu'un ARNm contient un (ou plusieurs) IRE dans sa partie 3'-UTR, la fixation d'IRP empêche la dégradation de cet ARNm, le stabilisant donc.

Par ailleurs, le fer influence l'activité des IRP. En effet, en présence de fer, différents mécanismes inhibent l'activité régulatrice des IRP :

- IRP1 est une protéine qui peut avoir plusieurs fonctions. En présence de fer, elle se complexe avec un agrégat fer-soufre [4Fe-4S], acquérant alors une fonction aconitase. Cette complexation rend inaccessible le site de reconnaissance des IRE et empêche leur fixation [49].
- La protéine FBXL5 est une protéine possédant un site spécifique sur lequel peut se fixer du fer. Cette fixation de fer stabilise la protéine, autrement instable. Cette protéine stabilisée se fixe sur IRP2 et catalyse son ubiquitination, provoquant donc sa dégradation par le protéasome [59].

2.3 Entrée et sortie du fer au sein de la cellule

La transferrine, que nous désignerons Tf, est une molécule circulant dans le sang et les liquides extra-cellulaires [54], et qui a la capacité de fixer deux atomes de fer, formant un complexe Tf-Fe₂.

Les récepteurs de la transferrine, dont le symbole est TfR1 (pour *Transferrin Receptor 1*), sont des dimères situés sur la membrane des cellules sur lesquels viennent se fixer les complexes

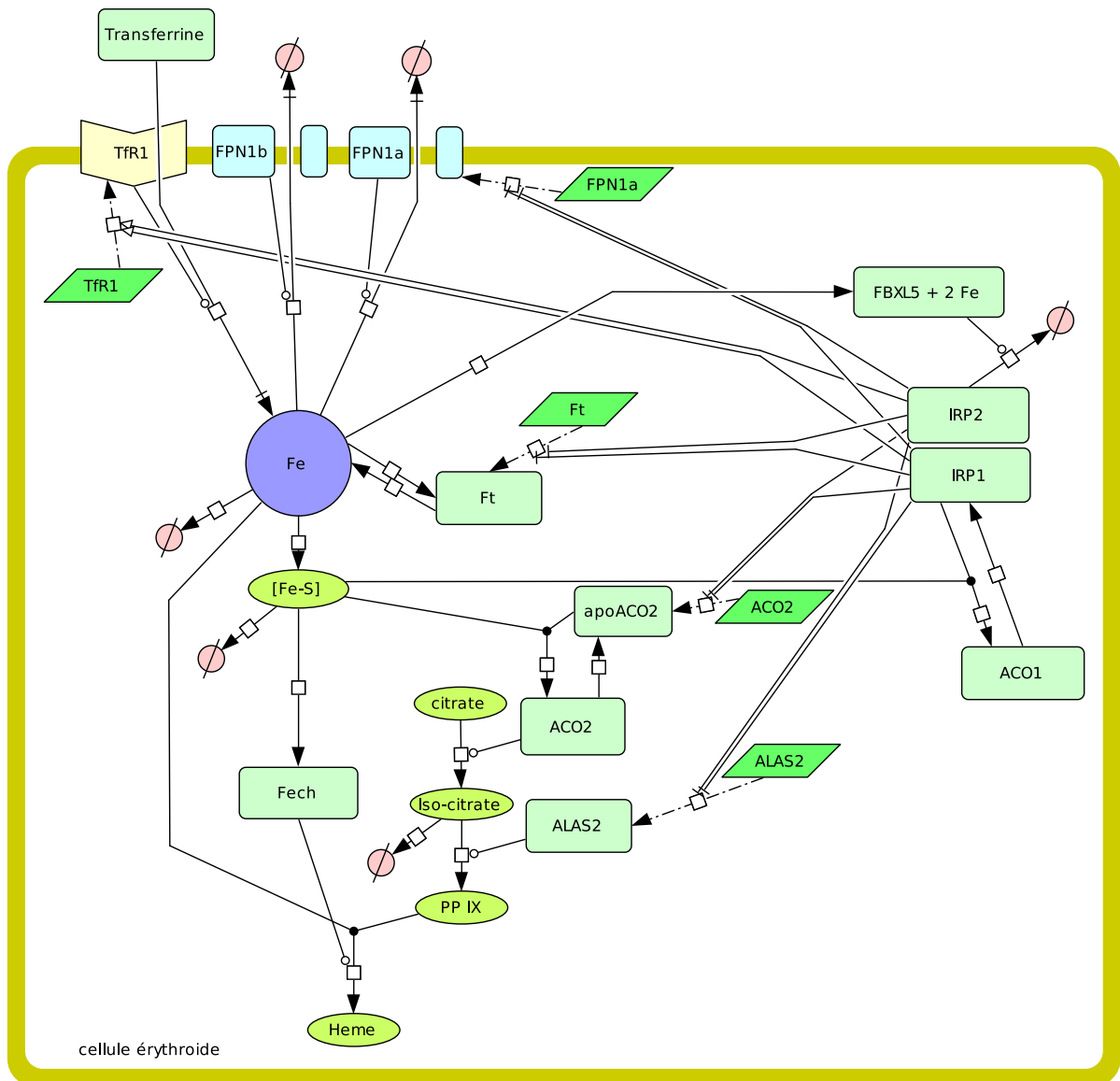


FIGURE 2.1 – Cette figure décrit le réseau de régulation du fer. La large bande jaune définit la membrane cellulaire. Les rectangles aux angles arrondis représentent des protéines, les parallélogrammes représentent des ARN messagers et l'hexagone concave représente un récepteur membranaire. Le rond portant le label « Fe » symbolise le pool de fer disponible et les deux symboles portant les labels FPN1a et FPN1b représentent les exporteurs du fer. Les ovales représentent les autres molécules (des métabolites). Les flèches en pointillés représentent la traduction de l'ARN messenger en protéine. Les flèches se terminant par une barre perpendiculaire et celles se terminant par une flèche vide traduisent respectivement une inhibition ou une activation de la traduction. Celles se terminant par un rond vide représentent une catalyse. Les flèches menant à un symbole vide représentent soit un export (du fer), soit une utilisation pour les besoins internes de la cellule (de fer, sous forme atomique ou d'agrégat fer-soufre, et d'iso-citrate), soit une dégradation (de IRP2). La dégradation des autres protéines n'est pas représentée. Les flèches se finissant par un trait perpendiculaire et une flèche symbolisent le transport du fer. Les autres flèches traduisent une réaction de transformation, excepté les deux flèches entre le fer et la ferritine qui traduisent le chargement de la ferritine en fer et le relargage de fer à partir de ferritine. Les flèches contenant un rond noir en leur milieu représentent une complexation. La signification des symboles est donnée dans le texte. Cette figure a été réalisée avec le logiciel CellDesigner [55].

Tf-Fe₂. Cette fixation provoque l'endocytose⁴ du complexe TfR1-(Tf-Fe₂). Le fer est séparé de la transferrine et est libéré au sein de la cellule [60, 61]. Les récepteurs retournent à la surface de la cellule, permettant la fixation d'autres transferrines. Ce mécanisme d'entrée du fer est régulé : l'ARNm de TfR1 possède cinq structures IRE situées dans la zone 3'-UTR, permettant ainsi une régulation par les IRP [62, 63].

Il existe divers autres transporteurs permettant l'entrée du fer au sein des cellules. L'importance de ces récepteurs varie selon, entre autres, le type cellulaire. Chez les érythrocytes, ces transporteurs sont peu exprimés [53, 54, 64] et participent peu à l'entrée en fer [65], le mécanisme principal d'entrée du fer passant par TfR1 [53, 66].

La ferroportine, dont le symbole est FPN, est une protéine située à la membrane cellulaire, permettant l'export du fer hors de la cellule. Cette protéine, qui est un dimère⁵ [67], est le seul exporteur de fer connu [68, 69]. Il existe plusieurs types de transcrits pour la ferroportine. Un de ces transcrits, nommé FPN1a, possède une boucle IRE située dans la zone 5'-UTR, permettant ainsi une régulation par les IRP [70]. FPN1b est un autre type de transcrit qui ne possède pas d'IRE et n'est donc pas régulé par les IRP. La quantité de transcrits de la ferroportine ne possédant pas d'IRE est au moins aussi importante que celle en possédant un [71].

2.4 Stockage du fer

La ferritine (dont l'abréviation est Ft) est une molécule permettant le stockage du fer qui se retrouve chez de nombreux organismes vivants, allant des bactéries aux mammifères [72]. La structure de cette molécule est bien caractérisée chez les mammifères : une ferritine est constituée de 24 sous-unités formant une coquille creuse, lui conférant une capacité de stockage. Chaque protéine peut contenir jusqu'à 4500 atomes de fer [73]. Le processus d'entrée du fer dans la ferritine se fait par une succession de réactions d'oxydo-réduction, impliquant certaines sous-unités [74, 75]. Le fer est ainsi stocké sous une forme non toxique pour la cellule et peut être conservé pour des besoins ultérieurs. La sortie du fer hors de la ferritine est suivie par l'ubiquitination de la coquille vide, provoquant sa dégradation [76].

La production de cette protéine est également contrôlée par le mécanisme de régulation étudié : l'ARN messager de la ferritine contient un IRE dans la région 5'-UTR, permettant donc une régulation de sa traduction par les IRP [77, 78].

Le fer stocké dans les ferritines n'est pas disponible immédiatement puisqu'il faut qu'il soit d'abord extrait de ces structures avant d'être utilisable. Une autre réserve de fer, nommée *Functional Iron Concentration* (abrégé en FIC) [79], est généralement considérée. La nature chimique de cette réserve est sujette à débat, ainsi que sa localisation dans la cellule [61, 80, 81]. Le fer contenu dans ce FIC, accessible directement, est celui utilisé pour les besoins de la cellule. Ces besoins incluent la synthèse d'agrégats fer-soufre (notés [Fe-S]) qui seront insérés au sein de protéines. Nous représentons explicitement ces agrégats, bien qu'ils n'existent pas sous forme libre dans la cellule.

4. L'endocytose est un mécanisme durant lequel une petite partie de la membrane de la cellule est invaginée et forme une vésicule à l'intérieur de la cellule. Cette vésicule (également appelée endosome) peut transporter des complexes ligand-récepteur comme TfR1-(Tf-Fe₂) dans diverses localisations intra-cellulaires.

5. Un dimère est une protéine comportant deux sous-unités.

2.5 Synthèse de l'hème

De par leur nature, les cellules érythroïdes synthétisent de grandes quantités d'hémoglobine. Le site actif de l'hémoglobine est constitué par un hème. Celui-ci est synthétisé au sein de la mitochondrie et résulte de l'insertion d'un atome de fer dans une molécule de protoporphyrine IX (abrégée en PP IX). La synthèse de cette protoporphyrine utilise un intermédiaire du cycle de Krebs et nécessite l'action catalytique de l'enzyme ALAS2 [82].

En amont dans le cycle de Krebs, la succession de réactions menant à la synthèse de cet intermédiaire comporte la transformation du citrate en iso-citrate sous l'action catalytique de l'aconitase ACO2 [83]. Bien que cette réaction soit réversible, l'abondance du citrate et la consommation de l'iso-citrate au sein du cycle de Krebs font que la conversion d'iso-citrate en citrate n'a, en pratique, pas besoin d'être considérée. La protéine ACO2 est produite par une réaction dans laquelle l'apo-protéine apo-ACO2 intègre un agrégat fer-soufre. La synthèse de PP IX est fortement liée au mécanisme de régulation par les IRP puisque les ARN messagers du catalyseur ALAS2 et de l'apo-protéine apo-ACO2 contiennent un IRE dans la région 5'-UTR, provoquant donc une diminution de leur traduction lorsque la concentration d'IRP augmente [83, 84].

L'insertion de l'atome de fer au sein de la protoporphyrine PP IX est catalysée par une enzyme : la ferrochélatase (abrégée en Fech). La stabilité de cette dernière est fer-dépendante : l'insertion de deux agrégats fer-soufre $[2\text{Fe}-2\text{S}]$ au sein de l'enzyme la stabilise [85, 86]. La synthèse d'hème est donc fortement régulée par le fer.

2.6 Autres acteurs

Dans cette présentation, certaines réactions sont simplifiées car les détails de ces réactions n'ajoutent pas d'information. Par exemple, la transformation de l'iso-citrate en protoporphyrine IX se fait via la synthèse d' α -cétoglutarate par certaines isozymes d'isocitrate deshydrogénase [87], puis via la transformation d' α -cétoglutarate en succinyl-CoA.

Différents signaux influencent aussi, à différents niveaux, le réseau de régulation du fer. Au niveau cellulaire, ces signaux incluent le monoxyde d'azote [88], plusieurs facteurs de transcription [69, 89] ou le réseau de régulation du stress oxydant [90]. De plus, des micro-ARNs interagissent avec des acteurs du réseau de régulation du fer [91]. Par exemple, le micro-ARN miR-485-3p régule post-transcriptionnellement la ferroportine [92]. Par ailleurs, l'hepcidine est un peptide sécrété principalement par les cellules hépatiques. Ce peptide, circulant dans l'organisme entier, régule la ferroportine. Lorsqu'il est sécrété, il vient se fixer sur la ferroportine, provoquant l'internalisation, suivi de la dégradation de cette dernière [93, 94]. Nous ne prenons pas en compte ces signaux dans nos modèles car nous considérons qu'ils sont constants au cours du temps dans les conditions expérimentales auxquelles le modèle doit s'appliquer.

2.7 Contexte d'étude et considérations biologiques

2.7.1 Contexte d'étude

La situation à laquelle nous nous intéressons ici est la réaction du système face à une situation de carence en fer. Les expériences biologiques sur lesquelles nous nous basons sont la mise en carence de cellules de lignées cellulaires (principalement de la lignée K562⁶) à partir d'une

6. Les cellules de la lignée K562 sont des cellules leucémiques pluripotentes cancéreuses issues d'une patiente ayant une leucémie myéloïde chronique.

situation où le fer est en abondance. Les expériences étant faites pour de grandes quantités de cellules, nous modélisons le comportement moyen d'une population de cellules, plutôt que le comportement d'une cellule unique.

2.7.2 Considérations biologiques

Différents compartiments existent au sein de la cellule, comme la mitochondrie, le noyau, le cytosol, etc. Pour la plupart des réactions présentées ici, l'ensemble des réactifs provient du compartiment où a lieu la réaction. Néanmoins, la complexation de IRP1 avec un agrégat fer-soufre nécessite que ce dernier passe de la mitochondrie vers le cytosol. Afin de ne pas augmenter la complexité des modèles, durant la modélisation de ce système biologique, nous n'avons pas considéré de séparation en différents compartiments, bien que certaines des réactions aient lieu dans la mitochondrie et d'autres dans le cytosol. Si on désire prendre en compte le changement de compartiment de l'agrégat fer-soufre avant la complexation avec IRP1, il suffit de considérer une méta-réaction incluant le changement de compartiment et la complexation et d'ajuster les paramètres cinétiques de cette méta-réaction en conséquence. Toutefois, ceci n'est pas fait dans les différentes modélisations présentées.

Par ailleurs, du fait de l'incertitude de la localisation du pool de fer, nous considérons un pool de fer qui peut être utilisé autant pour des réactions ayant lieu au sein de la mitochondrie (comme la création d'agrégat fer-soufre) que hors de la mitochondrie (comme la stabilisation de FBXL5).

Chapitre 3

Modèle préliminaire

À partir de la description du système biologique faite dans le chapitre 2, nous créons un modèle préliminaire, intégrant la totalité des processus décrits. Ce modèle permet de représenter mathématiquement l'ensemble des processus impliqués et de tester diverses méthodes pour sa simplification et son étude. Ce modèle ne fait pas l'objet d'une étude approfondie, mais sert de point de départ pour la création de modèles plus simples, dont les études sont présentées dans les chapitres 4 et 5.

Dans la suite de ce chapitre, nous décrivons, dans la section 3.1, le système d'équations différentielles modélisant l'utilisation et la régulation du fer pour la synthèse de l'hème. Puis la modélisation du scénario d'étude ainsi que des pistes pour l'analyse de ce système d'équations sont présentées dans la section 3.2.

3.1 Obtention d'un modèle mathématique

Dans cette section, nous résumons les hypothèses de modélisation faites, puis nous explicitons les conventions d'écriture utilisées. Ensuite, nous décrivons et expliquons les équations différentielles modélisant le système biologique.

3.1.1 Hypothèses de modélisation

Nous considérons que la dégradation des aconitases ou leur décomplexation avec l'agrégat fer-soufre provoque aussi la dégradation de l'agrégat.

Le mécanisme de stockage de fer dans les ferritines, ainsi que celui du re-largage de ce fer au sein de la cellule, n'est pas complètement connu. Nous faisons l'hypothèse que les ferritines synthétisées sont immédiatement remplies en fer et que ce fer n'est relâché que lorsque les ferritines sont dégradées. Par ailleurs, nous considérons que toutes les ferritines stockent le même nombre d'atomes de fer.

De plus, la cinétique de plusieurs réactions, telles la régulation par les IRP ou l'insertion d'agrégats fer-soufre n'est pas connue. Ces cinétiques sont alors supposées, comme indiqué dans les paragraphes respectifs.

En dernier lieu, comme indiqué dans la sous-section 2.7.2, les différents compartiments des cellules ne sont pas pris en compte dans cette modélisation et un seul pool de fer mobilisable, disponible pour l'ensemble des réactions nécessitant du fer, est considéré.

3.1.2 Conventions de notation

Afin de faciliter la lecture de ces équations, un ensemble de conventions d'écriture ont été fixées. Nous les listons ci-dessous.

- Les vitesses de production sont notées $k_{\dots_{prod}}$ où « ... » est remplacé par le nom de la protéine concernée. La vitesse de production de FBXL5, dans l'équation (3.10), est donc notée $k_{FBXL5_{prod}}$.
- Les constantes de vitesse de dégradation spontanée sont notées $k_{\dots_{deg}}$ où « ... » est remplacé par le nom de la protéine concernée. La constante de vitesse de dégradation de ACO2, dans l'équation (3.6), est donc notée $k_{ACO2_{deg}}$.
- Les constantes de vitesse des réactions métaboliques sont indicées par la description de la réaction. La constante de vitesse de second ordre de la complexation de IRP1 et de l'agrégat fer-soufre en ACO1, dans les équations (3.7) et (3.8), est donc notée $k_{IRP1+FeS \rightarrow ACO1}$.
- Les constantes de vitesse des régulations sont indicées par une flèche « régulant \rightarrow régulé ». La constante de vitesse de la régulation de la traduction de la ferroportine par IRP1, dans l'équation (3.2), est donc notée $k_{IRP1 \rightarrow FPN1a}$.
- Les fonctions sigmoïdes sont définies par $\text{sig}^+(x, \theta) = \frac{x^n}{x^n + \theta^n}$ où x est une variable, θ est le seuil de la sigmoïde et n définit la raideur de la sigmoïde. Nous utilisons également $\text{sig}^-(x, \theta) = 1 - \text{sig}^+(x, \theta)$. Les seuils des sigmoïdes sont indexés par une description de la régulation. Par exemple, le seuil de la régulation de IRP1 sur la ferroportine est noté $\theta_{IRP1 \rightarrow FPN1a}$.
- Les réactions métaboliques catalysées ont une cinétique analogue à celle donnée par les équations de Michaelis-Menten [95]. La vitesse de synthèse du produit est donnée par la formule $v = \frac{[E]_0 \cdot kcat \cdot [S]}{K_M + [S]}$ où $[S]$ est la concentration du substrat, $[E]_0$ la concentration de l'enzyme, $kcat$ la constante de vitesse de premier ordre de la transformation du complexe enzyme+substrat en produit et K_M la constante de Michaelis de la réaction. Les paramètres de ces réactions sont notés $kcat$ et KM , avec en indice un descriptif de la réaction et en exposant le catalyseur. Les paramètres de la conversion du citrate en iso-citrate, dans l'équation (3.12) sont donc notés $kcat_{cit \rightarrow isocit}^{ACO2}$ et $KM_{cit \rightarrow isocit}^{ACO2}$.

3.1.3 Régulation par les IRP

Le modèle vise à décrire la réponse de la cellule à une carence en fer ainsi que l'utilisation du fer pour la synthèse d'hème. En situation de carence, l'utilisation du fer est régulée grâce à deux protéines : IRP1 et IRP2. Ces deux protéines régulent la traduction des ARN messagers de ACO2, ALAS2, FPN1a et TfR1. Les mécanismes de régulation par les deux IRP étant similaires, nous modélisons de manière identique la régulation par IRP1 et par IRP2.

L'ARN messager de TfR1 possède cinq IRE situés dans sa région 3'-UTR. La régulation, par les IRP, de la dégradation de cet ARNm ne suit pas une loi type « tout-ou-rien », mais est en lien avec l'activité des IRP (IRP1 comme IRP2) [63]. Nous décrivons alors la vitesse de production du récepteur TfR1 comme étant composée d'une vitesse basale, qui est augmentée proportionnellement à la concentration en IRP.

Dans le cas où l'ARN messager possède un IRE dans sa région 5'-UTR (comme les ARNm de ALAS2, FPN1a, Ft et ACO2), la présence d'IRP actif inhibe la traduction de cet ARNm. La cinétique de cette régulation n'est pas connue. En se basant sur les travaux de Omholt *et al.* [96], nous considérons que cette cinétique est sigmoïdale. Cette régulation est décrite par deux termes de la forme $k_{IRP \rightarrow \dots} \cdot \text{sig}^+(IRP, \theta_{IRP \rightarrow \dots})$, un pour chaque IRP, où le $k_{IRP \rightarrow \dots}$

représente la diminution de la traduction et où $\theta_{IRP \rightarrow \dots}$ est le seuil de la sigmoïde.

Les équations de *ALAS2*, *FPN1a*, *Ft* et *TfR1* peuvent alors être décrites, respectivement, par les équations (3.1), (3.2), (3.3) et (3.4). Chacune d'elles contient, en plus des deux termes de régulation par les IRP, un terme décrivant la vitesse de production basale (décrit par le premier terme de chacune des équations) et un terme de dégradation spontanée (décrit par le dernier terme des équations, de la forme $k_{\dots_deg} \times \dots$).

$$\begin{aligned} \frac{d ALAS2}{dt} &= k_{ALAS2_prod} - k_{IRP1 \rightarrow ALAS2} \cdot \text{sig}^+(IRP1, \theta_{IRP1 \rightarrow ALAS2}) \\ &\quad - k_{IRP2 \rightarrow ALAS2} \cdot \text{sig}^+(IRP2, \theta_{IRP2 \rightarrow ALAS2}) \\ &\quad - k_{ALAS2_deg} \cdot ALAS2 \end{aligned} \quad (3.1)$$

$$\begin{aligned} \frac{d FPN1a}{dt} &= k_{FPN1a_prod} - k_{IRP1 \rightarrow FPN1a} \cdot \text{sig}^+(IRP1, \theta_{IRP1 \rightarrow FPN1a}) \\ &\quad - k_{IRP2 \rightarrow FPN1a} \cdot \text{sig}^+(IRP2, \theta_{IRP2 \rightarrow FPN1a}) \\ &\quad - k_{FPN1a_deg} \cdot FPN1a \end{aligned} \quad (3.2)$$

$$\begin{aligned} \frac{d Ft}{dt} &= k_{Ft_prod} - k_{IRP1 \rightarrow Ft} \cdot \text{sig}^+(IRP1, \theta_{IRP1 \rightarrow Ft}) \\ &\quad - k_{IRP2 \rightarrow Ft} \cdot \text{sig}^+(IRP2, \theta_{IRP2 \rightarrow Ft}) \\ &\quad - k_{Ft_deg} \cdot Ft \end{aligned} \quad (3.3)$$

$$\begin{aligned} \frac{d TfR1}{dt} &= k_{TfR1_prod} + k_{IRP1 \rightarrow TfR1} \cdot IRP1 + k_{IRP2 \rightarrow TfR1} \cdot IRP2 \\ &\quad - k_{TfR1_deg} \cdot TfR1 \end{aligned} \quad (3.4)$$

Pour la synthèse de la protéine ACO2, l'apo-protéine⁷ est traduite à partir de l'ARN messager, cette traduction étant régulée par les IRP. La fixation d'un agrégat fer-soufre sur cette apo-protéine forme l'aconitase ACO2. Dans notre modélisation, afin de décrire correctement la régulation de la traduction par les IRP ainsi que la complexation avec l'agrégat fer-soufre, nous séparons ces deux étapes. D'une part, la traduction de l'ARN messager en apo-protéine est décrite par l'équation (3.5). Cette équation se compose d'un terme décrivant la vitesse de production basale, de deux termes décrivant la régulation par les IRP et d'un terme de dégradation spontanée. D'autre part, l'équation (3.6) décrivant la dynamique de ACO2, est constituée de deux termes modélisant la complexation/décomplexation de l'apo-protéine et de l'agrégat fer-soufre, ainsi que d'un terme modélisant la dégradation spontanée de l'aconitase.

$$\begin{aligned} \frac{d apoACO2}{dt} &= k_{apoACO2_prod} - k_{IRP1 \rightarrow apoACO2} \cdot \text{sig}^+(IRP1, \theta_{IRP1 \rightarrow apoACO2}) \\ &\quad - k_{IRP2 \rightarrow apoACO2} \cdot \text{sig}^+(IRP2, \theta_{IRP2 \rightarrow apoACO2}) \\ &\quad - k_{apoACO2+FeS \rightarrow ACO2} \cdot apoACO2 \cdot FeS + k_{ACO2 \rightarrow apoACO2} \cdot ACO2 \\ &\quad - k_{apoACO2_deg} \cdot apoACO2 \end{aligned} \quad (3.5)$$

$$\begin{aligned} \frac{d ACO2}{dt} &= k_{apoACO2+FeS \rightarrow ACO2} \cdot apoACO2 \cdot FeS - k_{ACO2 \rightarrow apoACO2} \cdot ACO2 \\ &\quad - k_{ACO2_deg} \cdot ACO2 \end{aligned} \quad (3.6)$$

7. Une apo-protéine est une molécule sur laquelle viendra se fixer une autre molécule (appelée cofacteur). Le complexe formé par ces deux molécules est la protéine fonctionnelle.

3.1.4 Régulation des IRP

L'activité des IRP dépend de la concentration en fer au sein de la cellule. Les deux IRP sont régulées de manière différente, la cinétique de chacune de ces régulations n'étant pas connue précisément.

En présence de fer, IRP1 est convertie en ACO1 par la fixation d'un agrégat fer-soufre. La cinétique *in vivo* de cette réaction n'étant pas connue, nous considérons qu'elle suit une loi d'action de masse, c'est-à-dire que la vitesse de réaction est proportionnelle au produit des concentrations des réactifs. Cette complexation est réversible et nous considérons la décomplexation comme étant proportionnelle à la concentration d'ACO1. Les deux termes décrivant ces deux réactions apparaissent avec un signe opposé dans les équations de *IRP1* (équation (3.7)) et de *ACO1* (équation (3.8)). Par ailleurs, l'aconitase ACO1 ne pouvant être créée que grâce à ce mécanisme de complexation, son équation ne contient pas de terme de vitesse de production basale. Le premier terme de l'équation de *IRP1* décrit cette production basale. De plus, les protéines IRP1 comme ACO1 sont spontanément dégradées, ce qui est modélisé par un terme de dégradation (le dernier terme des équations (3.7) et (3.8)).

$$\begin{aligned} \frac{d \text{IRP1}}{dt} = & k_{\text{IRP1}_{\text{prod}}} - k_{\text{IRP1}+\text{FeS}\rightarrow\text{ACO1}} \cdot \text{IRP1} \cdot \text{FeS} \\ & + k_{\text{ACO1}\rightarrow\text{IRP1}} \cdot \text{ACO1} - k_{\text{IRP1}_{\text{deg}}} \cdot \text{IRP1} \end{aligned} \quad (3.7)$$

$$\begin{aligned} \frac{d \text{ACO1}}{dt} = & k_{\text{IRP1}+\text{FeS}\rightarrow\text{ACO1}} \cdot \text{IRP1} \cdot \text{FeS} - k_{\text{ACO1}\rightarrow\text{IRP1}} \cdot \text{ACO1} \\ & - k_{\text{ACO1}_{\text{deg}}} \cdot \text{ACO1} \end{aligned} \quad (3.8)$$

La concentration d'IRP2 est régulée par le fer via la protéine FBXL5 : la présence de cette protéine accroît la dégradation de IRP2. En situation de carence, le taux de dégradation de FBXL5 augmente, provoquant une baisse de la dégradation de IRP2. La cinétique de ces deux interactions n'étant pas connue, nous représentons la dégradation de FBXL5, dans l'équation (3.10), par, d'une part, une constante de vitesse de dégradation en présence de fer ($k_{\text{FBXL5}_{\text{deg}}}$) et, d'autre part, un terme traduisant l'augmentation de dégradation en situation de carence ($k_{\text{Fe}\rightarrow\text{FBXL5}} \cdot \text{sig}^-(\text{Fe}, \theta_{\text{Fe}\rightarrow\text{FBXL5}})$). Par ailleurs, dans l'équation (3.9), nous considérons que la dégradation de IRP2 provoquée par FBXL5 suit une loi sigmoïdale.

$$\begin{aligned} \frac{d \text{IRP2}}{dt} = & k_{\text{IRP2}_{\text{prod}}} \\ & - k_{\text{FBXL5}\rightarrow\text{IRP2}} \cdot \text{sig}^+(\text{FBXL5}, \theta_{\text{FBXL5}\rightarrow\text{IRP2}}) \cdot \text{IRP2} \\ & - k_{\text{IRP2}_{\text{deg}}} \cdot \text{IRP2} \end{aligned} \quad (3.9)$$

$$\begin{aligned} \frac{d \text{FBXL5}}{dt} = & k_{\text{FBXL5}_{\text{prod}}} \\ & - k_{\text{Fe}\rightarrow\text{FBXL5}} \cdot \text{sig}^-(\text{Fe}, \theta_{\text{Fe}\rightarrow\text{FBXL5}}) \cdot \text{FBXL5} \\ & - k_{\text{FBXL5}_{\text{deg}}} \cdot \text{FBXL5} \end{aligned} \quad (3.10)$$

3.1.5 Production et consommation des agrégats fer-soufre

Dans ce modèle, la consommation de fer pour la synthèse d'agrégats fer-soufre est représentée explicitement. Ces agrégats sont utilisés dans de nombreux processus, dont la synthèse de la ferrochelatase et des aconitases ACO1 et ACO2. La cinétique des réactions consommant ces agrégats n'est pas connue. En l'absence d'information et dans un souci de simplicité, nous

considérons que la cinétique de ces réactions est la même et choisissons une cinétique type loi d'action de masse, bien que d'autres types de cinétique, telle celle de Michaelis-Menten, puissent être considérées [97]. Le premier terme de l'équation (3.11) décrit la synthèse de ces agrégats et le dernier terme décrit la consommation de ces agrégats pour les besoins internes de la cellule. Les autres termes décrivent la complexation de ces agrégats pour former, respectivement, les aconitases ACO1, ACO2 et la ferrochélatase.

$$\begin{aligned} \frac{d FeS}{dt} = & k_{Fe \rightarrow FeS} \cdot Fe - k_{IRP1+FeS \rightarrow ACO1} \cdot IRP1 \cdot FeS \\ & - k_{apoACO2+FeS \rightarrow ACO2} \cdot apoACO2 \cdot FeS - k_{FeS \rightarrow Fech} \cdot k_{Fech_prod} \cdot FeS \\ & - k_{FeS_cons} \cdot FeS \end{aligned} \quad (3.11)$$

3.1.6 Synthèse de l'hème

Cette partie, modélisant la synthèse d'hème décrit, d'une part, la synthèse de la protoporphyrine, d'autre part, la synthèse de la ferrochélatase, et finalement, la synthèse de l'hème. Pour l'ensemble de ces réactions métaboliques, nous considérons une cinétique proche de celle décrite par les équations de Michaelis-Menten.

L'équation (3.12), relative à l'iso-citrate, contient (dans l'ordre) deux termes de transformation du citrate en iso-citrate, sous la catalyse des aconitases 1 et 2 ; un terme de transformation de l'iso-citrate en protoporphyrine IX, réaction catalysée par de nombreuses enzymes dont l'enzyme ALAS2 ; et un terme d'utilisation de l'iso-citrate pour les besoins de la cellule. La dynamique de la protoporphyrine IX est décrite par l'équation (3.13). Le premier terme de cette équation traduit la synthèse de PP IX et le deuxième terme représente la complexation de PP IX avec du fer pour former de l'hème.

$$\begin{aligned} \frac{d isocit}{dt} = & \frac{kcat_{cit \rightarrow isocit}^{ACO1} \cdot ACO1 \cdot cit}{KM_{cit \rightarrow isocit}^{ACO1} + cit} + \frac{kcat_{cit \rightarrow isocit}^{ACO2} \cdot ACO2 \cdot cit}{KM_{cit \rightarrow isocit}^{ACO2} + cit} \\ & - \frac{kcat_{isocit \rightarrow PPIX}^{ALAS2} \cdot ALAS2 \cdot isocit}{KM_{isocit \rightarrow PPIX}^{ALAS2} + isocit} - k_{isocit_cons} \cdot isocit \end{aligned} \quad (3.12)$$

$$\frac{d PPIX}{dt} = \frac{kcat_{isocit \rightarrow PPIX}^{ALAS2} \cdot ALAS2 \cdot isocit}{KM_{isocit \rightarrow PPIX}^{ALAS2} + isocit} - \frac{kcat_{Fe+PPIX \rightarrow heme}^{Fech} \cdot Fech \cdot PPIX \cdot Fe}{KM_{Fe+PPIX \rightarrow heme}^{Fech} + PPIX \cdot Fe} \quad (3.13)$$

La ferrochélatase est une enzyme stabilisée par l'insertion de deux agrégats fer-soufre. Si ces agrégats sont absents, la ferrochélatase n'est pas stable. L'équation (3.14) décrit la concentration de la ferrochélatase stabilisée. Le premier terme représente la synthèse (incluant l'insertion des agrégats) de la ferrochélatase et le deuxième terme traduit la dégradation spontanée.

$$\frac{d Fech}{dt} = k_{FeS \rightarrow Fech} \cdot k_{Fech_prod} \cdot FeS - k_{Fech_deg} \cdot Fech \quad (3.14)$$

L'hème est synthétisée à partir de fer et de protoporphyrine. L'évolution de sa concentration est donnée par l'équation (3.15). Le deuxième terme de cette équation représente l'utilisation de l'hème par la cellule.

$$\frac{d heme}{dt} = \frac{kcat_{Fe+PPIX \rightarrow heme}^{Fech} \cdot Fech \cdot PPIX \cdot Fe}{KM_{Fe+PPIX \rightarrow heme}^{Fech} + PPIX \cdot Fe} - k_{heme_cons} \cdot heme \quad (3.15)$$

3.1.7 Entrée, utilisation et stockage du fer

L'entrée en fer au sein de la cellule se fait via le récepteur TfR1. La quantité de fer entrante est proportionnelle à la concentration du récepteur ainsi qu'à la quantité de fer lié aux transferrines extra-cellulaires. Ceci est traduit par le premier terme de l'équation (3.16).

L'export de fer hors de la cellule se fait via un exporteur présentant deux formes : d'une part FPN1a dont la traduction est régulée par les IRP ; d'autre part, FPN1b qui n'est pas régulé par les IRP et pour lequel nous considérons une concentration constante, représentée par le paramètre C_{FPN1b} . La quantité de fer exportée est proportionnelle à la concentration en fer et à la concentration de l'exporteur. Nous considérons que le coefficient cinétique de l'export du fer est le même pour les deux formes de la protéine. Le deuxième terme de l'équation traduit cette sortie.

Comme indiqué dans la sous-section 3.1.1, nous considérons qu'une ferritine nouvellement synthétisée se remplit immédiatement en fer et que ce fer n'est relâché que lors de la dégradation de la ferritine. Ces processus sont représentés par le troisième terme de l'équation de Fe . Le paramètre n_{Ft} représente le nombre d'atomes de fer stocké par ferritine.

Les quatre termes suivants de l'équation (3.16) représentent, dans l'ordre, la consommation de fer pour la stabilisation de FBXL5, pour la synthèse d'agrégat fer-soufre, pour la synthèse d'hème, et pour les besoins internes de la cellule.

$$\begin{aligned} \frac{d Fe}{dt} = & k_{Fe_input} \cdot TfR1 \cdot Tf_{sat} - k_{Fe_export} \cdot Fe \cdot (FPN1a + C_{FPN1b}) - n_{Ft} \cdot \frac{d Ft}{dt} \\ & - k_{Fe \rightarrow FBXL5} \cdot \text{sig}^+(Fe, \theta_{Fe \rightarrow FBXL5}) \cdot FBXL5 - k_{Fe \rightarrow FeS} \cdot Fe \\ & - \frac{k_{cat}_{Fe+PPIX \rightarrow heme}^{Fech} \cdot Fech \cdot Fe \cdot PPIX}{KM_{Fe+PPIX \rightarrow heme}^{Fech} + Fe \cdot PPIX} - k_{Fe_cons} \cdot Fe \end{aligned} \quad (3.16)$$

3.1.8 Entrées du modèle

Ce modèle possède deux entrées permettant de représenter les conditions expérimentales. La première entrée décrit la saturation en fer de la transferrine présente dans le milieu extra-cellulaire. Cette entrée, nommée Tf_{sat} , intervient dans l'équation (3.16) décrivant la dynamique du pool de fer. La modélisation d'une mise en carence en fer des cellules se fait alors facilement en mettant Tf_{sat} à zéro. L'équation régissant l'entrée est donnée par la formule (3.17) où t représente le temps, t_{cutoff} l'instant où l'entrée en fer est coupée et $Tf_{sat,0}$ la saturation de la transferrine avant la mise en carence.

$$\frac{d Tf_{sat}}{dt} = Tf_{sat,0} \cdot \text{sig}^-(t, t_{cutoff}) - Tf_{sat} \quad (3.17)$$

La mise en carence de la cellule peut provoquer des perturbations dans des processus non représentés dans notre modèle. Pour pouvoir considérer cela, la deuxième entrée du modèle traduit la concentration en citrate. Cette entrée permet de représenter, en combinaison avec le paramètre k_{isocit_cons} de l'équation (3.12), le cycle de Krebs, qui est un acteur majeur du métabolisme.

3.2 Pistes d'étude du modèle

Dans la section 3.1, nous avons donné les équations différentielles régissant l'évolution de la concentration des seize acteurs considérés. Ces acteurs sont de différentes natures puisque nous

représentons autant des enzymes que des protéines membranaires ou encore des métabolites. Les équations différentielles contiennent 67 paramètres auxquels s'ajoutent deux entrées définissant la saturation de la transferrine et la concentration du citrate. De plus, ce modèle contient différents types d'interactions puisque nous y trouvons autant des régulations traductionnelles que des réactions métaboliques, catalysées ou non catalysées.

Ce modèle est étudié dans le cadre de l'homéostasie du fer. Pour cela, nous utilisons le scénario d'étude présenté dans la sous-section 2.7.1. La sous-section 3.2.1 présente comment nous le mettons en place dans le cadre de notre modélisation.

Par ailleurs, ce modèle préliminaire contient un grand nombre de paramètres et de variables. Nous présentons trois méthodes visant à réduire ce nombre, permettant ainsi de faciliter l'étude de ce modèle. La première, présentée dans la sous-section 3.2.2, est l'adimensionnement et nécessite seulement que les unités des inconnues du modèle soient correctement décrites. La deuxième (sous-section 3.2.3) se base sur la séparation d'échelle de temps des réactions. Appliquées à ce modèle, aucune de ces deux méthodes n'a permis une réduction significative du nombre d'inconnues, mais cela nous a permis de les évaluer. La troisième, introduite dans la sous-section 3.2.4, est plus spécifique à notre étude car elle s'appuie sur la situation de départ du scénario d'étude, à savoir la situation où le fer est en abondance, et vise à décrire les concentrations en fonction des paramètres.

3.2.1 Modélisation du scénario d'étude

Comme indiqué dans la sous-section 2.7.1, le scénario que nous étudions débute par une situation où le fer est en abondance, puis consiste à appliquer une perturbation qui est une mise en carence en fer.

Dans une situation où le fer est présent, comme la situation initiale de notre scénario, les cellules se divisent. Il est connu que les concentrations de nombreuses protéines varient au cours du cycle cellulaire et sont aussi soumises au rythme circadien [98]. Cependant, dans notre étude, les expériences sont effectuées sur des populations de cellules qui ne sont pas synchronisées. Les résultats expérimentaux sont donc des moyennes sur des populations et les variations de concentration entre cellules (dues au fait qu'elles sont dans des phases différentes du cycle) ne sont pas observables. Nous considérons donc que les concentrations de protéines sont constantes (en moyenne sur une population) lorsque les cellules sont dans un milieu riche en fer. Ce régime de fonctionnement est modélisé par un état stationnaire. Autrement dit les modèles développés dans notre étude visent à représenter le comportement de populations de cellules. Dans cet état stationnaire, la saturation de la transferrine est égale à $Tf_{sat,0}$ et est donc non nulle. De plus, pour que cette modélisation soit cohérente, l'état stationnaire doit être stable. Par ailleurs, nous supposons que, dans la situation non-carencée, la concentration en fer est supérieure au seuil de régulation sur FBXL5 $\theta_{Fe \rightarrow FBXL5}$ et que la concentration de FBXL5 est supérieure au seuil $\theta_{FBXL5 \rightarrow IRP2}$. Nous supposons également que la concentration d'IRP1 (resp. IRP2) actif est inférieure aux seuils d'IRP1 (resp. d'IRP2) tels $\theta_{IRP1 \rightarrow ALAS2}$ (resp. $\theta_{IRP2 \rightarrow ALAS2}$).

Pour permettre au système d'équations différentielles d'atteindre l'état stationnaire, nous laissons une durée de six heures entre le début de la simulation et l'application de la perturbation. Ceci permet d'utiliser des conditions initiales qui diffèrent légèrement de la valeur des variables à l'état stationnaire. Ensuite, à $t_{cutoff} = 6$ h, nous mettons à zéro l'entrée Tf_{sat} pour simuler la mise en carence.

L'étude du modèle s'attache donc à deux points : d'un côté l'existence et la stabilité d'un état stationnaire, de l'autre, la dynamique du modèle face à la mise en carence en fer.

3.2.2 Adimensionnement

L'adimensionnement est une méthode consistant à transformer des équations représentant un système physique en équations équivalentes ayant moins de variables et ne contenant plus que des quantités sans dimension [99]. Pour cela, à partir d'une équation de n quantités physiques $q_j, j \in [1, n]$ ayant des unités⁸, telle l'équation (3.18), un ensemble de k unités linéairement indépendantes est sélectionné. Ces unités, nommées unités fondamentales, sont notées $L_l, l \in [1, k]$. Afin de simplifier les notations, nous permutons l'ordre des arguments de la fonction f afin que $L_l = [q_l], l \in [1, k]$.

$$f(q_1, q_2, \dots, q_k, q_{k+1}, \dots, q_n) = 0 \quad (3.18)$$

Ensuite, la matrice D , de dimension $k \times n$ est construite, permettant de décrire l'ensemble des unités en fonction des unités fondamentales choisies, tel que décrit par l'équation (3.19).

$$D = \begin{pmatrix} D_{1,1} & \cdots & D_{1,n} \\ \vdots & D_{l,j} & \vdots \\ D_{k,1} & \cdots & D_{k,n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & D_{1,k+1} & \cdots & D_{1,n} \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & D_{k,k+1} & \cdots & D_{k,n} \end{pmatrix} \left| \begin{array}{l} j \in [1, n], \\ \prod_{l=1}^k L_l^{D_{l,j}} = [q_j] \end{array} \right. \quad (3.19)$$

Puis, nous créons $i = n - k$ nouvelles variables sans dimension nommées Π_1, \dots, Π_i selon l'équation (3.20). Intuitivement, ces nouvelles variables correspondent aux variables initiales $q_j, j \in [k + 1, n]$, normalisées selon les variables $q_l, l \in [1, k]$ de telle manière que les variables résultantes ($\Pi_j, j \in [1, i]$) n'aient plus d'unité. Pour un choix de k unités donné, cette normalisation est unique, les unités des variables $q_l, l \in [1, k]$ étant indépendantes. Par exemple, une variable qui, initialement, n'a pas d'unité, ne changera pas puisque la colonne correspondante dans la matrice D ne contiendra que des 0. Une variable qui, initialement, a la même unité que q_1 , sera divisée par q_1 , puisque la colonne correspondante dans la matrice D contiendra un 1 sur la première ligne et des 0 sur les autres lignes.

$$\forall j \in [k + 1, n], \Pi_{j-k} = q_j \cdot \prod_{l=1}^k q_l^{-D_{l,j}} \quad (3.20)$$

Nous obtenons alors une équation équivalente adimensionnée, décrite par l'équation (3.21).

$$\psi(\Pi_1, \Pi_2, \dots, \Pi_i) = 0, \quad [\Pi_j] = 1, \quad 1 \leq j \leq i \quad (3.21)$$

En choisissant judicieusement les unités fondamentales, il est possible de réduire le nombre de dimensions utilisées jusqu'à $\Delta = \text{rang}(D) + 1$ [100].

Cette méthode est applicable aux équations différentielles modélisant un système biologique, puisque, comme les équations représentant un système physique, elles ne contiennent généralement pas de fonctions telles que $\log(x)$ ou $\sin(x)$ où x n'est pas sans dimension. Dans notre modèle, les quantités physiques utilisées sont les paramètres et les variables de notre système d'équations. Les unités de ces paramètres et variables sont les cinq suivantes : « mol·L⁻¹ » (par exemple, les concentrations), « s⁻¹ » (par exemple, les constantes de vitesse des dégradations), « mol·(L·s)⁻¹ » (par exemple, les vitesses de production), « L·(mol·s)⁻¹ » (par exemple, les constantes de vitesse des réactions de deuxième ordre) et « s » (pour le temps). Le nombre d'unités fondamentales est donc $k = 2$ (par exemple, l'unité seconde et l'unité mole/litre), impliquant

8. Dans cette sous-section, la notation $[x]$ représente l'unité de la quantité physique x — à ne pas confondre avec la notation $[a,b]$ qui représente un intervalle fermé.

que le rang de la matrice D est égal à deux. L'adimensionnement de notre système d'équations permet donc de réduire le nombre de dimensions de notre modèle de 83 à 80. Néanmoins, cette réécriture nécessite de déterminer le bon choix d'unités fondamentales et complexifie l'interprétation du sens biologique des variables et paramètres. Cette piste n'est donc pas poursuivie.

3.2.3 Système lent-rapide

Plusieurs méthodes de réduction de modèle, classiquement utilisées en biologie systémique, permettent de procéder à des simplifications, tout en conservant les traits principaux de la dynamique. Parmi celles-ci, nous pouvons citer les méthodes d'analyse des systèmes lents-rapides [101], qui se basent sur la théorie des perturbations singulières. Cette théorie s'applique à des systèmes dans lesquels un petit paramètre ϵ multiplie certaines dérivées du système d'équations différentielles. Les variables correspondantes sont dites rapides. La théorie permet, sous certaines conditions, de réduire l'analyse au sous-système des variables lentes, les variables rapides s'ajustant très rapidement aux variations des variables lentes (après une phase transitoire).

Une difficulté dans l'application de la théorie des perturbations singulières est l'identification des sous-ensembles de variables lentes et rapides ainsi que du paramètre ou de la combinaison de paramètres qui va jouer le rôle du ϵ de la théorie. Dans le modèle présenté ici, la cinétique de nombreuses réactions n'est pas connue. Il est possible de créer deux classes de réactions, celle contenant les réactions rapides (réactions menant à la synthèse de l'hème⁹ et réactions d'insertion d'agrégat fer-soufre¹⁰), ainsi que celle contenant les réactions lentes (régulations traductionnelles par les IRP). Toutefois, il n'est pas possible de créer une classe intermédiaire qui contiendrait les réactions restantes, dont la vitesse est inconnue, puisqu'à cause des larges incertitudes sur les paramètres du modèle, ces réactions peuvent être aussi rapides que des réactions de la classe rapide ou aussi lentes que des réactions de la classe lente.

Cette méthode de réduction de modèle peut s'appliquer à des réseaux de taille importante mais nécessite de disposer d'informations suffisamment précises sur la vitesse des réactions pour les classer. Cette classification est d'autant plus difficile que certaines réactions considérées dans les modèles sont des regroupements de plusieurs réactions. Dans notre modèle, des constantes de vitesse peuvent n'être connues qu'avec une forte incertitude, empêchant l'utilisation de cette méthode.

3.2.4 Simplification

Comme indiqué dans la sous-section 3.2.1, notre scénario d'étude débute par une situation où le fer est en abondance, qui est modélisée par un état stationnaire. Il est donc nécessaire de donner des valeurs aux conditions initiales ainsi qu'aux paramètres, correspondant à un état stationnaire du système. Nous utilisons les équations algébriques issues des équations différentielles où les termes sigmoïdaux sont remplacés par 0 ou par 1 (simplifiant ainsi les équations de *ALAS2*, *FPN1a*, *Ft*, *apoACO2*, *IRP2*, *FBXL5*, *FeS* et de *Fe*) pour obtenir une expression analytique exprimant les conditions initiales en fonction des paramètres et ainsi réduire le nombre d'inconnues à déterminer. Cette approximation des termes sigmoïdaux est bonne si les concentrations sont suffisamment éloignées de leur seuil respectif, et si les sigmoïdes sont assez raides.

Pour cela, nous commençons par utiliser les cinq équations algébriques issues des équations différentielles de *ALAS2*, *FPN1a*, *Ft*, *IRP2* et *FBXL5* pour obtenir des expressions analytiques pour ces cinq variables, en fonction des paramètres. Ces expressions sont données dans les

9. Ces réactions sont : $cit \rightarrow isocit$, $isocit \rightarrow PPIX$, $PPIX + Fe \rightarrow heme$.

10. Ces réactions sont les suivantes : $FeS \rightarrow Fech$, $IRP1 + FeS \rightarrow ACO1$, $apoACO2 + FeS \rightarrow ACO2$.

équations (3.22) à (3.26).

$$ALAS2 = \frac{k_{ALAS2_prod}}{k_{ALAS2_deg}} \quad (3.22) \quad IRP2 = \frac{k_{IRP2_prod}}{k_{FBXL5 \rightarrow IRP2} + k_{IRP2_deg}} \quad (3.25)$$

$$FPN1a = \frac{k_{FPN1a_prod}}{k_{FPN1a_deg}} \quad (3.23) \quad FBXL5 = \frac{k_{FBXL5_prod}}{k_{FBXL5_deg}} \quad (3.26)$$

$$Ft = \frac{k_{Ft_prod}}{k_{Ft_deg}} \quad (3.24)$$

Nous recherchons ensuite une expression analytique décrivant les onze variables restantes en fonction des paramètres. L'idée générale que nous suivons consiste à exprimer l'ensemble des variables en fonction de paramètres et de la variable FeS , puis à obtenir et à résoudre une équation en FeS .

Nous commençons par considérer les équations algébriques issues des équations différentielles de $ACO1$ et $IRP1$ (resp. $ACO2$ et $apoACO2$) afin d'obtenir une expression analytique décrivant $ACO1$ et $IRP1$ (resp. $ACO2$ et $apoACO2$) en fonction de FeS et de paramètres. Les expressions ainsi obtenues sont données dans les équations (3.27) et (3.28) (resp. (3.29) et (3.30)).

$$IRP1 = \frac{k_{IRP1_prod} \cdot (k_{ACO1 \rightarrow IRP1} + k_{ACO1_deg})}{FeS \cdot k_{ACO1_deg} \cdot k_{IRP1+FeS \rightarrow ACO1} + (k_{ACO1 \rightarrow IRP1} + k_{ACO1_deg}) \cdot k_{IRP1_deg}} \quad (3.27)$$

$$ACO1 = \frac{FeS \cdot k_{IRP1+FeS \rightarrow ACO1} \cdot k_{IRP1_prod}}{FeS \cdot k_{ACO1_deg} \cdot k_{IRP1+FeS \rightarrow ACO1} + (k_{ACO1 \rightarrow IRP1} + k_{ACO1_deg}) \cdot k_{IRP1_deg}} \quad (3.28)$$

$$apoACO2 = \frac{k_{apoACO2_prod} \cdot (k_{ACO2 \rightarrow apoACO2} + k_{ACO2_deg})}{FeS \cdot k_{ACO2_deg} \cdot k_{apoACO2+FeS \rightarrow ACO2} + (k_{ACO2 \rightarrow apoACO2} + k_{ACO2_deg}) \cdot k_{apoACO2_deg}} \quad (3.29)$$

$$ACO2 = \frac{FeS \cdot k_{apoACO2+FeS \rightarrow ACO2} \cdot k_{apoACO2_prod}}{FeS \cdot k_{ACO2_deg} \cdot k_{apoACO2+FeS \rightarrow ACO2} + (k_{ACO2 \rightarrow apoACO2} + k_{ACO2_deg}) \cdot k_{apoACO2_deg}} \quad (3.30)$$

Puis nous utilisons l'équation algébrique issue de l'équation différentielle de la ferrocélatase pour obtenir une expression exprimant $Fech$ en fonction de paramètres et de FeS . Cette expression est donnée par l'équation (3.31).

$$Fech = \frac{FeS \cdot k_{FeS \rightarrow Fech} \cdot k_{Fech_prod}}{k_{Fech_deg}} \quad (3.31)$$

L'étape suivante consiste à remplacer, dans l'équation algébrique issue de l'équation différentielle de $TfR1$, les variables $IRP1$ et $IRP2$ par les expressions analytiques (3.27) et (3.25). Nous obtenons alors une expression décrivant $TfR1$ en fonction de FeS et de paramètres. Cette expression est donnée par l'équation (3.32). Dans un souci de lisibilité, le numérateur et le dénominateur de cette expression sont donnés séparément.

$$TfR1 = \frac{n_TfR1}{d_TfR1} \quad (3.32)$$

où

$$\begin{aligned}
n_TfR1 = & FeS \cdot k_{ACO1_deg} \cdot k_{IRP1+FeS \rightarrow ACO1} \\
& \cdot (k_{TfR1_prod} \cdot (k_{FBXL5 \rightarrow IRP2} + k_{IRP2_deg}) + k_{IRP2 \rightarrow TfR1} \cdot k_{IRP2_prod}) \\
& + (k_{ACO2 \rightarrow IRP1} + k_{ACO1_deg}) \\
& \cdot (k_{IRP1_deg} \cdot (k_{TfR1_prod} \cdot (k_{FBXL \rightarrow IRP2} + k_{IRP2_deg}) + k_{IRP2 \rightarrow TfR1} \cdot k_{IRP2_prod}) \\
& + k_{IRP1_prod} \cdot k_{IRP1 \rightarrow TfR1} \cdot (k_{FBXL5 \rightarrow IRP2} + k_{IRP2_deg}))
\end{aligned}$$

et

$$\begin{aligned}
d_TfR1 = & (FeS \cdot k_{ACO1_deg} \cdot k_{IRP1+FeS \rightarrow ACO1} + (k_{ACO1 \rightarrow IRP1} + k_{ACO1_deg}) \cdot k_{IRP1_deg}) \\
& \cdot (k_{FBXL5 \rightarrow IRP2} + k_{IRP2_deg}) \cdot k_{TfR1_deg}
\end{aligned}$$

Ensuite, nous remplaçons, dans l'équation algébrique issue de l'équation différentielle de FeS , les variables $IRP1$ et $apoACO2$ par les expressions analytiques données dans les équations (3.27) et (3.29) afin d'obtenir une expression analytique de Fe en fonction de la variable FeS et de paramètres. La forme de cette expression est donnée dans l'équation (3.33), où les $c_i, i \in [1, 6]$ représentent des combinaisons de paramètres. Pour exemple, l'expression factorisée de c_3 est donnée par l'équation (3.34).

$$Fe = \frac{c_1 \cdot FeS + c_2 \cdot FeS^2 + c_3 \cdot FeS^3}{c_4 + c_5 \cdot FeS + c_6 \cdot FeS^2} \quad (3.33)$$

$$c_3 = k_{ACO1_deg} \cdot k_{ACO2_deg} \cdot k_{IRP1+FeS \rightarrow ACO1} \cdot k_{apoACO2+FeS \rightarrow ACO2} \cdot (k_{FeS \rightarrow Fech} \cdot k_{Fech_prod} + k_{FeS_cons}) \quad (3.34)$$

En remplaçant les variables $TfR1$, $FPN1a$ et Ft par leurs expressions analytiques dans l'équation algébrique de Fe , nous obtenons une expression analytique exprimant $PPIX$ en fonction de FeS et de paramètres. La forme de cette expression est donnée dans l'équation (3.35) où les $c_i, i \in [7, 18]$ sont des sommes de produits de paramètres.

$$PPIX = \frac{c_7 + c_8 \cdot FeS + c_9 \cdot FeS^2 + c_{10} \cdot FeS^3 + c_{11} \cdot FeS^4 + c_{12} \cdot FeS^5}{c_{13} \cdot FeS + c_{14} \cdot FeS^2 + c_{15} \cdot FeS^3 + c_{16} \cdot FeS^4 + c_{17} \cdot FeS^5 + c_{18} \cdot FeS^6} \quad (3.35)$$

Dans l'avant-dernière étape, nous remplaçons les variables $ALAS2$, $Fech$ et $PPIX$ par leurs expressions analytiques, dans l'équation algébrique issue de l'équation différentielle de $PPIX$. Ceci nous permet d'obtenir une expression analytique décrivant $isocit$ en fonction de la variable FeS et de paramètres. La forme de cette expression est donnée dans l'équation (3.36) où les $c_i, i \in [19, 26]$ sont des sommes de produits de paramètres.

$$isocit = \frac{c_{19} + c_{20} \cdot FeS + c_{21} \cdot FeS^2 + c_{22} \cdot FeS^3}{c_{23} + c_{24} \cdot FeS + c_{25} \cdot FeS^2 + c_{26} \cdot FeS^3} \quad (3.36)$$

La dernière étape consiste à utiliser l'équation algébrique issue de l'équation différentielle de $isocit$ pour obtenir une expression analytique de FeS ne dépendant que de paramètres. Par conséquent, en remplaçant la variable FeS par cette expression dans les dix expressions algébriques (3.27) à (3.36), nous obtenons des expressions décrivant les variables en fonction de paramètres uniquement. Afin d'obtenir cette expression de FeS , nous remplaçons les variables $ACO1$, $ACO2$, $ALAS2$ et $isocit$ par leurs expressions analytiques dans l'équation d' $isocit$. Nous obtenons alors une équation de la forme $f(FeS)/g(FeS) = 0$ où le degré de $f(FeS)$ est de six.

Pour obtenir une expression analytique exprimant les conditions initiales en fonction des paramètres, il est alors nécessaire de résoudre le polynôme de degré six en fonction de FeS : $f(FeS) = 0$. Le degré de ce polynôme étant trop élevé pour pouvoir être résolu analytiquement, nous n'avons pu poursuivre cette piste. Toutefois, pour un modèle plus simple, il est envisageable de décrire les concentrations en fonction des paramètres, à partir du système d'équations algébriques obtenu à l'état stationnaire. Notons qu'ici il a été nécessaire de guider manuellement la résolution de ces équations pour obtenir ces expressions, la résolution automatique de ce système d'équations, en utilisant le solveur symbolique Sage [102], n'aboutissant pas à cause de la complexité du système.

Chapitre 4

Modèle spécifique de la régulation par les IRP

Ce chapitre présente un modèle centré sur le mécanisme de régulation du fer intracellulaire par les IRP. Ce modèle, obtenu par simplification du modèle préliminaire présenté dans le chapitre 3, est constitué de cinq acteurs et d'une entrée. Nous présentons ensuite la méthodologie semi-formelle créée pour l'étude des systèmes biologiques et nous l'appliquons à ce modèle. Un résultat majeur de cette étude est la mise en avant de l'importance du seuil de la régulation du fer sur les IRP.

Dans ce chapitre, nous décrivons, dans la section 4.2, l'intégralité de la méthodologie que nous avons développée, ainsi que le modèle, et son étude. Cette section est la recopie d'un article en préparation pour une soumission au journal BMC Systems Biology. Elle est introduite par la section 4.1 qui donne un aperçu de la méthodologie. À la suite de la recopie de l'article, nous présentons les différentes pistes qui ont été explorées et décrivons des détails méthodologiques supplémentaires dans la section 4.3. Ce chapitre se termine par une comparaison avec le modèle présenté par Chifman *et al.* [103], ainsi qu'avec le modèle présenté par Omholt *et al.* [96] et avec celui de Mitchell *et al.* [104] (section 4.4).

4.1 Introduction

4.1.1 Définition du modèle

Le modèle préliminaire, présenté dans le chapitre 3, comporte 16 variables et 67 paramètres, rendant son étude complexe. Face à cette complexité, nous créons un modèle simplifié, qui se concentre sur le mécanisme de régulation du fer intracellulaire face à une situation de carence. Pour cela, nous procédons à trois changements principaux :

- nous supprimons la partie métabolique du réseau préliminaire, c'est-à-dire l'utilisation des agrégats fer-soufre pour synthétiser les deux aconitases ACO1 et ACO2, ainsi que la voie de synthèse de l'hème. Cette simplification nous paraît raisonnable étant donné que la synthèse d'hème est faible chez les cellules K562. En effet, cette lignée cellulaire, que nous modélisons, est peu différenciée. La synthèse des agrégats fer-soufre utilisant une quantité significative de fer et cette consommation de fer n'étant plus explicitée, nous modifions la signification biologique du terme de consommation de fer pour les besoins internes de la cellule et considérons qu'il englobe également la consommation de fer pour la synthèse de ces agrégats ;

- les mécanismes de régulation par les deux IRP étant similaires, ainsi que leurs affinités pour les IRE [51], nous regroupons ces deux protéines. Nous simplifions, par conséquent la régulation des IRP par le fer : nous supprimons la protéine FBXL5 et, la cinétique de la régulation des IRP n'étant pas connue, nous considérons que le fer régule IRP selon une cinétique sigmoïdale ;
- nous ne considérons plus explicitement l'export non régulé de fer par l'exporteur FPN1b, mais nous regroupons dans un même terme cet export non régulé et la consommation interne de fer.

Du fait de ces changements, nous ne considérons plus la régulation par les IRP de la consommation du fer pour les besoins de la cellule (la consommation de fer pour la synthèse d'hème chez les précurseurs des globules est régulée par les IRP).

Ce nouveau modèle, comporte cinq variables qui sont Ft , Fe , $TfR1$, $FPN1a$ et IRP , ainsi que 24 paramètres et l'entrée Tf_{sat} . Les quatre premières variables et l'entrée ont la même signification biologique que les variables portant le même nom dans le modèle préliminaire. La variable IRP représente la somme des variables $IRP1$ et $IRP2$ du modèle préliminaire.

4.1.2 Méthodologie d'étude

La section 4.2 présente la méthodologie que nous avons créée, ainsi que l'étude du modèle, utilisant le scénario d'étude décrit dans la sous-section 3.2.1. Cette section est la recopie d'un article en préparation. Cet article commence par une rapide présentation du système biologique considéré, puis introduit le système d'équations différentielles le modélisant. Les 24 paramètres et conditions initiales de ces cinq équations différentielles, formant l'ensemble des inconnues du modèle, sont également décrits. Ensuite, les différentes parties de la méthodologie, visant à obtenir l'ensemble des *jeux de paramètres valides*, sont présentées. Un jeu de paramètre est défini comme une valeur pour chaque paramètre et condition initiale du modèle. Nous disons qu'un jeu de paramètres est valide s'il satisfait toutes les contraintes déduites des données disponibles sur le système biologique.

La première partie de cette méthodologie est la traduction de l'ensemble des données disponibles en 1) intervalles de valeurs possibles pour les paramètres et conditions initiales, 2) inégalités sur les paramètres et/ou conditions initiales ou 3) formules STL décrivant les données comportementales. Nous pouvons alors définir *l'espace de recherche* comme étant un hyperrectangle, de grande dimension, où chaque dimension représente un paramètre ou une condition initiale. Chaque point de cet espace est un jeu de paramètre.

La deuxième partie de cette méthodologie consiste à étudier l'état stationnaire, d'une part, en vérifiant sa stabilité et, d'autre part, en utilisant un solveur sur intervalles pour réduire les intervalles des inconnues. De plus, les équations algébriques, définissant le ou les états stationnaires, sont simplifiées selon les données disponibles sur l'état stationnaire. Ces équations algébriques sont alors utilisées pour calculer les conditions initiales en fonction des paramètres, de manière approchée. Ceci permet de réduire la dimension de l'espace de recherche et de s'assurer que les simulations atteignent l'état stationnaire avant la mise en carence. L'espace exploré est alors nommé l'espace des paramètres et les points de cet espace sont appelés des jeux de paramètres.

La troisième partie présente la recherche d'une *région valide*, c'est à dire une région de l'espace de recherche ne contenant que des jeux de paramètres valides. Cela consiste à rechercher un premier jeu de paramètres valide en utilisant l'algorithme d'optimisation développé par Nelder-Mead. La fonction objectif utilisée pour évaluer un jeu de paramètres est l'évaluation quantitative des formules STL sur la simulation correspondante. À partir de ce jeu de paramètres valide, un algorithme type Metropolis-Monte-Carlo est utilisé, fournissant une région valide grossière, puis

un algorithme d'expansion d'ellipsoïde est utilisé pour raffiner cette région. Pour rappel, travailler avec une région valide plutôt qu'un unique point valide permet d'éviter ce choix arbitraire d'un point, dont l'unicité n'est soutenue par aucune donnée. La région valide, sa forme et son volume, traduit directement le contenu informationnel des données. En d'autres termes, elle traduit un état des connaissances. En conséquence les propriétés qui sont vraies pour tout un ensemble d'instanciations valides correspondent à ce qui peut être réellement déduit des données, sans ajouter d'hypothèses incontrôlées (choix d'un point représentatif unique).

La dernière partie de cette méthodologie consiste à décrire la forme et le volume de la région valide trouvée ; à faire des analyses statistiques sur un échantillonnage uniforme de cette région, telles le calcul de corrélations entre les inconnues du modèle ; et à décrire la diversité des simulations obtenues. Cette analyse des résultats permet alors d'apporter de nouvelles connaissances sur le système biologique et de guider les expérimentations à venir. Pour le système de l'homéostasie du fer, ces analyses ont permis de découvrir la quantité minimale de fer pour qu'une cellule survive.

Les deux difficultés majeures de ce travail sont donc 1) l'exploration d'un espace de grande dimension à la recherche des points de cet espace respectant des propriétés données et 2) la représentation d'un ensemble géométrique potentiellement non convexe, potentiellement non connexe, en haute dimension.

4.1.3 Implémentation

Cette méthodologie est implémentée par un script écrit en Python ainsi que par un ensemble de fonctions Matlab qui ont été intégrées dans le logiciel Breach. Le script Python facilite l'étude du modèle à l'état stationnaire puisqu'à partir d'un fichier d'entrée écrit dans un format spécifique, il calcule les équations algébriques à l'état stationnaire, réduit les intervalles grâce au solveur sur intervalle, recherche l'existence de contradictions à l'état stationnaire et indique quelles sont les contraintes à supprimer pour lever la contradiction. De plus, il génère automatiquement un ensemble de fichier Matlab pour la recherche d'un jeu de paramètres valide. Cette implémentation permet ainsi d'automatiser l'étude de modèles.

4.2 Présentation des travaux

Semi-Formal Methodology for Parameter Space Exploration : Application to the Iron Homeostasis Network

Nicolas Mobilia^{*5,6}, Alexandre Donzé³, Emmanuel Pourcelot^{4,5,6,7}, Pascal Mossuz^{4,5,6}, Jean-Marc Moulis^{1,2,5,7}, Éric Fanchon^{*5,6}

¹Univ. Grenoble-Alpes, Environmental and Systems Biology (BEeSy), Grenoble France

²Laboratoire de Bioénergétique Fondamentale et Appliquée (LBFA), Inserm U1055

³EECS Department, University of California Berkeley, Berkeley, CA 94720 USA

⁴Laboratoire d'Hématologie, Institut de Biologie et Pathologie, CHU

⁵Université Joseph Fourier-Grenoble I, F-38041 Grenoble, France

⁶CNRS, TIMC-IMAG, UMR 5525, F-38041 Grenoble, France

⁷CEA, iRTSV, F-38054 Grenoble, France

Email: Nicolas Mobilia* - nicolas.mobilia@imag.fr; Alexandre Donzé - donze@eecs.berkeley.edu; Emmanuel Pourcelot - epourcelot@chu-grenoble.fr; Pascal Mossuz - pmossuz@chu-grenoble.fr; Jean-Marc Moulis - jean-marc.moulis@cea.fr; Éric Fanchon* - eric.fanchon@imag.fr;

*Corresponding author

Abstract

Background: Ordinary differential equations (ODEs) are widely used to model biological processes at all scales, in particular at the sub-cellular scale to model reaction networks. These models contain many parameters which need to be given a value in order to perform simulations and analyze the dynamical properties. This classical approach encounters difficulties in Systems Biology because the experimental data needed to build dynamical models are presently too scarce and heterogeneous. The parameter values are either unknown or uncertain when some data exist. The current practice is to search the best qualitative solution, that is to search for parameter values corresponding to a simulated behavior that best satisfy observations, and to analyze the dynamics of this 'best solution'. This can be misleading as many other solutions may exist which are nearly as good with respect to agreement with observed behavior. We propose an approach to tackle this problem, in which temporal logic formulas are used to represent behavioral properties and exploited to identify a region of parameter space (instead of a single point).

Results: We present a new approach for the modeling of biological networks with uncertain parameters. It is based on recent tools analyzing the satisfaction of properties of (continuous or hybrid) dynamical systems. The knowledge, or the assumptions, related to the behavior of the system are first formalized in the Signal Temporal Logic (STL) using the tool Breach. An algorithm is dedicated to the search of a parameter point satisfying the STL formula encompassing all the knowledge on the system. From this initial point, we then try to identify the largest possible region made of parameter points satisfying the same STL formula. This expansion is performed

with the algorithm of Zamora-Sillero et al. (2011). In the last step, we analyze the set of possible behaviors compatible with the STL constraints. We apply this methodology in an effort to model and better understand the complex network of cellular iron homeostasis in mammals with focus on progenitors of erythroid cells. The problem of identifying parameters is largely under-determined in the current state of knowledge, and consequently the valid domain is quite large. We show that it is nevertheless possible to extract some biological insight.

Conclusions: The data required to build dynamical models are scarce. We describe an approach to exploit different kinds of biological data and assumptions, combining automatic interval reduction, simulation, and temporal logic formula verification. Using this approach we present a new model of iron homeostasis in mammalian cells. The analysis of this model provides insight on the set of behaviors which are compatible with the observational constraints.

Modeling, Differential equations, Interval solver, STL formulas, Iron homeostasis

1 Background

1.1 Introduction

Ordinary differential equations (ODE), and more recently hybrid systems [105] (combining finite states automata with ODE), are widely used in quantitative modeling approaches in cellular biology. These models involve many parameters, the numerical values of which are required to perform simulations and analyze the behavior. When a direct measurement of a parameter exists, the level of uncertainty can be high. It is well known, for instance, that values of kinetic constants obtained from *in vitro* measurements on purified enzymes can hardly be used to describe the behavior *in vivo*, due for example to interactions with other cell components (depending on cellular state), to the possibility of preferential localization or sequestration, or to anomalous diffusion. Two sets of measurements performed on identical cell types placed in supposedly identical conditions can also be qualitatively different for multiple reasons (undetected heterogeneity of cell populations, different batches of antibodies used for detection, for instance). But the biggest problem to date is the paucity of data needed to build dynamical models. In the present state of biological knowledge a large part of the needed values are often unavailable (not measured and not computable from first principles). To improve the situation the modeler is thus led to use data from other cell types, other organisms or other experimental conditions. In other words, models in biology are generally build using very heterogeneous data (in terms of conditions, cell types, organisms) at the cost of increased uncertainty on parameter values. But even with such an inclusive strategy the amount of data may remain inadequate to produce a fully instantiated model. This means that a set of possible models should be considered instead.

Our approach rests on the following points: experimental data on parameters are represented by intervals (inequalities) instead of a numerical value. When the parameter values are highly uncertain due to heterogeneity a large margin is assumed on each side of the measured value in order to account for the

uncertainty. If no measurement is available at all, interval bounds are chosen to define a physiologically reasonable domain. Known and unknown parameters are thus represented formally in the same way, by intervals, but the interval associated to an unmeasured parameter is several orders of magnitude larger. A second aspect of the approach is that to integrate constraints representing knowledge, or hypotheses, on the behavior of the system. This information can be quantitative or qualitative. It is formalized using a temporal logic formalism. In fact, temporal logic is used as a language to express all available information: parameter inequalities, other relationships involving model parameters, and behavioral knowledge. Once formalized, this information can be exploited in an automatic manner to select models which satisfy all these constraints. A third aspect is that we do not look for a single optimal parameter instantiation. We try instead to identify the region of parameter space corresponding to traces which satisfy the constraints. All the associated models satisfy the constraints and consequently the whole set should be considered in any interpretation of the data. Clearly, the larger the volume of the valid parameter region, the weaker the interpretation. But it is misleading to base the reasoning on a single supposedly representative model, as is often done. The volume and shape of the valid parameter region, and associated behaviors, are in themselves a form of information which can be used to identify different classes of behaviors, and to select informative experiments, in the sense that they will allow the largest volume reduction.

These principles are combined in the present paper using a *semi-formal* approach. On the one hand, we use methods and tools from the domain of formal methods, such as an interval solver [106] to reduce the domain of valid parameters, and Signal Temporal Logic (STL) [38] to formalize the available experimental knowledge or hypotheses. On the other hand, we use extensive simulation and heuristic-based optimization to explore the parameter space - which, if it cannot prove formally the validity of the properties for a whole region, provides instead a scalable, flexible and incremental approach to tackle the exploration of large search space in the search of models which are consistent with the known data.

This methodology is applied to cellular iron homeostasis in animal cells, more precisely in erythroid progenitors, from which red blood cells are produced. Iron is mandatory and abundant in mammals. Most is used for hemoglobin synthesized by erythroid cells and it is channeled to the active site of iron-proteins and enzymes by a series of molecules protecting cells from the deleterious reactions iron ions may catalyze [107, 108]. Organismic and cellular iron use is highly regulated to optimize supply and avoid losses [109].

A simple model is presented, containing 19 parameters. This case is typical of the situation described above, namely the need to incorporate heterogeneous data, and the fact that no data are available at all for some of the parameters. This results in a high-dimensional search space.

The biological network we consider for iron homeostasis is presented in Section 1.2. Then, Section 2 describes the method used to model the biological system and to obtain the results. This section splits into the presentation of the mathematical model (Section 2.1), the formalization of the data and dynamical properties gathered in the literature (Sections 2.2 and 2.3) and the steps of our methodology for finding a valid region in parameter space (Sections 2.4, 2.5 and 2.6). The results obtained from the model are

exhibited in Section 3. Finally, a conclusion is given (Section 4).

1.2 Biological network

We are considering regulation of iron handling in precursors of erythroid cells, and, more specifically, we focus on the human immortalized cell line K562. Cellular iron input from the circulation occurs by endocytosis of the transferrin-transferrin receptor complex, and export is via ferroportin. Cellular iron sensing involves the mRNA-binding Iron Regulatory Proteins, the main targets of which define the molecular nodes of the considered network.

1.2.1 Molecular species and processes

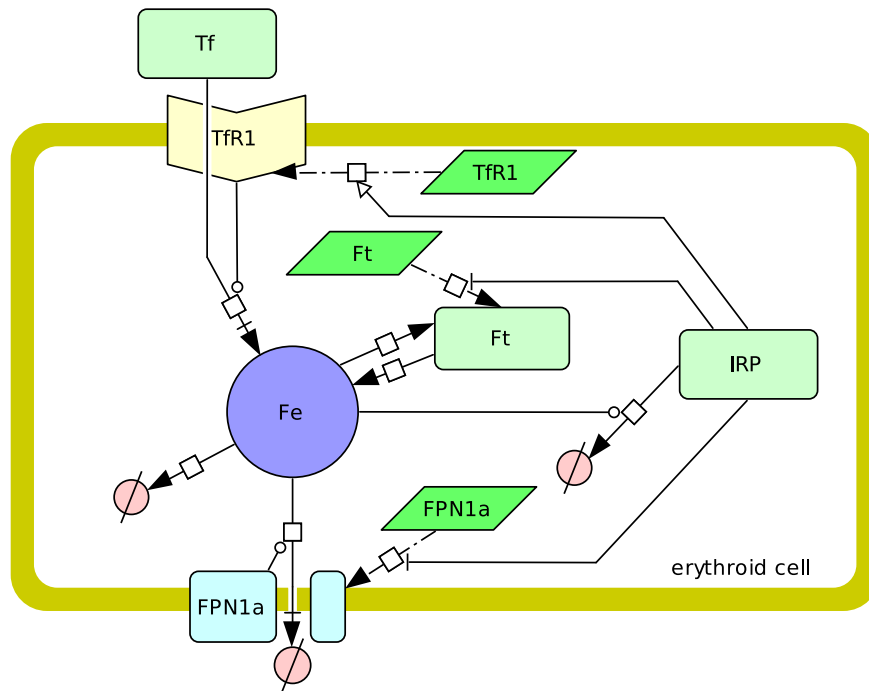


Figure 1: The acronyms and abbreviations used in this graph are explained in Section 1.2.1. The dashed arrows represent translation of mRNA into proteins. The arrow ending in an empty white head represents translational enhancement, while the lines with a broken ending arrowhead represent inhibition of translation. The lines ending with a combined perpendicular stroke and arrow represent iron transport through membranes. The arrow leading to an empty set symbol indicates either degradation (of IRP) or use (consumption and export of iron). Finally, the two regular arrows represent the loading/unloading of iron into/from the ferritins (Ft). The rounded rectangles represent proteins, the parallelograms represent mRNA, and the circle labeled Fe represents the pool of available iron. The concave hexagon represents the transferrin receptor (TfR1). This diagram was drawn with the software CellDesigner [55].

The main components of the cellular iron regulation network are displayed in Figure 1. They include five species: Iron Regulatory Protein (IRP , no difference is made between the two known proteins displaying this function), iron (Fe), ferritin (Ft), the transferrin receptor ($TfR1$) and ferroportin ($FPN1a$). Transferrin

(*Tf*) is a protein which carries iron in the blood stream, and which can bind one or two iron atoms.

The IRPs bind to a specific hairpin-like structures on mRNA, called the Iron Responsive Elements (IRE). Depending on the position of the IRE on the non-coding sequence of mRNA, two main consequences result:

- If the IRE is located in the 5' untranslated region (UTR) of the mRNA, binding of the IRP prevents the translation machinery from initiating mRNA translation. Thus, IRP represses the synthesis of the corresponding protein (negative effect on the target);
- If the IRE is located in the 3'-UTR of the mRNA, IRP binding inhibits mRNA degradation and consequently enhances the amount of translated protein (positive effect on the target).

Two known proteins, *IRP1*, and *IRP2*, display the mRNA-binding property. The concentration of both *IRP1* and *IRP2* responds to iron level within the cell, but different mechanisms are involved. *IRP2* is regulated by FBXL5, an iron sensor which promotes *IRP2* ubiquitination which in turn leads to *IRP2* degradation [59]. *IRP1* is a bifunctional protein. In the absence of iron it has the mRNA-binding property, but when iron level is high enough, iron promotes the assembly of an iron-sulfur cluster and the protein acquires an enzymatic function, loosing its mRNA-binding activity [49]. In the model we chose to represent *IRP1* and *IRP2* by a single entity called *IRP* which can be active or inactive. Only the active form has a regulatory effect. The underlying assumption is that the kinetics of *IRP2* degradation in the presence of iron is similar to the kinetics of iron-sulfur cluster assembly on *IRP1*. In Figure 1, the node *IRP* represents the active form.

In the model the iron species (*Fe*) represents the cellular iron which is available for cellular processes. The actual (bio)chemical nature of this iron in biological cells is still a topic of discussion [110]. A virtual species, which we call "*Fe*", is introduced in order to represent the readily available iron in the cell. If this amount of intracellular iron becomes too low, the cell cannot maintain vital functions and dies.

Ferritin (*Ft*) is a protein which stores iron [111]. It is a 24-subunit assembly with a large cavity. Thousands of iron ions can be internalized in a given *Ft* protein. Not all ferritins in a cell contain the same amount of iron [112]. For the sake of simplicity in the model the *Ft* symbol refers to ferritin loaded with a fixed amount n_{Ft} of iron ions. The complex mechanisms of iron loading into, and release from, ferritin are taken into account here in a simplified way. The mRNA of both ferritin subunits contain an IRE on their 5'-UTR regions, and the IRP consequently inhibits their translation.

The transferrin receptor (*TfR1*) is located at the surface of the cell, including erythroid ones [113]. Transferrin binding to its receptor induces endocytosis and leads to cellular iron uptake. The transferrin receptor mRNA contains five IRE in the 3'-UTR region, and is stabilized upon IRP binding. The density of transferrin receptor at the cell surface is thus correlated with the IRP activity.

Finally, ferroportin (*FPN1a*) is an iron exporter located at the surface of some cells [114]. The form considered in this work (*FPN1a*) contains an IRE in its 5'-UTR mRNA region, thus its translation is

inhibited by IRP.

The structure of this network corresponds to the core of the IRP regulated system with iron input and output present in many cell types. Initial intervals for the parameters were derived as much as possible from experimental data obtained with erythroid progenitors and corresponding cellular models (but some data from other cell types were also used).

1.2.2 Behavior

A qualitative understanding of the system has been obtained through a large body of biological experiments and we selectively summarize them here.

If the amount of iron is sufficient (*iron-replete* situation), the system is in a stationary state. Assuming that the iron provision stays constant, the cellular concentrations of species described in Section 1.2.1 do not vary. In this state, the IRP activity is expected to be relatively low. In the case of excessive iron amounts (a situation which is not explicitly studied here), the percentage of transferrin binding two iron atoms increases, the IRP activity reaches a minimum and inhibits translation of the iron importer (TfR1) to minimize cellular iron input.

From the iron-replete stationary state, if the cells become *iron-depleted*, the inhibition of IRP activity by iron decreases. This leads to increased transferrin receptor concentration and decreased ferroportin concentration, with the consequence of increasing iron uptake and retention. Due to a higher ferritin degradation rate, the iron stored in ferritin is, at least partially, released and becomes available for biosynthetic purposes, hence delaying iron starvation. The active IRP level is high as the iron level remains low, and the RNA-binding activity fades away upon iron replenishment.

2 Methods

In this section, we first present the system of differential equations which models the biological system. The two subsections 2.2 and 2.3 show the translation of available biological data into intervals in parameter space and temporal formulas. Then we characterize the steady-state of the system, corresponding to the iron-replete situation. We finally describe the algorithm used to obtain a valid parameter set and then a valid region.

2.1 ODE modeling

In this section we describe the differential equations that we derived to specifically account for the evolution of the concentration of each species over time. Some processes in the iron homeostasis network are in fact a composition of many reactions, and the choice of an analytical expression to represent them is not straightforward. This is the case of the iron entry mediated by TfR1, which involves endocytosis, and of the loading of iron in ferritin, for example.

2.1.1 Ferritin equation

To model the temporal evolution of ferritin, we assume that each ferritin protein is rapidly filled with iron atoms. In other words the amount of empty or partially filled ferritin is negligible. The second assumption is that iron stored within ferritins is only released when the ferritin is degraded. The variable Ft thus represents the concentration of filled ferritin proteins within the cell. Salgado et al. [112] have produced a very detailed model in which every molecular species (e.g. ferritin with all intermediate iron loads) is represented. This leads to a model with 5,003 differential equations which they reduced to 103 equations by considering that loading occurs by packages of 50 iron ions, instead of one iron at a time. This is still much too detailed for our purpose and we chose to use the following phenomenological equation for the dynamics of ferritin:

$$\frac{d Ft}{d t} = k_{Ft_prod} - k_{IRP \rightarrow Ft} \cdot \text{sig}^+(IRP, \theta_{IRP \rightarrow Ft}) - k_{Ft_deg} \cdot Ft \quad (1)$$

We consider a basal production rate described by k_{Ft_prod} . The second term describes the regulation by IRP. We use a sigmoidal regulation here, but we could consider other dynamics. If the concentration of IRP is above the threshold $\theta_{IRP \rightarrow Ft}$, the production rate is lowered by $k_{IRP \rightarrow Ft}$. It follows that $k_{IRP \rightarrow Ft}$ has to be lower or equal to k_{Ft_prod} since a production rate is necessarily positive. Finally, the third term ($k_{Ft_deg} \cdot Ft$) describes the spontaneous degradation of ferritin.

The sigmoid function is defined such that: $\text{sig}^+(x, \theta) = \frac{x^n}{x^n + \theta^n}$, where x is a concentration variable, θ is the threshold and n defines the steepness of the sigmoid. In our model, we use $n = 20$, which corresponds to a very steep sigmoid.

2.1.2 Iron equation

In the cell, when the level of iron becomes low, some iron-containing proteins release their iron content, by protein degradation for instance, which contributes to refill the iron pool; however, only iron release from ferritin is considered in our model without affecting the general validity of the reasoning because iron release by other protein degradation is small regarding to the quantity released by ferritin. The equation defining the dynamics of the iron concentration is:

$$\frac{d Fe}{d t} = k_{Fe_input} \cdot TfR1 \cdot Tf_{sat} - n_{Ft} \cdot \frac{d Ft}{d t} - k_{Fe_export} \cdot Fe \cdot FPN1a - k_{Fe_cons} \cdot Fe \quad (2)$$

The import of iron into the cell is proportional to the amount of iron-bound transferrin internalized with the receptor. This is represented by the product of the receptor concentration and the fraction of iron in transferrin, i.e. the transferrin saturation Tf_{sat} . The second term describes the storage or release of iron due to the synthesis or degradation of ferritin, respectively. This variation is equal to the number of iron atoms per ferritin molecule (aka n_{Ft}) times the variation in the ferritin concentration (aka $\frac{d Ft}{d t}$). Here, we consider that (i) iron release occurs only when ferritin is degraded (a process of loss from intact ferritin has been reported *in vitro* but there is no evidence that this process is significant *in vivo*); (ii) synthesized ferritins are very quickly filled with iron so that only ferritin loaded with iron exist in significant amount. The third term describes the export of iron out of the cell as proportional to the iron and FPN1a concentration. The export parameter is k_{Fe_export} . Finally, within the cell, iron is used for many purposes;

the global consumption of iron is represented by the last term in our equation ($k_{Fe_cons} \cdot Fe$). For the sake of simplicity, we consider this consumption term (which represents a whole set of cellular processes) to be proportional to the iron concentration. The export of iron by FPN1 forms that are not regulated by IRP is included in this term.

2.1.3 IRP equation

In this section, we develop the equation describing the temporal evolution of active IRP concentration. The term “IRP concentration” will designate the concentration of active IRP only. The equation describing the IRP concentration is:

$$\frac{d IRP}{d t} = k_{IRP_prod} - k_{Fe \rightarrow IRP} \cdot \text{sig}^+(Fe, \theta_{Fe \rightarrow IRP}) \cdot IRP - k_{IRP_deg} \cdot IRP \quad (3)$$

We define a constant rate of production of *IRP* by k_{IRP_prod} , thus neglecting transcriptional regulation. As explained above the concentration of *IRP* depends on iron level. The degradation of IRP is described by an iron-independent term ($k_{IRP_deg} \cdot IRP$) and an iron-dependent term ($k_{Fe \rightarrow IRP} \cdot \text{sig}^+(Fe, \theta_{Fe \rightarrow IRP}) \cdot IRP$). In this way, if the iron level is significantly below the threshold $\theta_{Fe \rightarrow IRP}$, the degradation rate is $k_{IRP_deg} \cdot IRP$, whereas if the iron concentration is significantly above this threshold, the degradation rate is $(k_{Fe \rightarrow IRP} + k_{IRP_deg}) \cdot IRP$, where $k_{Fe \rightarrow IRP}$ is the parameter describing the enhanced degradation of *IRP* in presence of iron.

2.1.4 Ferroportin equation

The form of the equation describing the ferroportin concentration is similar to that of ferritin. This equation is:

$$\frac{d FPN1a}{d t} = k_{FPN1a_prod} - k_{IRP \rightarrow FPN1a} \cdot \text{sig}^+(IRP, \theta_{IRP \rightarrow FPN1a}) - k_{FPN1a_deg} \cdot FPN1a \quad (4)$$

The first term describes the basal production of FPN1a. The second term expresses the regulation of IRP on the translation of FPN1a. With the same reasoning as previously, the regulation parameter $k_{IRP \rightarrow FPN1a}$ has to be lower or equal to the production parameter k_{FPN1a_prod} . The dynamics of IRP regulation on Ft and FPN1a are considered to be similar because both FPN1a and Ft mRNA have a single IRE located in their 5'-UTR region. The sequences of these IRE are slightly different but there is no experimental evidence of major differences between the two. The third term describes the spontaneous degradation of ferroportin.

2.1.5 Transferrin receptor equation

The concentration of the transferrin receptor evolves according to the following equation:

$$\frac{d TfR1}{d t} = k_{TfR1_prod} + k_{IRP \rightarrow TfR1} \cdot IRP - k_{TfR1_deg} \cdot TfR1 \quad (5)$$

The production of TfR1 includes a basal rate k_{TfR1_prod} , and it is considered to increase proportionally to the IRP concentration by the term $k_{IRP \rightarrow TfR1} \cdot IRP$. This is a rough approximation [63]: its analytical translation is different from those describing the regulation of ferritin and ferroportin in an effort to consider the binding of several IRP to a single TfR mRNA molecule (since it contains several IREs). The last term of the equation represents the spontaneous degradation.

2.1.6 Transferrin saturation

The transferrin saturation, which represent the percentage of transferrin sites loaded with iron, is considered to be an externally controlled entity. It starts at the value $Tf_{sat,0}$ and stays at this value until a perturbation is applied, namely the addition of an iron chelator to the culture medium. As a consequence, the initial condition $Tf_{sat}(0)$ is equal to $Tf_{sat,0}$. The perturbation occurs at a time $t = t_{cutoff}$. The equation modeling the evolution of Tf_{sat} upon chelation is:

$$\frac{d Tf_{sat}}{d t} = 10000 \cdot Tf_{sat,0} \cdot \text{sig}^-(t, t_{cutoff}) - 10000 \cdot Tf_{sat} \quad (6)$$

The coefficient 10000 in the two terms allows the parameter Tf_{sat} to reflect the quick change of iron availability.

2.1.7 Interaction graph

The interaction graph is a useful tool to visualize the pattern of interactions in a system of differential equations. It is defined from the Jacobian matrix of the system. Each non-zero element in this matrix indicates that a species influences another species. The nodes of the interaction graph are the dynamical variables of the system (concentration of biological species), and each arc represents the influence of one node on another, i.e. a non-zero element of the Jacobian matrix. The sign of a partial derivative determines whether the corresponding influence is positive or negative (each arc is labeled with a sign). Of course, it is possible that the signs of the elements depend on the point of phase space at which the Jacobian is evaluated, but in our model the signs are all constant. The interaction graph built in this manner is shown in Figure 2.

In graph theory, a circuit in a graph is a succession of nodes, starting and ending at the same node such that, for each pair of consecutive nodes of the circuit, there exists an arrow directed from the first to the second node of the pair. We define a negative (resp. positive) circuit as a succession of arrows in the interaction graph such that the product of the signs labeling these arrows is negative (resp. positive). The graph associated to the current model contains three negative circuits of length greater than one ($\text{IRP} \rightarrow \text{Fe} \rightarrow \text{IRP}$; $\text{IRP} \rightarrow \text{FPN1a} \rightarrow \text{Fe} \rightarrow \text{IRP}$ and $\text{IRP} \rightarrow \text{TfR1} \rightarrow \text{Fe} \rightarrow \text{IRP}$), and five negative loops corresponding to spontaneous degradation or consumption. It also contains one positive circuit ($\text{IRP} \rightarrow \text{Ft} \rightarrow \text{Fe} \rightarrow \text{IRP}$). The presence of these circuits, together with the presence of non-linear terms in the equations, potentially generates a complex dynamical behavior: the existence of the positive circuit can lead to multiple stable conditions, whereas the negative circuits can lead to oscillations [115].

2.2 Definition of the parameter search space

The parameters of the equations and the concentrations at $t = 0$ (initial conditions) are both considered as unknowns of the system. Consequently, the parameter (search) space in the following should be understood as the space constituted by parameters values and species concentration under initial conditions.

The parameter space of the system is thus twenty-four dimensional: 19 model parameters and 5 initial conditions. To define an initial parameter search space, we combined published data from different sources

as shown in Table 1 (when available, we use data obtained with erythroleukemic cells). As a reminder, these cells proliferate when iron is available, i.e., in the iron-replete conditions, a situation which can be modeled as a steady state. This iron-replete steady state is taken as the initial state to which a perturbation is applied. We then obtained intervals or values for eleven unknowns, including seven intervals for parameters ; the value of the input initial condition ; and intervals for three other initial conditions. Also, an inequality relating k_{Fe_cons} , k_{Fe_export} and $FPN1a$ initial condition was derived on the basis of a physiologically sound assumption.

In order to define the search space for the fourteen remaining unknowns, we reasoned by similarity. For example, we had information on the production rate of the transferrin receptor (Table 1) but none on the basal production rate of ferritin, ferroportin and IRP. We then assumed that the basal production rate of these three proteins is neither more than 1 000 times higher than the highest possible transferrin receptor basal production rate, nor 1 000 times lower than the lowest. The interval describing the possible values for the parameters k_{Ft_prod} , k_{FPN1a_prod} and k_{IRP_prod} were thus defined as $[1.0 \times 10^{-18}, 1.0 \times 10^{-10}] \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$. Moreover, we restrict intervals to meaningful values. Typically, a degradation rate lower than $1.0 \times 10^{-8} \text{ s}^{-1}$ means that the half-life is longer than one year. As a consequence, this value is taken as the lower limit for the parameters k_{Ft_deg} and k_{FPN1a_deg} . We also set the lower value for the storage capacity of ferritin (n_{Ft}) to 50 to reflect the genuine storage capacity of ferritin. Concerning the threshold parameters, we simply assumed that their intervals are the same as those of the corresponding protein concentrations. Furthermore, the lower value for iron concentration at steady state is set to $1.0 \times 10^{-10} \text{ mol}\cdot\text{L}^{-1}$. Finally, the intervals for Ft and $FPN1a$ initial conditions are set to $[1.0 \times 10^{-13}, 1.0 \times 10^{-5}] \text{ mol}\cdot\text{L}^{-1}$ since $1.0 \times 10^{-13} \text{ mol}\cdot\text{L}^{-1}$ represents less than one molecule per cell and $1.0 \times 10^{-5} \text{ mol}\cdot\text{L}^{-1}$ is a relatively high intracellular concentration for a soluble protein.

2.3 Formal expression of behavioral observations

Since the information extracted from the literature, in the form of inequalities as explained in the previous subsection, is very partial, it is essential to be able to use the knowledge (or hypotheses) on the behavior of the system. To express formally this type of knowledge we use the formalism of Signal Temporal Logic (STL). This section shows how the expression power of STL eases the translation of behavioral data and hypothesis into temporal formulas. We first describe STL syntax, and then formally define the expected behavior for the two modes described in Section 1.2.2: the *iron-replete* mode and the *iron-depleted* mode into STL formulas.

2.3.1 Signal Temporal Logic

For our analysis, we used Breach [39,47], a tool based on Matlab/C++. This tool allows users not only to simulate differential systems, but also to express temporal logic formulas in Signal Temporal Logic (STL) and to check whether a trajectory satisfies a given STL formula. Here we briefly introduce the parts of STL used in this work. A more complete description can be found in [39]. Initially created for verifying the correctness of reactive programs, temporal logics provide languages to express properties over time-dependent phenomena for any kind of dynamic system. STL and its quantitative semantics [116]

specialize into systems with continuous time and with real-valued states. An STL formula φ is written according to the following grammar, where φ_1 , φ_2 and φ_0 are STL formulas:

$$\varphi ::= \mu \mid (\varphi_1) \text{ and } (\varphi_2) \mid \text{ev_}[a,b] (\varphi_0) \mid \text{alw_}[a,b] (\varphi_0) \quad (7)$$

where μ is an inequality constraint between expressions involving variables (written $\text{var}[t]$), constants or derivatives of a variable over time (written $\text{ddt}\{\text{var}\}[t]$). A formula is evaluated at each requested time (thereafter called the *current time*). A formula with an **and** statement is true if both φ_1 and φ_2 are true at the current time. The $\text{ev_}[a,b] (\varphi_0)$ statement is true if φ_0 holds at least once between the current time plus a and the current time plus b . If the interval is omitted, then a is 0 and b is inf , that is, the formula is true if φ_0 holds at least once, anytime after the current time. The statement $\text{alw_}[a,b] (\varphi_0)$ is true if φ_0 holds between current time plus a and current time plus b . Again, if the interval is omitted then a is 0 and b is inf meaning in that case that the formula must hold all the time.

For a given simulation x and a formula φ , Breach can compute a real satisfaction value $\rho(x, \varphi) \in \mathbb{R}$ (quantitative semantics) with the property that $\rho(x, \varphi) \geq 0$ iff x satisfies φ [116]. The problem of finding a parameter set that satisfies an STL formula can be thus reduced to maximizing ρ for φ , in order to find a simulation for which ρ is positive.

2.3.2 Dynamics of the induction of the iron-depleted situation

The experiment we want to simulate is the cut-off of iron entry for cells initially in the iron-replete situation. The total simulated duration is 48 hours, as the qualitative description of the iron-depleted situation tells us that significant changes can be seen after 24 hours. First, the simulations must stabilize in a steady state corresponding to the iron-replete situation. Then, at time $t=6$ hours, the entry of iron is cut by setting Tf_{sat} to zero. We expect the system to increase the IRP regulatory protein concentration in order to maintain a correct iron level for at least ten hours. We use temporal logic formulas to describe these properties. They will then be used to identify possible parameter values.

We first want the system to be in a steady state corresponding to the iron-replete situation. To avoid an inappropriately long stabilization phase, we specify initial conditions close to the value of the stationary state. To ensure that the system reaches a steady state, we use the following temporal logic formula (8). The value $1.0\text{e-}6$ is sufficiently small to ensure that no significant variable variations occur.

$$\begin{aligned} \varphi_{Sstable} = & \text{abs}(\text{ddt}\{\text{FPN1a}\}[t] / \text{FPN1a}[t]) + \\ & \text{abs}(\text{ddt}\{\text{Fe}\}[t] / \text{Fe}[t]) + \\ & \text{abs}(\text{ddt}\{\text{Ft}\}[t] / \text{Ft}[t]) + \\ & \text{abs}(\text{ddt}\{\text{IRP}\}[t] / \text{IRP}[t]) + \\ & \text{abs}(\text{ddt}\{\text{TfR1}\}[t] / \text{TfR1}[t]) < 1.0\text{e-}6 \end{aligned} \quad (8)$$

(To make the reading easier, all formulas dealing with the iron-replete steady state will begin by “ φ_S ”, formulas related to parameters will begin by “ φ_P ” and formula related to the behavior after the cut-off of iron entry will begin by “ φ_B ”). To check that the value of the variables at steady state are in the intervals

shown in Table 1, or deduced as shown in Section 2.4.2, we use the following formulas:

$$\begin{aligned}
\varphi_{S1} &= (1.04e-13 < \text{FPN1a}[t]) \text{ and } (\text{FPN1a}[t] < 1.0e-5) \\
\varphi_{S2} &= (1.0e-10 < \text{Fe}[t]) \\
\varphi_{S3} &= (1.0e-13 < \text{Ft}[t]) \quad \text{and } (\text{Ft}[t] < 1.0e-5) \\
\varphi_{S4} &= (3.0e-9 < \text{IRP}[t]) \quad \text{and } (\text{IRP}[t] < 1.07e-8) \\
\varphi_{S5} &= (1.0e-8 < \text{TfR1}[t]) \quad \text{and } (\text{TfR1}[t] < 8.7e-8)
\end{aligned} \tag{9}$$

As the maximal concentration of iron should remain below 2.0×10^{-6} mol·L⁻¹ upon depletion, and not only at the steady state, we do not enforce this property here.

When the iron input is switched off, the iron level decreases and ferritin releases iron. This release is due to a lower ferritin production rate, which provokes a decrease of ferritin concentration. This can only be triggered a crossing by *IRP* of the threshold $\theta_{IRP \rightarrow Ft}$. This agrees with the previous statement that the variable describing *IRP* concentration is lower than $\theta_{IRP \rightarrow Ft}$ under iron-replete conditions. We translate this property in the following formula:

$$\varphi_{S6} = \text{IRP}[t] < \text{theta_IRP_Ft} \tag{10}$$

The increase of *IRP* concentration can only be triggered by a decrease in iron concentration important enough to cross the threshold $\theta_{Fe \rightarrow IRP}$. This is in line with the already indicated inequality $Fe^{eq} > \theta_{Fe \rightarrow IRP}$ under iron-replete conditions. The corresponding formula is:

$$\varphi_{S7} = \text{Fe}[t] > \text{theta_Fe_IRP} \tag{11}$$

Moreover, when the iron level becomes too low, the *FPN1a* concentration is expected to become lower. This, also, can only be triggered by a crossing by *IRP* of the threshold $\theta_{IRP \rightarrow FPN1a}$. That means that in the iron-replete steady state the *FPN1a* concentration is above this threshold, which is expressed by this formula:

$$\varphi_{S8} = \text{IRP}[t] < \text{theta_IRP_FPN1a} \tag{12}$$

Nevertheless, as explained in Section 2.1.4 (resp. Section 2.1.1), the regulation of *FPN1a* by *IRP* ($k_{IRP \rightarrow FPN1a}$) (resp. of *Ft* ($k_{IRP \rightarrow Ft}$)) cannot generate a term larger than the basal production rate of *FPN1a* (k_{FPN1a_prod}) (resp. *Ft* (k_{Ft_prod})). The formulas expressing these facts are:

$$\varphi_{P1} = k_IRP_FPN1a < k_FPN1a_prod \tag{13}$$

and

$$\varphi_{P2} = k_IRP_Ft < k_Ft_prod \tag{14}$$

The increase of active *IRP* concentration leads to the stabilization of the iron level. We express that by enforcing the existence of a plateau at this level after shutting the iron import off (second line of formula (15)). We impose that this plateau lasts for at least ten hours (first line of formula (15)). As the simulation duration is 48 hours, we ask this plateau to start at least 10 hours before the end of the simulation. To avoid incorrect simulations for which the concentration of this plateau is zero, we set it at,

at least, one hundredth of the concentration in the iron-replete situation (third line of formula (15)). As it is enforced that the system have reached the steady state at time $t=4$ hours (see formula φ_{Sall} below), the steady state iron concentration is expressed by $Fe[4*3600]$. Moreover, as this plateau cannot last forever, we require that, at the end of the simulation, the iron concentration is less than half of its value at the plateau (fourth line of formula (15)).

$$\begin{aligned} \varphi_{B1} = & \text{ev_}[6*3600, (48-10)*3600] (\text{alw_}[0, 10*3600] (\\ & ((\text{abs}(\text{ddt}\{\text{Fe}\}[t] / \text{Fe}[t]) < 1.0\text{e-}5) \text{ and} \\ & (\text{Fe}[t] > 0.01*\text{Fe}[4*3600])) \text{ and} \\ & (\text{Fe}[t] > 2*\text{Fe}[47*3600]) \\ &)) \end{aligned} \quad (15)$$

In addition, we experimentally determined (unpublished) that, in human erythroleukemic KG1 cells, the ferritin concentration is at least three times lower 24 hours after iron scavenging by a chelator than it is at steady state under iron-replete conditions (ferritin concentration at steady state is expressed by $Ft[4*3600]$). This is expressed by:

$$\varphi_{B2} = Ft[30*3600] < Ft[4*3600]/3 \quad (16)$$

When the iron stored in ferritin is exhausted, the iron concentration eventually decreases to zero. Ferritin disappearance precedes iron depletion which is not rescued by replenishment from any internal store. Otherwise, it would mean that the cell has stored iron but is not using it. Moreover a ferritin production rate not close to zero means that released iron from degraded ferritins is used to fill newly created ferritins and not for the cell needs. This implies that the parameter describing the strength of the regulation by the IRP (i.e. $k_{IRP \rightarrow Ft}$) is slightly lower than the production rate (i.e. k_{Ft_prod}), and we set the ratio of these values at 95% at most. This constraint is translated by the following formula:

$$\varphi_{P3} = k_{Ft_prod}*0.95 < k_{IRP_Ft} \quad (17)$$

The following formulas express that the iron concentration is never higher than $2 \times 10^{-6} \text{ mol}\cdot\text{L}^{-1}$, and enforces the inequality $k_{Fe_cons} \geq k_{Fe_export} \cdot FPN1a$, described in Table 1, that holds in the iron-replete situation:

$$\varphi_{B3} = \text{alw} (\text{Fe}[t] < 2\text{e-}6) \quad (18)$$

$$\varphi_{S9} = k_{Fe_cons} > k_{Fe_export}*FPN1a[t] \quad (19)$$

Finally, a formula must assemble all previous ones. For this purpose, we first define the formula φ_{Sall} which aggregates all formulas related to the steady state. As we want to be sure that, before the iron input is switched off, the system reaches and stays at the steady state, we enforce that the properties related to the steady state hold at least for two hours before the sixth hour:

$$\varphi_{Sall} = \text{ev_}[0,6*3600] (\text{alw_}[0,2*3600] (\varphi_{Sstable} \text{ and } (\varphi_{S1} \text{ and } (\dots \text{ and } \varphi_{S9})))) \quad (20)$$

We also introduce the formulas φ_{Ball} (resp. φ_{Pall}) enforcing that all formulas related to behavior (resp. parameter) are satisfied:

$$\varphi_{Ball} = \varphi_{B1} \text{ and } (\varphi_{B2} \text{ and } \varphi_{B3}) \quad (21)$$

$$\varphi_{Pall} = \varphi_{p1} \text{ and } (\varphi_{P2} \text{ and } \varphi_{P3}) \quad (22)$$

And finally, the formula φ_{all} , which enforces that all properties are satisfied:

$$\varphi_{all} = \varphi_{Sall} \text{ and } (\varphi_{Ball} \text{ and } \varphi_{Pall}) \quad (23)$$

2.4 Iron-replete steady state

The model being now defined and the available biological data formalized we start the study of the model by the analysis of the iron-replete situation, which is modeled by a steady state. To increase the readability, the concentrations at this steady state are represented by adding the superscript “ eq ” to the name of the variables. In this state, the iron concentration is expected to be significantly higher than the threshold $\theta_{Fe \rightarrow IRP}$. We can thus infer that the sigmoidal term $\text{sig}^+(Fe^{eq}, \theta_{Fe \rightarrow IRP})$ is very close to one. The IRP concentration is also expected to be significantly below the thresholds $\theta_{IRP \rightarrow Ft}$ and $\theta_{IRP \rightarrow FPN1a}$, so that the two sigmoidal functions $\text{sig}^+(IRP^{eq}, \theta_{IRP \rightarrow Ft})$ and $\text{sig}^+(IRP^{eq}, \theta_{IRP \rightarrow FPN1a})$ are close to zero.

2.4.1 Analytical analysis

From this simplified system of equations, we check whether the system has a unique steady state. For this purpose, we set the derivatives in the model to zero and use the symbolic solver of the open source software Sage [117] to check whether it is possible to express the steady state concentrations as functions of the model parameters. If this is possible then each instantiation of the parameters corresponds to a unique steady state given by these equations. The iron-replete situation being modeled by a steady state, to get a coherent model, it is necessary to prove that this steady state is stable. For this purpose, we compute the Jacobian matrix of the system, simplified as explained in the beginning of Section 2.4, using Sage [117]. Stationary points in regions such that the eigenvalues of this Jacobian are real and negative are proven to be stable.

2.4.2 Reduction of search space

Exploiting the symbolic expression of the steady state concentrations and using the specific data we have on the parameters and the concentration (see Table 1), we then aim at reducing the volume of the parameter space to be searched. To perform this task we use an interval solver. An interval solver uses intervals on unknowns and algebraic relations (equalities and inequalities) between these unknowns to reduce the provided intervals, classically using a branch-and-prune algorithm [106]. The software Realpaver [106] is a recent interval solver which manages real variables and non-linear constraints. By considering the algebraic equations between parameters and variables, as well as the defined intervals and constraints as inputs for the calculation, Realpaver provides a way to reduce the estimated intervals and, consequently, to reduce the parameter search space.

2.5 Exploring the parameter space

In this subsection, we present our approach for exploring the parameter space defined in the previous steps.

In order to reduce the dimension of the search space to explore and to avoid some invalid parameter values, we use the expressions found in Section 2.4.1 to compute the initial conditions of the system as a function of the parameters. These expressions rest on the hypothesis that the IRP concentration in the iron-replete steady state (IRP^{eq}) is lower than the thresholds $\theta_{IRP \rightarrow Ft}$ and $\theta_{IRP \rightarrow FPN1a}$, and that the iron concentration (Fe^{eq}) is higher than the threshold $\theta_{Fe \rightarrow IRP}$. To impose these considerations, we compare the computed value of IRP^{eq} (resp. Fe^{eq}) and the value of the thresholds. Then, each threshold value is set to the maximum (resp. minimum) of its value and of $1.258 \times IRP^{eq}$ (resp. $Fe^{eq}/1.258$), in order for the sigmoidal terms to be lower than 0.01 (resp. greater than 0.99). In addition, to ensure that the formulas φ_{P2} and φ_{P3} are satisfied, we enforce that $k_{IRP \rightarrow Ft}$ is set equal to $0.975 \times k_{Ft_prod}$ (any value between 0.95 and 1 being valid, we chose the mean). Finally, if the parameter k_{FPN1a_prod} is smaller than $k_{IRP \rightarrow FPN1a}$, we swap their values to force the formula φ_{P1} to be satisfied. Thus, the dimension of the search space is reduced from twenty-four to eighteen. As a consequence, some sub-formulas of φ_{all} does not need to be verified as they hold by definition. The formulas φ_{Sall} and φ_{all} are rewritten in:

$$\varphi_{Sall} = \text{ev_}[0,6*3600] (\text{alw_}[0,2*3600] (\varphi_{Sstable} \text{ and } (\varphi_{S1} \text{ and } (\varphi_{S2} \text{ and } (\varphi_{S3} \text{ and } (\varphi_{S4} \text{ and } (\varphi_{S5} \text{ and } \varphi_{S9})))))))) \quad (24)$$

$$\varphi_{all} = \varphi_{Sall} \text{ and } \varphi_{Ball} \quad (25)$$

Despite the associated reduction in the parameter space, an automatic search method is still necessary. For this purpose, we have implemented an optimization algorithm to maximize the satisfaction value of φ over the parameters of the system. As the optimization algorithm must handle a non-linear, multi-variable objective function (here, the evaluation of φ as a function of the parameters), we use a variant of the Nelder-Mead [118] algorithm which fulfills these requirements. It proceeds as follows:

1. it creates a simplex by picking randomly $n+1$ parameter sets $\mathcal{P}_0 \dots \mathcal{P}_n$ (where n is the number of parameters) in the neighborhood of a parameter set \mathcal{P}_U provided by the user;
2. it computes $y_i = \rho(x_{\mathcal{P}_i}, \varphi)$ for $i \in [0, n]$ where $x_{\mathcal{P}_i}$ is the simulation associated to the parameter set \mathcal{P}_i ;
3. it sorts out the $n+1$ parameter sets such that \mathcal{P}_0 is the best (i.e. y_0 is the highest), and \mathcal{P}_n is the worst (i.e. y_n is the lowest);
4. it computes $\bar{\mathcal{P}}$, the average of $\{\mathcal{P}_0, \dots, \mathcal{P}_{n-1}\}$, the n best parameter sets;
5. it computes $y^* = \rho(x_{\mathcal{P}^*}, \varphi)$, where \mathcal{P}^* is the reflection of \mathcal{P}_n with respect to $\bar{\mathcal{P}}$;
6. if $y^* > y_n$, it either moves or extends the simplex, according to the order between y^* and y_0 , by replacing y_n by a parameter set on the half-line $[\bar{\mathcal{P}}, \mathcal{P}^*]$;
7. else, it computes \mathcal{P}^{**} , belonging to the segment $[\mathcal{P}_n, \bar{\mathcal{P}}]$. If $y^{**} > y_n$, it contracts the simplex by replacing \mathcal{P}_n by \mathcal{P}^{**} ;

8. else, it makes an homothetic transformation of all \mathcal{P}_i , $i \in [1, n]$, with a scale factor -0.5 and \mathcal{P}_0 as center;
9. it loops to step 3 until the maximum number of iteration is reached or the size of the simplex is small enough.

As shown in Section 2.2, the parameter range may extend over many orders of magnitude. The $n+1$ initial parameter sets picked in step 1 of the algorithm fall in the same range as the one provided by the user (\mathcal{P}_U) which limits the width of the exploration. To avoid this limit, we randomly choose, on a logarithmic scale, a set of parameter sets and run many times the above optimization algorithm using each of them as \mathcal{P}_U . Thus, an extensive range is explored for the parameter sets. In order to optimize these multiple runs, φ is evaluated for each randomly picked parameter set and the algorithm is first run with the parameter set providing the highest evaluation.

When the formula φ is complex, which is the case in our study, it can be advantageous to decompose it into sub-formulas and search incrementally for valid parameter set. Assume for instance that $\varphi = \varphi_1$ and φ_2 . We start with N initial parameter sets $\mathcal{P}_{U_1} \dots \mathcal{P}_{U_N}$ and run the optimization algorithm for φ_1 (resp. φ_2) to obtain a set of optimized parameter sets $\mathcal{P}'_{U_{1,1}}, \dots, \mathcal{P}'_{U_{N,1}}$ (resp. $\mathcal{P}'_{U_{1,2}}, \dots, \mathcal{P}'_{U_{N,2}}$). Then we use the parameter sets $\mathcal{P}'_{U_{1,1}}, \dots, \mathcal{P}'_{U_{N,1}}, \mathcal{P}'_{U_{1,2}}, \dots, \mathcal{P}'_{U_{N,2}}$ as initial sets for the optimization of $\varphi = \varphi_1$ and φ_2 . In our study, when applying the optimization based strategy with φ_{all} , we were unsuccessful in finding a valid parameter set. Consequently, we first optimize for φ_{B1} , φ_{B2} , φ_{B3} and φ_{Sall} and used the obtained parameter sets as initial parameter sets for the search of parameter sets for φ_{all} .

2.6 Identifying a valid region

Having defined the formula φ which expresses the relevant properties and found a valid instantiation of the parameters, we want to define a region containing only valid parameter sets by exploring the neighborhood of this instantiation, and expanding from there.

Due to the high-dimensionality of the search space, a brute force algorithm, like a systematic sampling around the valid parameter sets found, must be avoided. In their recent paper, Zamora-Sillero et al. [119] provide a method based on ellipsoids to search, sample and characterize a valid parameter region. This method has the advantage that its performances scale linearly with the number of dimensions. Moreover, when transforming the search space on a log scale, this method allows the exploration of a very large search space. Finally, as it uses many ellipsoids to characterize the valid parameter region, the method makes no assumptions on the convexity and even the connexity of the valid parameter region. As we have no information on the shape of the valid region and as the parameter space to explore is large, this method is well suited to our problem.

The method proceeds in two steps: 1) it computes a coarse-grained sampling of the valid parameter region used to 2) gain a finer-grained sampling of the valid parameter region. The first step consists in applying an out-of-equilibrium adaptive Metropolis Monte-Carlo algorithm which aims at acquiring a sample of parameter sets distributed all over the valid parameter region. Then, from these parameter sets, the second

step consists in a new algorithm called “multiple ellipsoid-based sampling”. It iteratively explores the space locally around a valid parameter set and constructs ellipsoids that enclose the valid parameter sets found. This last step provides a finer sampling of the valid parameter region. An optional last step consists in a characterization of the valid parameter region with a set of ellipsoids enclosing it and a uniform sampling of these ellipsoids (namely, a set of parameter sets sampling this parameter region such that the values for each parameter follow a uniform distribution), thus providing a uniform sampling of the valid parameter region. This uniform sampling is used to compute the volume of the union of ellipsoids and the volume of the valid parameter region found.

We aim at describing the set of ellipsoids enclosing the valid parameter region to gain information on its shape. For this purpose, we modify the last step of the method described by Zamora-Sillero [119] in two ways. First, we save the set of ellipsoids enclosing the valid parameter region. Thus, we can compute, for each ellipsoid, its volume and its eccentricity, providing insight on the shape of the valid parameter region. Second, we improve the function computing the uniform sampling in order to get intersections between the ellipsoids, so we can address the question of the connexity of the valid parameter space.

We also aim at describing some statistical properties of the set of valid parameter sets. As the parameter values span many orders of magnitude, we first translate them on a logarithmic scale. To investigate the existence of relationship between parameters, we compute the correlation between each couple of parameters, using all the valid parameter sets found. The correlation is computed with Kendall’s rank correlation method [120]. Unlike many methods which require a normal distribution of the data, this method has no requirements on the distribution of the parameters, and is thus well suited to our problem. The Matlab statistical toolbox [121] provides a function that computes Kendall’s tau. For each couple of parameters, the function also computes a *p-value* which can be used to test the significance of the null hypothesis (i.e. no correlation). We consider that two parameters are correlated if the p-value is lower than a specified threshold.

Finally, to determine the influence of each parameter on the formula satisfaction value, we consider the computation of a *multiple linear regression*. A multiple linear regression tries to explain the values taken by a *response variable* in function of the values of *predictor variables*. Basically, it searches the values of the coefficients $\beta_0, \beta_1, \dots, \beta_n$ of the following linear function (also called linear model):

$$y_j = \beta_0 + \beta_1 \cdot x_{1,j} + \beta_2 \cdot x_{2,j} + \dots + \beta_n \cdot x_{n,j}$$
, where y is the response variable, and the $x_{i,j}$ are the predictor variables, such that the computed y has the best fit of the real y for all the observations (the j index is the observation index). Each coefficient β_i comes with a p-value that measures the significance of the coefficient. From this linear model, we can compute the adjusted R-square statistic which describes the percentage of the response variable variance explained by the linear model [122]. This adjusted R-squared gives an insight into the reliability of the linear model. Then, using the analysis of the variance, we describe the percentage of variance of the response variable explained by each predictor variable [122]. A preliminary step consists in determining the set of predictor variables that compose the linear model. Indeed, an increased number of predictor variables provides a better R-square statistic, but may overfit the response. For this purpose, we use a slightly simplified version of the method described by

Venables & Ripley [123], implemented in the statistical software R [124] which is based on the Akaike information criterion (AIC). This criterion is a measure of the relative quality of a statistical model for a given set of data that deals with the goodness of the fit and the number of predictor variables. This criterion allows us to automatically compare different models and choose the best one (namely, the one with the lowest AIC value) [125]. Using a multiple linear regression, with the formula satisfaction value, $\rho(\varphi_{all})$, as the predicted variable and the unknowns of the system (i.e. parameters and concentrations at steady state) as the predictor variables, provides a description of the impact of each parameter on the formula evaluation. As for the correlation computation, the parameter values are translated on a logarithmic scale. Once the best set of predictor variables is chosen, the linear regression provides a value for all the regression coefficients, their p-values and the percentage of variance of $\rho(\varphi_{all})$ explained by the corresponding parameters. An interface between R and Matlab is provided by the package R.matlab [126].

3 Results

In the previous sections, we have described the model and defined the expected dynamic using the STL logic that provides a convenient framework to express behavioral properties. In this section, we will first summarize the result of the modeling of iron-replete cells considered as a steady state. Then, we will examine the dynamics of the system upon iron withdrawal, including the biological insight provided by the theoretical analysis.

3.1 Steady state

We explore the properties of the steady state modeling the iron-replete situation.

3.1.1 Uniqueness

When considering the steady state, the system of differential equations that rule the system is rewritten in a system of algebraic equations simplified in accordance to collected data related to the iron-replete steady state, as explained in Section 2.4. The analytical resolution of this system of algebraic equations by the symbolic solver Sage [117] provides a unique solution, described by equations (26) to (30). The uniqueness of this solution proves that for each set of parameter values (generated as indicated in Section 2.6), a unique steady state exists (provided the concentrations satisfy the other constraints). Moreover, the analytical expressions provide these values for every parameter set.

$$FPN1a^{eq} = \frac{k_{FPN1a_prod}}{k_{FPN1a_deg}} \quad (26)$$

$$Fe^{eq} = \frac{((k_{IRP_deg} + k_{Fe \rightarrow IRP}) \cdot k_{TfR1_prod} + k_{IRP \rightarrow TfR1} \cdot k_{IRP_prod}) \cdot k_{Fe_input} \cdot k_{FPN1a_deg} \cdot Tf_{sat}}{(k_{FPN1a_prod} \cdot k_{Fe_export} + k_{FPN1a_deg} \cdot k_{Fe_cons}) \cdot (k_{IRP_deg} + k_{Fe \rightarrow IRP}) \cdot k_{TfR1_deg}} \quad (27)$$

$$Ft^{eq} = \frac{k_{Ft_prod}}{k_{Ft_deg}} \quad (28)$$

$$IRP^{eq} = \frac{k_{IRP_prod}}{k_{IRP_deg} + k_{Fe \rightarrow IRP}} \quad (29)$$

$$TfR1^{eq} = \frac{(k_{IRP_deg} + k_{Fe \rightarrow IRP}) \cdot k_{TfR1_prod} + k_{IRP \rightarrow TfR1} \cdot k_{IRP_prod}}{(k_{IRP_deg} + k_{Fe \rightarrow IRP}) \cdot k_{TfR1_deg}} \quad (30)$$

3.1.2 Stability

The stability of the stationary point is then addressed by computing the Jacobian matrix of the system of algebraic equations, using the software Sage [117]. The obtained matrix is shown in (31) with the following order of variables: $(Fe, TfR1, FPN1a, Ft, IRP)$. We observe that this matrix is an upper-triangular matrix. It follows that the eigenvalues are equal to the diagonal elements. It is thus clear that they are all real and negative. This formally prove that, for all parameter sets, the computed steady state is stable.

$$\begin{pmatrix} -FPN1a^{eq} \cdot k_{Fe_export} - k_{Fe_cons} & Tf_{sat} \cdot k_{Fe_input} & -Fe^{eq} \cdot k_{Fe_export} & k_{Ft_deg} \cdot n_{Ft} & 0 \\ 0 & -k_{TfR1_deg} & 0 & 0 & k_{IRP \rightarrow TfR1} \\ 0 & 0 & -k_{FPN1a_deg} & 0 & 0 \\ 0 & 0 & 0 & -k_{Ft_deg} & 0 \\ 0 & 0 & 0 & 0 & -k_{IRP_deg} - k_{Fe \rightarrow IRP} \end{pmatrix} \quad (31)$$

3.1.3 Interval reduction

The interval solver Realpaver [106] gives a way to reduce the intervals using the parameter and variable intervals, as the steady-state algebraic equations and constraints. The results returned by the solver are shown in Table 2. We observe that returned intervals for the five parameters k_{IRP_prod} , $k_{Fe \rightarrow IRP}$, $\theta_{IRP \rightarrow FPN1a}$, $\theta_{IRP \rightarrow Ft}$ and k_{Fe_export} are reduced by many orders of magnitude and that six other intervals are reduced. This demonstrates that some region of the parameter space need not be considered when searching for a parameter set producing an adequate iron-replete steady state.

3.2 Existence of a valid parameter set

After expressing the known behavior of the system in the STL formalism, as described in Section 2.3, we study whether there exists at least a valid parameter set, thus proving that the data are not inconsistent. For this purpose, we use the variant of the Nelder-Mead optimization algorithm [118] as described in Section 2.5. For each of the sub-formula φ_{B1} , φ_{B2} , φ_{B3} and φ_{Sall} , 150 initial parameter sets are randomly picked on a logarithmic scale and the number of iterations of the algorithm is set to 70. Then, we search a parameter set verifying φ_{all} : we run the algorithm from each of the parameter sets obtained with a maximal number of iterations set to 250, until a parameter set verifying φ_{all} is found. We obtain a parameter set verifying all the data collected, in 3.5 hours on a desktop computer. This parameter set, called \mathcal{P}_1 and displayed in Table 3, demonstrates that the heterogeneous data extracted from the literature are not inconsistent as a whole.

The examination of this valid parameter set \mathcal{P}_1 provides some preliminary insight. In particular, the IRP regulatory parameters, for which little is known, deserve some comment. The values of the rate constants decreasing the production of proteins, the mRNA of which contains a 5'-UTR IRE, normalized by the

protein concentrations (i.e. $\frac{k_{IRP \rightarrow Ft}}{Ft^{eq}}$ and $\frac{k_{IRP \rightarrow FPN1a}}{FPN1a^{eq}}$) differ by two orders of magnitude. In other words, the impact of active IRP on ferritin is larger than that on FPN1a. Interestingly, the threshold values of IRP on each of these mRNA are close. Of course this is only a first indication since, as was explained in the introduction, the whole valid parameter region must be considered to draw safe conclusions

3.3 Valid parameter region

In the previous section, we found an initial parameter set, \mathcal{P}_1 , such that the simulation, shown in Figure 3, verifies the known behavioral data on the iron-depleted situation. Following the line of Section 2.6, \mathcal{P}_1 is used as a starting point (in parameter space) to find a region corresponding to simulations verifying the specified properties. To avoid to restrict the exploration to the order of magnitude of the parameter values in \mathcal{P}_1 , we consider a logarithmic scale.

3.3.1 Enclosing ellipsoids

The execution of the method presented in Section 2.6 returns a union of 18 ellipsoids enclosing the valid region found and a uniform sampling of this union, composed of 14917 valid parameter sets. The computation of this union of ellipsoids and its sampling takes four hours on a desktop computer.

For each ellipsoid, we compute the size and the ratio of the semi-major axis over semi-minor axis (eccentricity). The obtained results, shown in Table 4, indicate that the volume of the ellipsoid are close to one another, considering the dimensionality of the parameter space. The ratio of the highest volume (third ellipsoid) over the lowest one (the 17th), on a logarithmic scale, is around 4900, which would correspond to a multiplication by only 1.6 of the semi-axis length. Then we address the question of the connexity of the union of ellipsoids. We find that all ellipsoids intersect at least one other ellipsoid and that there are no separate clusters of ellipsoids.

Then one study how accurately this union of ellipsoid encompasses the valid parameter space. For this purpose, we compute the volume of the union of ellipsoids and the volume of the valid parameter region contained in this union. We find that $49.7\% \pm 2.3\%$ of the space described by the union is valid. With regard to the dimension of the space, this result highlights the accuracy of a representation based on ellipsoids. For comparison, we compute the volume of the hyper-rectangle which encompasses the valid parameter region found. The valid fraction of the hyper-rectangle volume is only $0.007\% \pm 0.0002\%$, four orders of magnitude lower than the valid fraction of the union of ellipsoids, proving the need for a more elaborated representation.

3.3.2 Parameter sets

From the uniform sampling composed of 14917 valid parameter sets, we explore the repartition of each parameter and initial condition over the initial search interval. For this purpose, we split, on a logarithmic scale, the search interval in sub-intervals and plot the quantity of valid parameter sets in each sub-interval. Figure 4, which depicts these results, shows that, for some parameters, valid values belong to a restricted interval. Typically, no valid parameter sets such that $k_{Fe \rightarrow IRP}$ is lower than 1.0×10^{-4} , or such that

k_{Fe_cons} is lower than 1.0×10^{-2} were found. On the opposite, we observe that parameters related to *FPN1a* span over the whole initial interval (no reduction seen on this 1D projection). We also notice that the values of the threshold $\theta_{Fe \rightarrow IRP}$ are mainly enclosed in the interval $[1.0 \times 10^{-12} ; 1.0 \times 10^{-10}]$.

Now we would like to characterize the typical behaviors associated to the valid parameter region. To answer this question, we compute the trajectories corresponding to the parameter sets that are the ellipsoid centers. All these trajectories, shown in Figure 6, exhibit a very similar behavior, except one for which the iron concentration decreases more quickly at the end of the plateau (purple curve). We examine the parameter set leading to this simulation and observe that the values for the parameters k_{Ft_prod} and k_{Ft_deg} are ten to a hundred times higher than their values in the other ellipsoid centers. This allows a rapid decrease of ferritin, even at low concentration, thus maintaining a stable iron plateau until the ferritin concentration is almost equal to its lowest possible value. We also observe that, among these trajectories, *FPN1a* initial concentration spans its whole initial search interval and that *FPN1a* concentration does not significantly decrease after iron cut-off. This observation is further discussed in Section 3.5.

Now that we got a set of trajectories corresponding to the expected behavior of the model, we ask whether we can find trajectories which verify the STL constraints, but which are qualitatively different from the ones shown in Figure 6. Such trajectories provide a different perspective for the system: either these trajectories are biologically impossible, thus leading to a better definition of the expected behavior, or they give an indication on the span of possible system behaviors. To search such trajectories, we select, from the uniform sampling of the valid region, all parameter sets having a parameter with an extremal value. We then manually pick trajectories which look qualitatively different. These trajectories are shown in Figure 7. In this figure, we notice some extremal behavior, like a plateau duration ranging from less than 15 hours (green curve) to almost 30 hours (purple curve). We also notice a high variability in the iron concentration in the plateau values of iron concentration. These values may differ by more than two orders of magnitude (red and blue curves). We also remark that some simulations present a strong transient iron depletion immediately following the removal of iron input from transferrin, that is rescued by replenishment from internal store iron in ferritin. The plateau can consequently start a long time (until 10 hours) after the iron cut-off. Otherwise, we notice that the IRP concentration also exhibits a high variability during the iron plateau, as its concentration may differ by more than two orders of magnitude (red and yellow curves). Finally, we note that, as in Figure 6, the IRP concentration converges toward more than 1.0×10^{-6} mol·L⁻¹, which is the concentration that IRP would have if the iron does no regulation. The comparison of this concentration with the IRP concentration at iron-replete steady-state shows that iron has a strong regulation on IRP.

3.3.3 Longer exploration

In the Sections 3.3.1 and 3.3.2, we presented the ellipsoids and the encompassed valid parameter sets obtained with the method described in Section 2.6. The two first steps of this method have two possible termination criteria: the first is the convergence of the step, the second consists of reaching the maximal number of iterations. In our calculations both steps stop because of the second criteria. A question is thus

whether our results depend on the number of iterations.

To answer this question, we run again the whole procedure described in Section 2.6 with a number of iterations forty times higher. The first step ends because the maximal number of iterations is reached, while the second ends because the algorithm has converged. The total computation takes 46 hours on a desktop computer and provides an uniform sampling of 53 463 valid parameter sets. Figure 8 shows the new repartition of parameter values. A comparison between this figure and Figure 4 shows that the parameters $k_{Fe \rightarrow IRP}$ and $k_{Fe_{cons}}$ range over a larger interval (reduction of the lower bound by a factor 19 and 554 respectively), and it is also the case, to a lesser extent, for the parameter $k_{Fe_{export}}$. As a consequence, the interval for the iron concentration at steady state is broader, and the interval for the threshold $\theta_{Fe \rightarrow IRP}$ is broader too. We also observe that the upper limit of the parameter $k_{Ft_{prod}}$ is increased and that the repartition of the values for parameter $k_{Ft_{deg}}$ changes, leading to a larger interval for the ferritin concentration at steady state. Finally, we note a decrease by more than one order of magnitude of the lower limit of the parameter $k_{IRP_{prod}}$. Consequently to these increases of the interval ranges, the computation of the valid parameter region as indicated in Section 2.6 shows that the volume of the valid parameter region, on a logarithmic scale, is increased by a factor 2 394 in comparison of the valid parameter region obtained in Section 3.3.1.

Observing with care the repartition of $\theta_{Fe \rightarrow IRP}$ values, we find that they are mainly in the interval $[1.0 \times 10^{-10} ; 1.0 \times 10^{-8}]$. Thus, the threshold of *IRP* regulation by iron ($\theta_{Fe \rightarrow IRP}$) is worth comparing to the concentration which can be mobilized for biosynthetic and other needs (variable Fe of our model). Previous experimental determinations of the latter using fluorescent probes have produced values in the $\mu\text{mol}\cdot\text{L}^{-1}$ range, often considered as an estimate of the *labile iron pool*. It turns out that $\theta_{Fe \rightarrow IRP}$ is mainly at least two order of magnitude smaller than the $\mu\text{mol}\cdot\text{L}^{-1}$ range, i.e. *IRP* is sensitive to far lower concentrations of iron than those readily accessible by most cellular reactions using the metal. The molecular reasons for such extreme sensitivity of the cellular iron detection system will have to be explored. Before that, it is worth mentioning that the major buffering component for iron is ferritin as shown in Figure 1, and that poly-r(C)-binding proteins (PCBP) have been recently proposed to act as iron chaperones for ferritin. It is thus likely that PCBP, and possibly other iron chaperones not yet identified, modulate the available iron under iron-sufficient conditions, whereas *IRP* become prominent at lower concentrations when iron is scarce. This is illustrated in the simulation of Figure 6 where it can be seen that the active *IRP* concentration only significantly increases when the iron resources are almost exhausted. The mechanism is also in agreement with the dissociation constants of iron and PCBP1 which lie in the $\mu\text{mol}\cdot\text{L}^{-1}$ range [75].

These results indicate that parameters already extended over the whole valid parameter interval, except for six of them. Moreover, as the second step ends due to convergence, it leads us to assume that considering a higher number of iterations will not provide large new valid region. The new valid parameter region found is considered in the following sections.

3.4 Statistical analysis

The next step consists in extracting information from the set of valid parameter sets obtained. This is done in two different ways. The first one consists in computing correlations between parameters. The second consists in computing the effect of the parameters on the formula evaluation, by mean of a multiple linear regression as explained in Section 2.6.

For each couple of parameters, we compute a correlation value, using the Kendall's tau correlation coefficient [120], based on the 526208 valid parameter sets provided by all the steps of the method described in Section 2.6. The results, obtained in 97 hours on a desktop computer, provides 252 significant correlations. The manual inspection of these correlations indicates that many of them are obvious, like the correlation between the parameter k_{Ft_prod} and $k_{IRP \rightarrow Ft}$, but that others are unexpected. Figure 9 shows three interesting correlations: k_{Fe_cons} versus $\theta_{Fe \rightarrow IRP}$, k_{Ft_deg} versus k_{Ft_prod} and n_{Ft} versus Ft^{eq} . The positive correlation between k_{Ft_deg} and k_{Ft_prod} (i.e. the higher k_{Ft_deg} is, the higher k_{Ft_prod} is) indicates that the iron-replete steady state concentration of ferritin must lie in a specific range. This is in line with the repartition of the initial concentration of the ferritin shown in Figure 8. Then, the negative correlation between Ft and n_{Ft} indicates that the quantity of iron stores in ferritins at steady state must stay within a specific interval. Finally, the surprising negative correlation between k_{Fe_cons} and $\theta_{Fe \rightarrow IRP}$ expresses that when the internal needs are high, the threshold of the iron regulation on IRP is low.

The next question is the following: how do the different parameters account for the variance of the evaluation of the STL formula? And consequently, how do they impact on the behavior of the system? For this purpose, we compute a multiple linear regression based on all valid parameter sets obtained from the uniform sampling of the valid region, as described in Section 2.6. In less than one minute on a desktop computer, the regression provides a linear model that predicts φ_{all} evaluation as a function of the value of sixteen unknowns of the system. This results shows that the nine unknowns $Tf_{sat,0}$, k_{FPN1a_prod} , k_{Fe_cons} , k_{Fe_export} , k_{Ft_prod} , $k_{IRP \rightarrow FPN1a}$, $k_{IRP \rightarrow Ft}$, k_{IRP_deg} and $\theta_{IRP \rightarrow FPN1a}$ do not increase the goodness of the fit by a sufficient amount to be included in the linear model. In other words they are found to be non-significant for the regression. Moreover, the p-value for the coefficients corresponding to the parameters $k_{Fe \rightarrow IRP}$ and k_{IRP_prod} are higher than 1.0×10^{-5} , thus indicating that these parameters do not explain the evaluation of the STL formula. The computation indicates that the adjusted R-square statistic is 0.567, meaning that 56.7% of the variance of φ_{all} evaluation is explained by the regression model. Finally, the percent of variance of φ_{all} evaluation explained by each parameter is shown in Figure 10. We observe that the four unknowns that explain the main part of the variance are the steady state values of ferritin, ferroportin and iron, and the value of the threshold $\theta_{Fe \rightarrow IRP}$. It may first be surprising that the ferroportin concentration at steady state explains part of the variance of φ_{all} evaluation as valid simulations are found whatever the ferroportin concentration at steady state is. Nevertheless $FPN1a^{eq}$ appears in the sub-formula φ_{S9} , and thus influences the evaluation of φ_{all} . The variance explained by Ft^{eq} and Fe^{eq} indicates that the amount of available iron (both directly available or stored) influences the evaluation of φ_{all} . Finally, the threshold $\theta_{Fe \rightarrow IRP}$ appears to significantly explain the variance of φ_{all} evaluation.

3.5 FPN1a insensitivity to IRP

The simulation exhibited In Figure 3, shows that *FPN1a* variable remains after iron withdrawal at levels close (within 8%) to its initial (iron-replete) value. It is the only variable of our model for which the minimal value is not close to zero during the simulation. Moreover, in both Figure 6 and Figure 7, we observe that, in the iron-depleted situation, the FPN1a concentration remains at a rather high level (relatively to its initial value), whatever its initial concentration. This implies that the considered cell keeps its ability to export iron despite shortage of the metal. This is due to the low ferroportin degradation rate that leads to the low decrease of the ferroportin concentration. This is a significant outcome of modeling which suggests that the IRP system is not sufficient to strongly regulate FPN1a, at least at this time scale and in the biological context (myeloid progenitors) emphasized herein. It may be proposed that the (IRE-containing) FPN1a expression must be maintained in myeloid progenitors even under conditions of iron deficiency and when IRE-less isoforms are not sufficiently produced [71], to signal the body iron need for erythropoiesis. Furthermore, ferroportin has been shown to be subjected to additional regulatory mechanisms involving the systemic hormone hepcidin [94], the interaction with FPN1a of which triggers degradation of the transporter, or of miRNA [92].

4 Conclusion

Three models of the core iron homeostasis system have been published previously. Omholt et al. [96] and Mitchell et al. [104] define differential equations and then manage to instantiate all the model parameters. Chifman et al. [103] has a radically different approach. They do not use experimental data and prove general mathematical properties but of a very limited type (bearing on the steady state and its shift when some condition is changed). Here we have chosen an approach intermediate between these two poles. The available experimental data allowed us to apply inequality constraints between the parameters, together with behavioral constraints expressed in a temporal logic formalism. We could thus study not only steady states, but also dynamical responses to perturbations, as the sharp decrease of iron input detailed herein. Regions of the parameter space for which the model exhibits the expected behavior have been characterized. It is worth mentioning that this characterization posed significant challenges due to the level of uncertainty in the parameters. We addressed these challenges by using a combination of formal methods and optimization heuristics exploring different orders of magnitudes in the parameter space. Also, our approach is incremental, in the sense that new constraints expressed in STL (coming, e.g., from new experiments) can be easily added in our framework, to further validate or improve the model.

Finally, the model of cellular iron homeostasis in myeloid progenitors presented herein highlights key steps in the handling of this essential nutriment that has been recently shown to participate in the fate of these cells, be it proliferation (mainly in the case of pathological hematopoiesis, e.g. leukemia), or differentiation, or death. An attainable perspective is to adjust this model and to extend it with connected regulatory networks to different stage along the differentiation pathways of hematopoietic cells. It may thus be ambitious to provide a comprehensive view of this complex biological process under normal and pathological conditions with innovative insights into critical parameters as first demonstrated in the present work. It also has been shown that not-well-understood links between iron and redox homeostasis

exist [127], and modeling of such an extended biological system may give insight into iron management under oxidative stress. In addition the regulatory mechanisms of *IRP* could be described more precisely by integrating mRNA concentrations in the model. These extensions will be performed in a controlled and stepwise manner. The qualitative behavior exhibited by extended models will be compared with that of the core model presented here.

Competing interests

The author(s) declare that they have no competing interests.

Authors contributions

NM, AD, EF and JMM wrote the manuscript.

EP, JMM and PM designed and carried out biological experiments.

NM and EF designed the method.

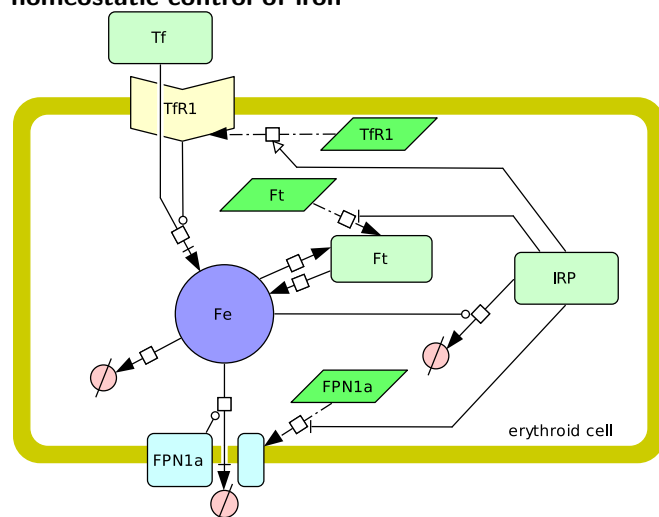
NM and AD developed the method.

Acknowledgements

This work was supported by Microsoft Research through its PhD Scholarship Programme. The authors thank for financial support: the Région Rhône-Alpes (project funded through the Cible 2010 Program) and the Rhône Alpes Complex Systems Institute (IXXI). We also thank Olivier François for his useful advices.

Figures

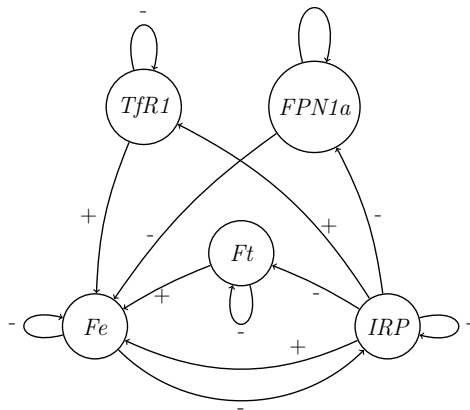
Figure 1 - Schematic representation of the main biological processes involved in the cellular homeostatic control of iron



The acronyms and abbreviations used in this graph are explained in Section 1.2.1. The dashed arrows represent translation of mRNA into proteins. The arrow ending in an empty white head represents translational enhancement, while the lines with a broken ending arrowhead represent inhibition of

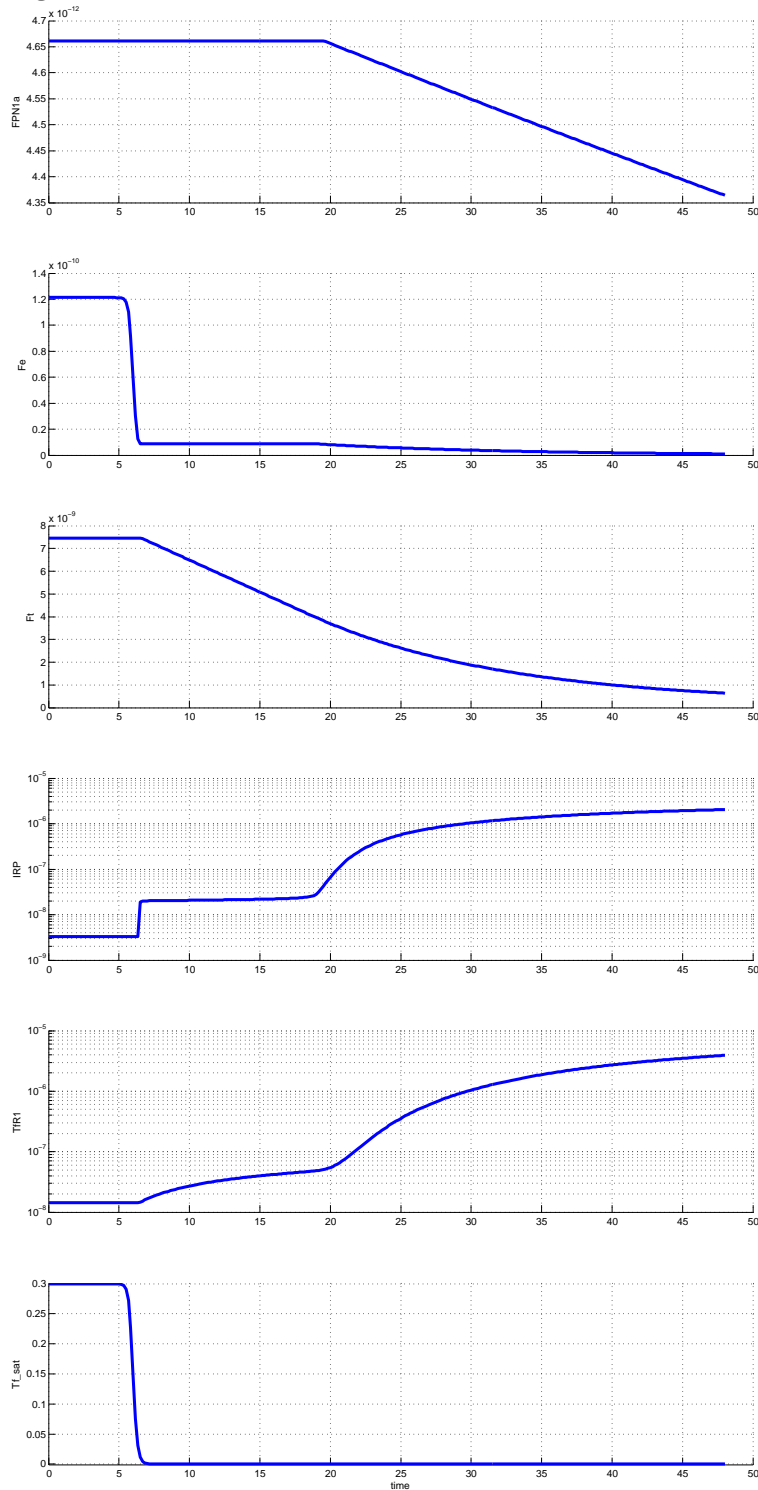
translation. The lines ending with a combined perpendicular stroke and arrow represent iron transport through membranes. The arrow leading to an empty set symbol indicates either degradation (of IRP) or use (consumption and export of iron). Finally, the two regular arrows represent the loading/unloading of iron into/from the ferritins (Ft). The rounded rectangles represent proteins, the parallelograms represent mRNA, and the circle labeled Fe represents the pool of available iron. The concave hexagon represents the transferrin receptor ($TfR1$). This diagram was drawn with the software CellDesigner [55].

Figure 2 - Interaction graph



Interaction graph defined by the equations (2) to (5). The signs on the arrows indicate whether the interaction is positive or negative. The interaction graph contains both positive and negative circuits, leading to a complex dynamical behavior.

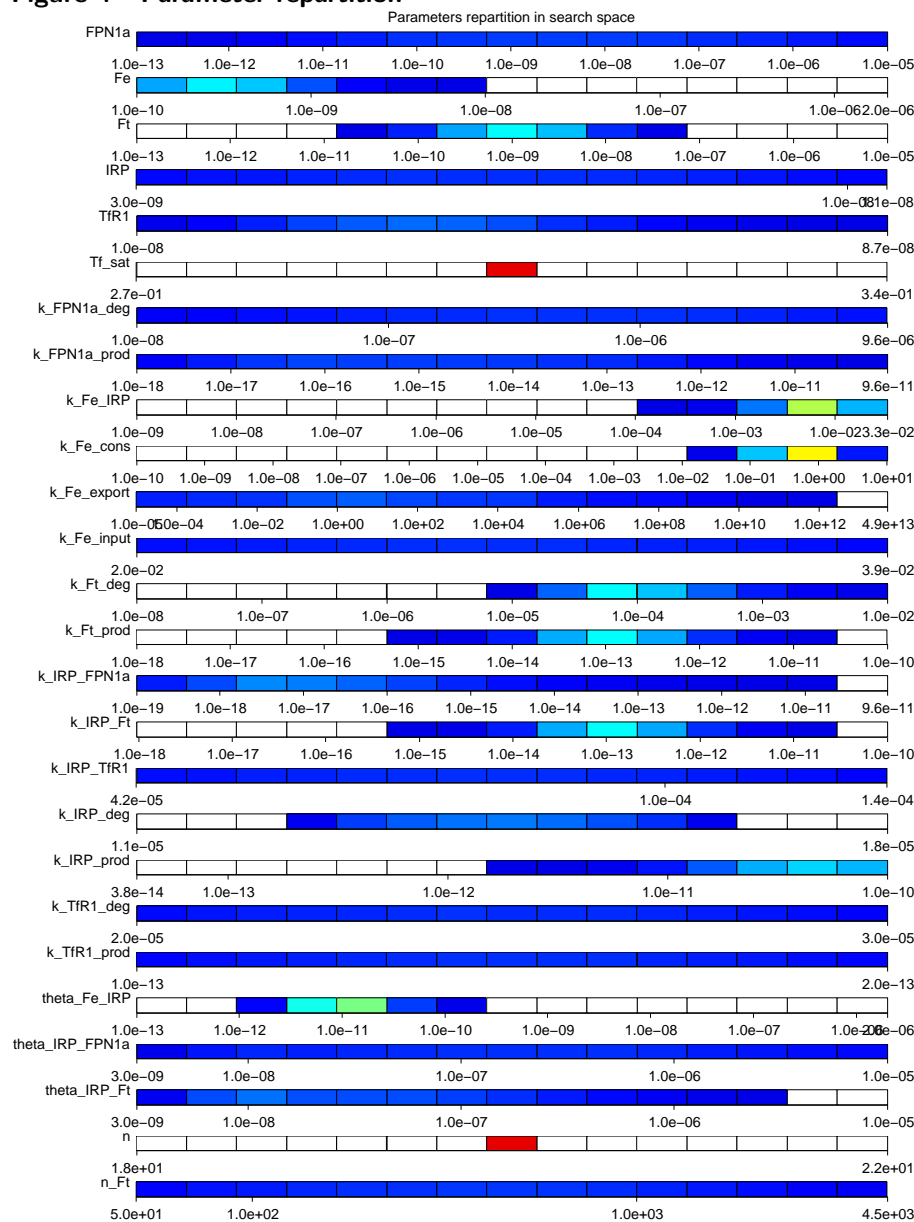
Figure 3 - Simulation of \mathcal{P}_1



Simulation associated to the parameter set \mathcal{P}_1 automatically found by the algorithm introduced in Section 2.5. The variable Tf_{sat}^i , which represents the iron input, is set to 0 at $t=6$ hours to simulate iron

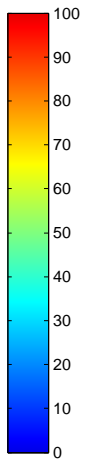
deprivation. The iron concentration consequently decreases, leading to an increase of IRP concentration. This increase leads to a decrease in ferritin (Ft) concentration, keeping the iron level constant. This also leads to an increase in transferrin receptor ($TfR1$) concentration and, later, to a decrease in ferroportin ($FPN1a$) concentration. During the phase when iron is kept constant, the concentration of iron is close to the threshold $\theta_{Fe \rightarrow IRP}$ and IRP concentration is close to the threshold $\theta_{IRP \rightarrow Ft}$. Once the ferritin concentration comes close to zero, the iron concentration also collapses to zero. The x-axis describes the time, in hour, the y-axis describes the concentration of each specie, in $\text{mol} \cdot \text{L}^{-1}$ on a log scale, except for Tf_{sat} which represents the transferrin saturation.

Figure 4 - Parameter repartition



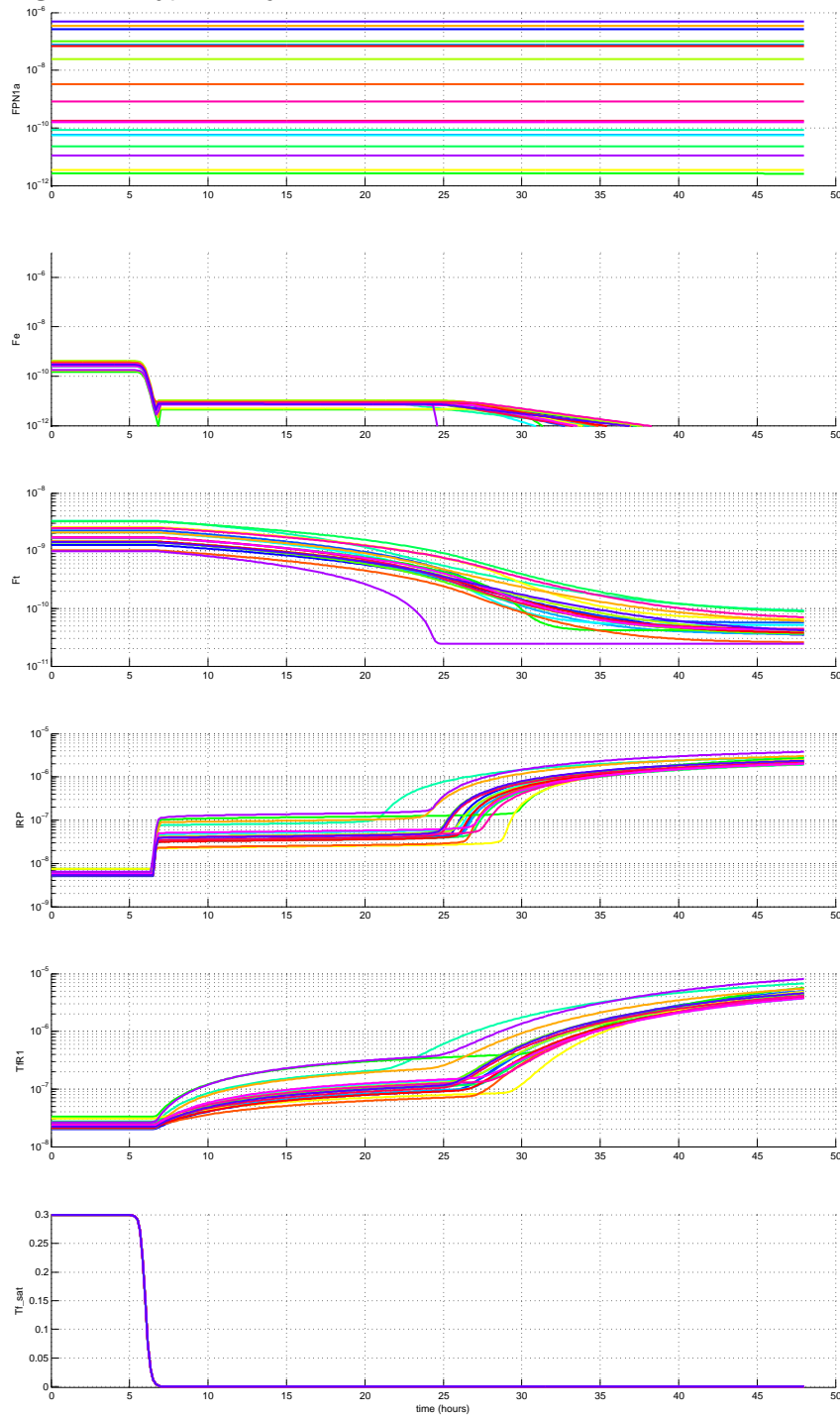
This figure shows the values repartition for each parameter and initial condition, on a logarithmic scale. The unity of the parameters are the same as in Table 3. The color of each interval indicates the percentage of valid parameter sets such that the parameter value lies in this interval. White intervals contain no parameter sets. The percentage corresponding to the other colors is depicted in Figure 5. The parameter n and the initial condition Tf_{sat} are fixed. The parameter $k_{IRP \rightarrow Ft}$ is computed from the parameter k_{IRP_prod} and the five initial conditions $FPN1a$, Fe , Ft , IRP and $TfR1$ are computed as shown by equation (26) to (30).

Figure 5 - Legend of Figure 4



This figure indicates the percentage of valid parameter set in each interval of Figure 4, according to its color. White intervals contain no valid parameter sets.

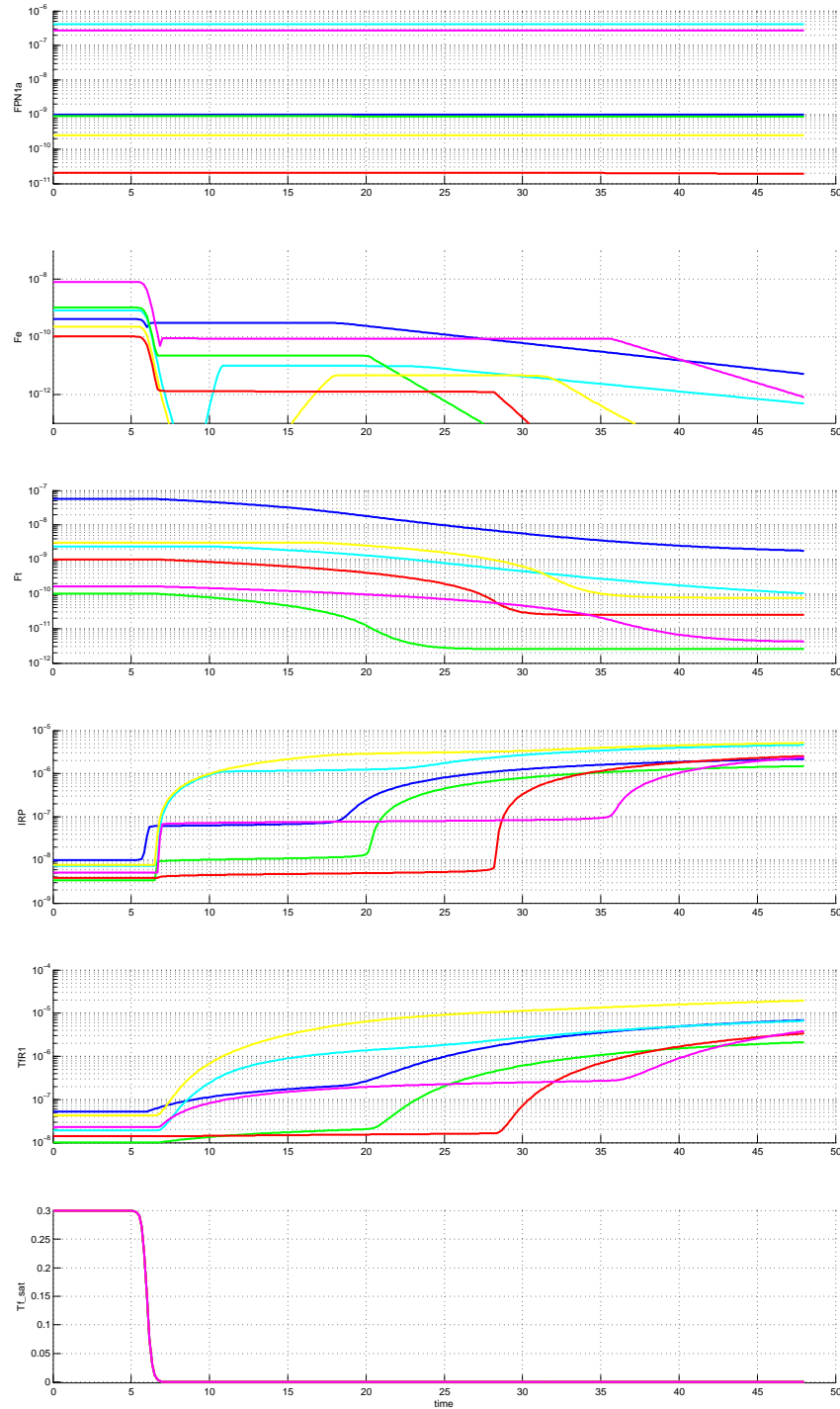
Figure 6 - Typical trajectories



This picture shows the trajectories corresponding to the centers of the 18 ellipsoids found by the method presented in Section 2.6. A color is associated to each parameter set. The x-axis represents the time, in hours, while the y-axis represents the concentration of each species in $\text{mol}\cdot\text{L}^{-1}$ on a log scale, except for Tf_{sat} which represents the transferrin saturation. All plots, except the one for Tf_{sat} are shown on a

logarithmic scale.

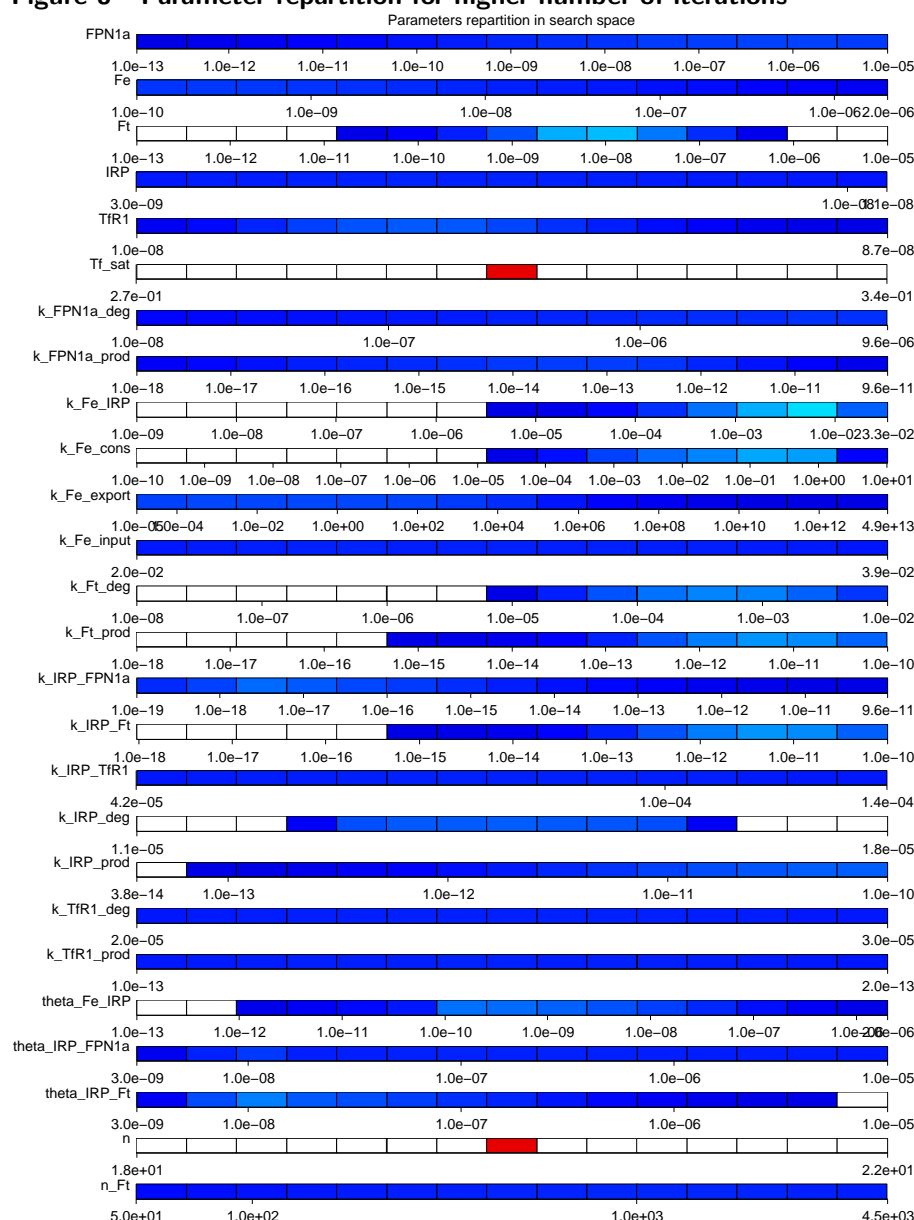
Figure 7 - Heterogeneity of the trajectories



This figure displays some trajectories showing specificity. The x-axis represents the time in hours, while

the y-axis represents the concentration of the species in $\text{mol}\cdot\text{L}^{-1}$ on a log scale, except for Tf_{sat} which represents the transferrin saturation. To each parameter set is associated a specific color. The dark blue curve shows a high plateau concentration, while the red curve shows a low plateau concentration. The green and purple curve have respectively a short and long plateau duration. The cyan and yellow curves show a late start of the plateau after the iron cut-off. The IRP concentration during the plateau between the red and the yellow curves differs by almost two orders of magnitude.

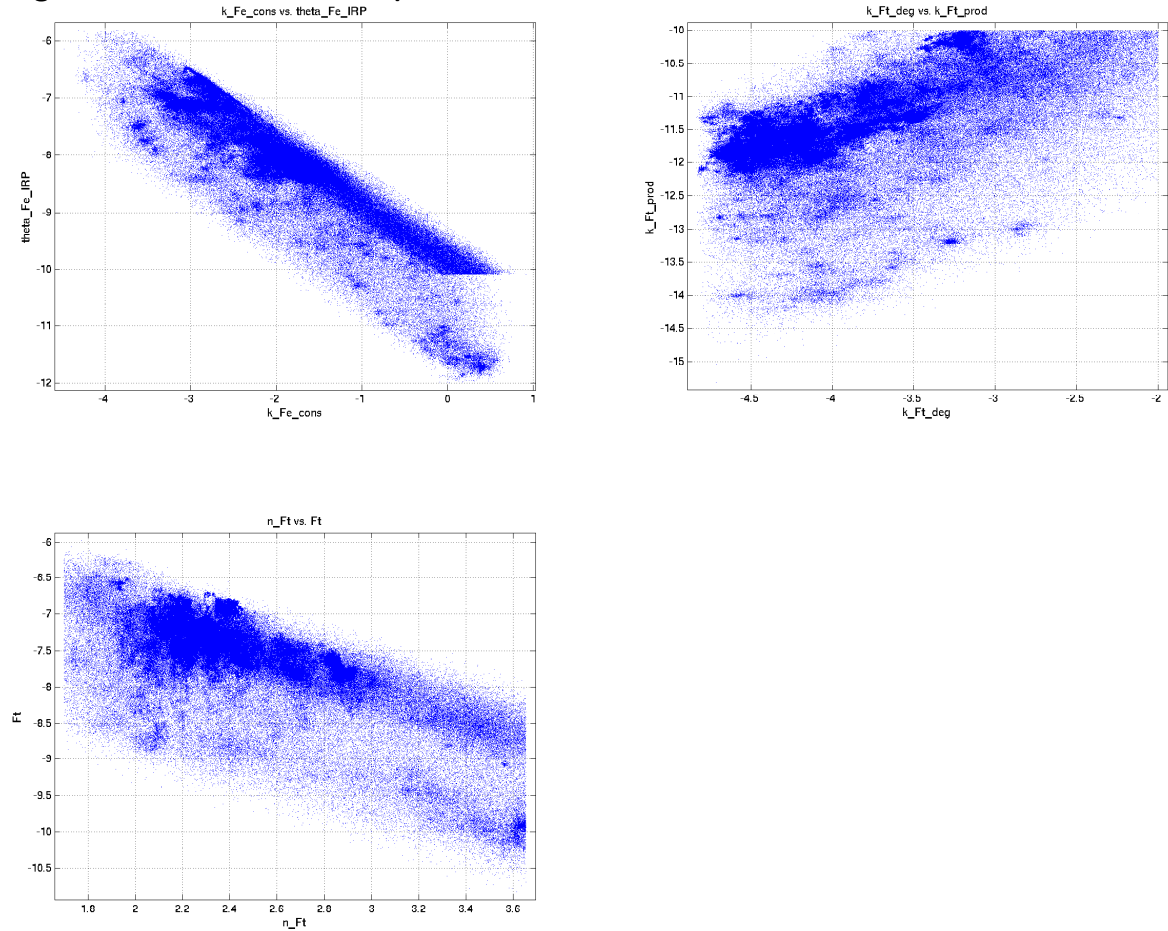
Figure 8 - Parameter repartition for higher number of iterations



This figure indicates the repartition of the valid parameter sets found by the method described in Section 2.6 with a high number of iteration, as explain in Section 3.3.3. The x-axes are on a logarithmic

scale. The percentage of valid parameter sets included in each sub-interval is provided by its color as indicated in Figure 5.

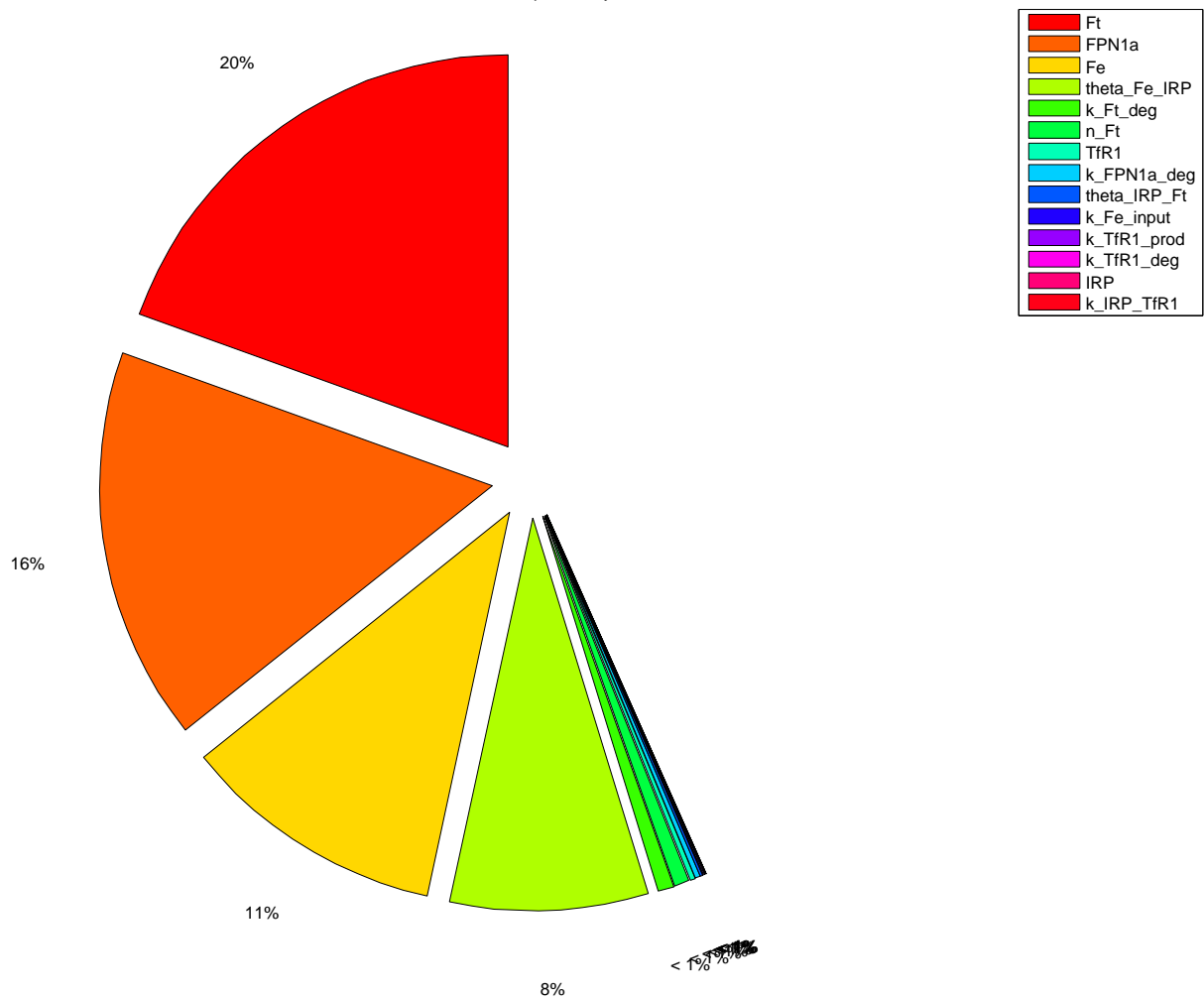
Figure 9 - Correlation between parameters



This figure shows strong correlation between three couple of parameters or initial conditions. The top left graph plots the values of the threshold $\theta_{Fe \rightarrow IRP}$ versus the values of the parameter k_{Fe_cons} . The top right graph plots the values of the parameter k_{Ft_prod} versus the values of k_{Ft_deg} . Finally, the bottom left graph plots the values of n_{Ft} versus the values of the concentration of Ft^{eq} at steady state. Each dot represent a valid parameter set. The plots are on a logarithmic scale.

Figure 10 - Variance of φ_{all} evaluation explained by parameters

Percent of explained variance (decreasingly sorted)
56.7% of variance explained by the model



This pie chart shows the percent of variance of φ_{all} evaluation explained by each of the fourteen parameter of the linear model. All these parameter significantly explain the variance of φ_{all} evaluation. The parameter names in the legend are decreasingly ordered by their influence on the variance. The ten unreadable percentages of explained variance are less than 1%

Tables

Table 1 - Experimental data used to define the parameter search space

The left column shows the data collected in the literature; the right column describes the corresponding interval on parameters and variables, and the deduced equalities and inequalities.

In mouse macrophages, the half-life of FPN1a seems to be more than 20 hours [128]	$k_{FPN1a_deg} \leq 9.6 \times 10^{-6} \text{ s}^{-1}$
The production rate of TfR1 was estimated at 7×10^{-6} pg/cell/min as in [96]	$1.0 \times 10^{-13} \leq k_{TfR1_prod} \leq 2.0 \times 10^{-13} \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$
The half-life of IRP is 12-15 hours in H1299 human lung cancer cells [129]	$1.28 \times 10^{-5} \leq k_{IRP_deg} \leq 1.6 \times 10^{-5} \text{ s}^{-1}$
In the human erythroleukemia cell line K562, the half-life of TfR1 is 8 hours [130]	$2.0 \times 10^{-5} \leq k_{TfR1_deg} \leq 3.0 \times 10^{-5} \text{ s}^{-1}$
The TfR1 production rate enhanced by IRP is 6×10^{-5} pg/cell/min [96]	$4.2 \times 10^{-5} \leq k_{IRP \rightarrow TfR1} \leq 1.44 \times 10^{-4} \text{ s}^{-1}$
In the human erythroleukemia cell line K562, the iron input is up to $7.2 \pm 2.4 \times 10^4$ Fe atoms/cell/min [80]	$2 \times 10^{-2} \leq k_{Fe_input} \leq 3.9 \times 10^{-2} \text{ s}^{-1}$
In mammalian cells, the number of iron atoms in ferritin is up to 4 500 [72]	$n_{Ft} \leq 4.5 \times 10^3$
Human normal saturation level of transferrin is within 25% and 45% [131], we take 30%	$Tf_{sat,0} = 0.3$ (in the iron-replete situation)
In human erythroid cells, a fraction of the iron exporter (FPN1) is not IRP-regulated [70]. The amount of iron needed for biosynthesis (“consumed”) is higher than the exported one.	$k_{Fe_cons} \geq k_{Fe_export} \cdot FPN1a$ (in the iron-replete situation)
A higher limit for the iron pool in erythroleukemic cells is set at $2 \mu\text{mol/L}$ [132]	$Fe \leq 2 \times 10^{-6} \text{ mol} \cdot \text{L}^{-1}$
In the rat liver, the IRP concentration is around 0.11 pmol/mg of cytosolic proteins under normal conditions [133]	$3 \times 10^{-9} \leq IRP \leq 1.07 \times 10^{-8} \text{ mol} \cdot \text{L}^{-1}$ (in the iron-replete situation)
In human erythroleukemic (K562) cells, there are 140 000 transferrin receptors by cell [134]	$1.0 \times 10^{-8} \leq TfR1 \leq 10.0 \times 10^{-8} \text{ mol} \cdot \text{L}^{-1}$ (in the iron-replete situation)

Table 2 - Reduction of search space by the interval solver RealPaver

This table shows the reductions on intervals got from the use of RealPaver on our data as explain in Section 2.4.2. The meaning of parameters is explained in Section 2.1. The superscript “*eq*” on

concentrations means concentrations at the iron-replete steady state.

Deduction	Initial interval
$k_{IRP_prod} \geq 3.84 \times 10^{-14} \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$	$[1.0 \times 10^{-18}, 1.0 \times 10^{-10}]$
$k_{Fe \rightarrow IRP} \leq 3.33 \times 10^{-2} \text{ s}^{-1}$	$[1.0 \times 10^{-9}, 1.0]$
$k_{IRP \rightarrow Ft} \geq 9.5 \times 10^{-19} \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$	$[1.0 \times 10^{-19}, 1.0 \times 10^{-10}]$
$\theta_{IRP \rightarrow FPN1a} \geq 3.0 \times 10^{-9} \text{ mol}\cdot\text{L}^{-1}$	$[1.0 \times 10^{-13}, 1.0 \times 10^{-5}]$
$\theta_{IRP \rightarrow Ft} \geq 3.0 \times 10^{-9} \text{ mol}\cdot\text{L}^{-1}$	$[1.0 \times 10^{-13}, 1.0 \times 10^{-5}]$
$k_{FPN1a_prod} \leq 9.6 \times 10^{-11} \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$	$[1.0 \times 10^{-18}, 1.0 \times 10^{-10}]$
$k_{IRP \rightarrow FPN1a} \leq 9.6 \times 10^{-11} \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$	$[1.0 \times 10^{-19}, 1.0 \times 10^{-10}]$
$k_{Fe_cons} \leq 1.0 \times 10^{+1} \text{ s}^{-1}$	$[1.0 \times 10^{-10}, 1.2 \times 10^{+1}]$
$k_{Fe_export} \leq 4.9 \times 10^{+13} \text{ L}\cdot(\text{mol}\cdot\text{s})^{-1}$	$[1.0 \times 10^{-5}, 1.0 \times 10^{+17}]$
$TfR1^{eq} \leq 8.7 \times 10^{-8} \text{ mol}\cdot\text{L}^{-1}$	$[1.0 \times 10^{-8}, 1.0 \times 10^{-7}]$
$FPN1a^{eq} \geq 1.04 \times 10^{-13} \text{ mol}\cdot\text{L}^{-1}$	$[1.0 \times 10^{-13}, 1.0 \times 10^{-5}]$

Table 3 - Parameter set \mathcal{P}_1

This table shows the parameter values for the valid parameter set \mathcal{P}_1 automatically found by the algorithm introduced in Section 2.5.

$FPN1a^{eq} = 4.66\text{e-}12 \text{ mol}\cdot\text{L}^{-1}$	$k_{Ft_prod} = 1.51\text{e-}13 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$
$Fe^{eq} = 1.21\text{e-}10 \text{ mol}\cdot\text{L}^{-1}$	$k_{IRP \rightarrow FPN1a} = 3.03\text{e-}18 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$
$Ft^{eq} = 7.46\text{e-}9 \text{ mol}\cdot\text{L}^{-1}$	$k_{IRP \rightarrow Ft} = 1.47\text{e-}13 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$
$IRP^{eq} = 3.33\text{e-}9 \text{ mol}\cdot\text{L}^{-1}$	$k_{IRP \rightarrow TfR1} = 6.86\text{e-}5 \text{ s}^{-1}$
$TfR1^{eq} = 1.44\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$	$k_{IRP_deg} = 1.35\text{e-}5 \text{ s}^{-1}$
$Tf_{sat,0} = 0.3 \text{ -}$	$k_{IRP_prod} = 3.72\text{e-}11 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$
$k_{FPN1a_deg} = 8.55\text{e-}7 \text{ s}^{-1}$	$k_{TfR1_deg} = 2.64\text{e-}5 \text{ s}^{-1}$
$k_{FPN1a_prod} = 3.98\text{e-}18 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$	$k_{TfR1_prod} = 1.52\text{e-}13 \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$
$k_{Fe \rightarrow IRP} = 1.12\text{e-}2 \text{ s}^{-1}$	$\theta_{Fe \rightarrow IRP} = 9.70\text{e-}12 \text{ mol}\cdot\text{L}^{-1}$
$k_{Fe_cons} = 8.88\text{e-}1 \text{ s}^{-1}$	$\theta_{IRP \rightarrow FPN1a} = 4.42\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$
$k_{Fe_export} = 2.51\text{e+}9 \text{ L}\cdot(\text{mol}\cdot\text{s})^{-1}$	$\theta_{IRP \rightarrow Ft} = 2.00\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$
$k_{Fe_input} = 2.53\text{e-}2 \text{ s}^{-1}$	$n = 20 \text{ -}$
$k_{Ft_deg} = 2.03\text{e-}5 \text{ s}^{-1}$	$n_{Ft} = 1.02\text{e+}2 \text{ -}$

Table 4 - Ellipsoids characterization

This table describes the volume and the ratio semi-major axis over semi-minor axis for the eighteen ellipsoids automatically found by the algorithm introduced in Section 2.6. The volume and the semi-axis length are computed in a logarithmic space and are intended to be used for comparison. The biggest and the smallest ellipsoids (namely, the third and the seventeenth) are boxed and specifically discussed in the

text in Section 3.3.1. The last line indicates the volume of union of all the ellipsoids. This volume differs from the sum of the individual volume as the ellipsoids intersect.

ellipsoid number	volume	ratio
01	$6.72 \times 10^{+5}$	61.3
02	$1.82 \times 10^{+6}$	48.9
03	$4.43 \times 10^{+6}$	63.4
04	$2.07 \times 10^{+6}$	58.3
05	$4.20 \times 10^{+5}$	56.0
06	$1.67 \times 10^{+6}$	43.6
07	$2.95 \times 10^{+3}$	55.2
08	$1.26 \times 10^{+6}$	60.8
09	$2.27 \times 10^{+6}$	50.2
10	$1.94 \times 10^{+4}$	35.7
11	$1.21 \times 10^{+5}$	74.0
12	$7.56 \times 10^{+5}$	53.2
13	$8.58 \times 10^{+5}$	72.1
14	$2.13 \times 10^{+6}$	49.8
15	$2.25 \times 10^{+6}$	60.7
16	$2.67 \times 10^{+6}$	50.6
17	$8.88 \times 10^{+2}$	88.0
18	$2.12 \times 10^{+5}$	62.3
Union	$7.11 \times 10^{+6} \pm 1.24 \times 10^{+5}$	

4.3 Détails de la méthodologie

La création de la méthodologie présentée dans cet article est issue de l'exploration de différentes pistes pour l'étude du modèle. Certaines de ces pistes ne fournissent pas de résultats utilisables et ne sont donc pas intégrées dans la méthodologie. Nous présentons, dans cette section, ces différentes pistes. De plus, certains points de la méthodologie ont été présentés succinctement dans l'article, et nous revenons dessus dans cette section pour donner plus de détails sur ces points. Cette section est divisée en deux parties, la première (sous-section 4.3.1) est relative à l'état stationnaire, la deuxième (sous-section 4.3.2), à la dynamique du modèle.

Par ailleurs, les résultats de ce travail ont orienté les expérimentations de nos collaborateurs dans l'étude de la concentration en fer régulant l'activité des IRP. Ces expérimentations ont validé nos résultats théoriques et ont permis d'affiner les connaissances sur le système biologique. Ces expérimentations ainsi que les résultats obtenus sont présentés dans la sous-section 4.3.3.

4.3.1 À l'état stationnaire

Cette sous-section débute par une description du fonctionnement du solveur sur intervalles Realpaver plus détaillée que celle faite dans la partie 2.4.2 de l'article. Nous présentons ensuite une recherche visant à augmenter le nombre de contraintes fournies à Realpaver, puis nous décrivons la gestion de contradictions entre les contraintes à l'état stationnaire et les intervalles, cette fonctionnalité n'étant pas décrite dans l'article. Nous finissons par présenter les différentes pistes explorées pour l'obtention d'un jeu de paramètres valide.

Fonctionnement de Realpaver

Realpaver [106] est un solveur de contraintes sur intervalles permettant de résoudre des problèmes de satisfaction de contraintes non-linéaires sur les réels. Un problème est défini par un ensemble d'inconnues, une valeur ou un intervalle (acceptant des bornes infinies) pour chaque inconnue et un ensemble de contraintes non-linéaires. Ces contraintes sont des égalités et des inégalités entre fonctions arithmétiques ayant pour arguments les inconnues du problème. À partir d'un problème, le solveur renvoie un ensemble d'hyper-rectangles dans l'espace des inconnues, nommés « boîtes », tel que l'union de ces boîtes contienne une partie ou l'ensemble des solutions du problème, selon les options indiquées au solveur. Pour calculer cet ensemble de boîtes, le solveur utilise un algorithme *branch and prune* qui consiste à itérativement diviser les boîtes en un ensemble de boîtes plus petites, puis, à réduire ces nouvelles boîtes en supprimant les valeurs violant les contraintes. Une boîte ne contenant pas de solution est supprimée. Un exemple simple de programme est donné par l'équation (4.1).

$$\begin{aligned} a &\in [0.1, 2] & c &\in [1, 10] \\ b &\in [0.1, 2] & a \times b &= c \end{aligned} \tag{4.1}$$

L'utilisation de Realpaver sur ce problème permet d'obtenir les déductions données dans l'équation (4.2).

$$a \in [0.5, 2] \quad b \in [0.5, 2] \quad c \in [1, 4] \tag{4.2}$$

Le comportement du solveur peut être modifié grâce à un ensemble d'options. Celles-ci permettent de spécifier :

- l'obtention d'un pavage ou non. Obtenir un pavage signifie que l'ensemble des solutions du problème est contenu dans l'union des boîtes renvoyées. Dans ce cas, le solveur prouve qu'il n'existe pas de solution hors de l'union des boîtes renvoyées ;

- le nombre maximal de boîtes. Si cette option n'est pas précisée, une valeur par défaut est utilisée. L'utilisation d'un nombre de boîtes plus élevé améliore la précision du résultat, mais demande un temps de calcul plus long ;
- la taille maximale des boîtes. Si un pavage est demandé et que le nombre maximal de boîtes n'est pas compatible avec la taille maximale demandée, alors cette option n'est pas prise en compte. S'il n'est pas demandé de pavage, pour toutes les boîtes renvoyées, la taille des intervalles des inconnues est inférieure à la taille maximale indiquée ;
- l'obtention d'une enveloppe convexe. Cette option, indépendante des trois précédentes, permet d'indiquer au solveur de calculer et renvoyer l'enveloppe convexe¹¹ des boîtes calculées.

La méthodologie que nous avons développée utilise Realpaver en demandant un pavage afin de ne pas perdre de solutions et en spécifiant le nombre de boîtes calculées, permettant de gérer le temps de calcul. L'enveloppe convexe de ce pavage définit l'espace de recherche que nous explorons.

Ajout de contraintes relatives au plateau

Une des premières étapes de notre méthodologie consiste à utiliser les données relatives à l'état stationnaire pour réduire l'espace à explorer, grâce au solveur Realpaver. Par conséquent, l'ajout de contraintes issues de données sur la dynamique peut permettre de réduire davantage les intervalles.

La contrainte imposant un plateau du fer en situation de carence permet un tel ajout de contraintes. Pour l'imposer, nous considérons les concentrations des cinq acteurs *IRP*, *FPN1a*, *TfR1*, *Ft* et *Fe* à dix instants consécutifs et lions les valeurs des concentrations des quatre premiers acteurs par leurs dérivées. Nous imposons ensuite que la dérivée de la concentration du fer soit nulle à ces dix différents instants.

Cette méthode n'a pas permis d'obtenir de meilleure réduction des intervalles et, du fait du plus grand nombre de variables et de contraintes, nécessite un temps de calcul significativement plus long puisque, pour un même nombre de boîtes, le temps de calcul est environ vingt fois plus long. De plus, un des objectifs de ce travail est de créer une méthode automatique réutilisable pour différents modèles. L'ajout de ces contraintes étant très spécifique à ce modèle où la réponse à la carence en fer est modélisée par un plateau où la dérivée du fer est quasiment nulle, nous choisissons de ne pas intégrer ces contraintes au problème fourni à Realpaver.

Gestion de contradictions

Dans le cas où Realpaver révèle une contradiction dans les contraintes à l'état stationnaire et les intervalles des inconnues (i.e. il n'y a pas de solutions au système de contraintes soumis), il est inutile de poursuivre l'étude du modèle. Dans une telle situation, il est intéressant de savoir quelle est la contrainte ou la borne d'intervalle à supprimer pour lever cette contradiction. Pour cela, en considérant que notre problème contient n contraintes, nous créons n problèmes contenant chacun $n - 1$ contraintes. Le i -ème problème ($i \in [1, n]$) contient les n contraintes initiales, excepté la i -ème. Nous utilisons ensuite Realpaver pour trouver l'ensemble des problèmes, parmi ces n , ne présentant pas de contradiction. De plus, Realpaver permet l'utilisation de bornes d'intervalles égales à $+\text{inf}$ ou $-\text{inf}$. Cela nous permet de tester si la suppression de la borne d'un intervalle permet de lever la contradiction.

11. L'enveloppe convexe est définie comme la plus petite boîte contenant toutes les boîtes calculées.

Nous obtenons alors l'ensemble des contraintes et bornes d'intervalles permettant de lever la contradiction. Dans le cas où aucune suppression de contrainte ou de borne d'intervalle ne le permette, nous recherchons l'ensemble de couples de contraintes ou de bornes d'intervalles permettant de la lever.

Recherche d'un jeu de paramètres valide

Dans la méthodologie que nous avons développée, après avoir réduit l'espace des paramètres, nous recherchons un jeu de paramètres valide en utilisant l'algorithme d'optimisation de Nelder-Mead, comme décrit dans la section 2.5 de l'article. Nous présentons ici une tentative d'utilisation plus fine du pavage fourni par le solveur Realpaver. Les résultats obtenus avec cette tentative ne sont pas probants.

En partant du constat que le volume de l'enveloppe convexe d'un pavage est plus grand que la somme des volumes des boîtes constituant ce pavage, nous décidons d'utiliser ce pavage pour définir l'espace au sein duquel l'algorithme d'optimisation recherche un jeu de paramètres valide. L'espace à explorer étant plus petit, nous supposons que l'algorithme d'optimisation sera plus efficace. Pour cela, afin de rechercher uniformément sur les différents ordre de grandeur des paramètres, nous transposons le problème sur une échelle logarithmique. Le programme donné dans l'équation (4.1) se transpose alors dans le programme présenté dans l'équation (4.3).

$$\begin{aligned} a &\in [\ln(0.1), \ln(2)] & c &\in [\ln(1), \ln(10)] \\ b &\in [\ln(0.1), \ln(2)] & e^{(a+b)} &= e^c \end{aligned} \tag{4.3}$$

Ensuite, nous utilisons Realpaver pour obtenir un pavage composé de cent boîtes contenant l'ensemble des solutions potentielles. Puis, au sein de chaque boîte, nous utilisons l'algorithme d'optimisation à partir d'un point de la boîte choisi aléatoirement.

Bien que cette recherche utilise différents jeux de paramètres initiaux pour l'algorithme d'optimisation, et se fasse dans un espace plus petit que celui de l'enveloppe convexe, nous observons qu'aucun jeu de paramètres valide n'est trouvé. Cela peut s'expliquer par le fait que, dans le pavage utilisé, certaines boîtes peuvent ne pas contenir de solution. Par conséquent, contrairement à la recherche au sein de l'enveloppe convexe, l'algorithme d'optimisation ne peut pas se déplacer vers des régions de l'espace contenant les solutions.

4.3.2 Étude de la dynamique

Dans cette sous-section, nous présentons deux optimisations de l'outil Breach qui ont été développées, rendant la recherche d'un jeu de paramètres valide plus efficace. Ces optimisations concernent la représentation interne des formules STL et l'échantillonnage de larges volumes. Ensuite, deux pistes relatives à la recherche et à la caractérisation de la région valide sont présentées.

Optimisation de la définition des formules STL

La recherche d'un jeu de paramètres valide utilise l'évaluation d'une formule STL décrivant le comportement en situation de carence. Cette formule contient de nombreux opérateurs logiques « ET » imbriqués. Nous avons observé que l'évaluation de chaque opérateur nécessite un appel à une fonction C externe qui provoque une fuite de mémoire. Afin d'optimiser la consommation mémoire ainsi que le temps de calcul, nous avons augmenté la grammaire des formules STL en

implémentant un nouvel opérateur, nommé « ETn », qui calcule la conjonction de plusieurs sous-formules. De plus, nous avons implémenté une fonction permettant de remplacer une imbrication de « ET » par un opérateur « ETn », permettant ainsi d’optimiser aisément des formules déjà existantes.

Nous avons comparé les temps de calcul de la vérification de deux formules STL équivalentes, la première étant constituée de cinq « ET » imbriqués, et la deuxième d’un « ETn ». En utilisant le « ETn », nous obtenons une diminution du temps de calcul de 11 %, le gain de temps de calcul augmentant avec le nombre de « ET » imbriqués.

Échantillonnage logarithmique

Une des fonctionnalités de Breach est la création d’un échantillonnage de l’espace des paramètres en utilisant une méthode de génération quasi-Monte Carlo [39]. Cette méthode quasi-aléatoire présente une meilleure convergence que les méthodes pseudo-aléatoires classiques et permet donc de mieux représenter l’ensemble de l’espace à explorer avec un nombre d’échantillons pas trop élevé.

Dans notre méthodologie, lors de la recherche d’un jeu de paramètres valide, nous avons besoin d’un échantillonnage de l’espace des paramètres. L’utilisation de cette méthode paraît appropriée, mais lorsque cet espace à échantillonner couvre plusieurs ordres de grandeur, la méthode présente deux défauts majeurs. En effet, d’une part, des erreurs numériques apparaissent et empêchent une définition précise de l’espace et, d’autre part, les différents ordres de grandeur ne sont pas échantillonnés de manière uniforme : pour un intervalle donné, les ordres de grandeur les plus élevés sont fortement échantillonnés alors que les petits ordres de grandeurs ne sont quasiment pas échantillonnés. Pour remédier à cela, nous créons une fonction permettant d’échantillonner de manière uniforme les différents ordres de grandeur. Elle consiste à convertir sur une échelle logarithmique les intervalles des inconnues, puis à échantillonner quasi-aléatoirement selon les nouveaux intervalles. Il suffit ensuite de convertir l’échantillonnage obtenu sur une échelle unitaire pour obtenir un échantillonnage où tous les ordres de grandeur sont correctement échantillonnés.

Utilisation d’hyper-rectangles

Une fois un jeu de paramètres valide (nommé \mathcal{P}_1) trouvé, nous recherchons une région valide et la caractérisons. Deux manières de représenter cette région valide sont étudiées. La première, qui suppose que l’espace valide est convexe, se base sur l’utilisation d’un hyper-rectangle, et est présentée ici. La seconde, basée sur une union d’ellipsoïdes, est présentée dans la section 2.6 de l’article.

Une fonctionnalité majeure proposée par Breach est le calcul de différentes sensibilités, dont la sensibilité locale de formules STL. Par ailleurs, Breach permet nativement de représenter des hyper-rectangles en haute dimension. Nous utilisons alors le calcul de sensibilité pour rechercher un hyper-rectangle valide. Pour cela, nous créons un algorithme qui définit itérativement un hyper-rectangle valide, que nous nommons HR_V . À chaque itération, une face de HR_V est déplacée de manière à agrandir le volume de l’hyper-rectangle. Le choix de la face qui sera déplacée utilise le calcul de la sensibilité de la formule STL φ_{all} au paramètre normal à chacune des faces. L’algorithme est le suivant :

1. définition de HR_V comme étant l'hyper-rectangle (dégénéré) \mathcal{P}_1 ;
2. définition de l'ensemble F composé de toutes les faces de HR_V ;
3. à chaque face f de F est associée la sensibilité de φ_{all} au paramètre normal à f ;
4. définition d'un ensemble vide F' ;
5. tant que F n'est pas vide :
 - 5.1. la face f ayant la sensibilité la plus grande est extraite de F ;
 - 5.2. faire quatre fois la boucle :
 - 5.2.1. la valeur extrême du paramètre normal à f est modifiée, proportionnellement à sa valeur, augmentant le volume de HR_V ;
 - 5.2.2. recherche de points non valides sur la nouvelle face obtenue ;
 - 5.2.3. si un point non valide est trouvé, la valeur précédente du paramètre est restituée et sortie de la boucle ;
 - 5.3. si, après quatre itérations, aucun point non valide n'est trouvé dans f , alors f est ajoutée à F' ;
6. si F' n'est pas vide :
 - 6.1. F est défini comme F' ;
 - 6.2. à chaque face f de F est associée la moyenne des sensibilités (de φ_{all} au paramètre normal à f) pour un ensemble de jeux de paramètres tirés aléatoirement dans f ;
 - 6.3. retour à l'étape 4.

Cet algorithme, utilisant le jeu de paramètres valide \mathcal{P}_1 , calcule itérativement un hyper-rectangle valide. Un de ses avantages est que la limitation du nombre d'itération de la boucle de l'étape 5 permet de faire tendre HR_V vers un hyper-cube et ainsi évite que l'hyper-rectangle valide obtenu soit (trop) dégénéré. De plus, à chaque itération, la valeur extrême d'un paramètre est modifiée proportionnellement à sa valeur, ce qui est adapté à un hyper-rectangle s'étendant sur plusieurs ordres de grandeur. Par ailleurs, il permet aisément de sauvegarder les hyper-rectangles valides intermédiaires (lors de l'étape 6). Toutefois cet algorithme est mal adapté à l'exploration d'espaces de dimension élevée. En effet, pour que l'étape 5.2.2. puisse vérifier que la nouvelle face ne contient pas de point non valide, un échantillonnage intense est nécessaire, réduisant dramatiquement les performances de l'algorithme. C'est pour cette raison que nous utilisons une représentation basée sur les ellipsoïdes.

Calcul d'intersection des ellipsoïdes

À partir d'une région valide caractérisée par une union d'ellipsoïdes, nous désirons extraire des informations de cette région. Une question naturelle concerne la connexité de ces ellipsoïdes. Pour répondre à cette question, nous avons utilisé deux approches. La première utilise un algorithme de calcul d'intersection et est présentée ici. La deuxième utilise l'échantillonnage uniforme proposé par la méthode de Zamora *et al.* [119] et est présentée dans la section 2.6 de l'article.

L'outil Ellipsoidal Toolbox [135] propose un ensemble de fonctions spécifique à la manipulation d'ellipsoïdes, dont une permettant de calculer si deux ellipsoïdes (non dégénérés) s'intersectent. Cette fonction traduit cette question en un problème quadratique et utilise la programmation quadratique¹² pour le résoudre et ainsi calculer si les ellipsoïdes s'intersectent.

12. La programmation quadratique consiste à résoudre un problème d'optimisation où les fonctions à optimiser sont quadratiques.

L'utilisation de la programmation quadratique paraît adaptée puisque la description mathématique d'un ellipsoïde est une fonction quadratique. Néanmoins, l'application de cette fonction à l'ensemble d'ellipsoïdes décrivant la région valide indique qu'il n'y a aucun couple d'ellipsoïdes qui s'intersectent. Ce résultat est incorrect puisque, lors de la génération de l'échantillonnage uniforme des ellipsoïdes, la méthode de Zamora indique que certains points sont inclus dans deux ellipsoïdes. Par ailleurs, l'utilisation de la programmation quadratique nécessite un temps de calcul conséquent. Ceci nous a mené à modifier la méthode de Zamora afin de récupérer les intersections d'ellipsoïdes obtenues.

4.3.3 Impact de l'étude du modèle

L'utilisation de notre méthodologie a mis en lumière le faible ordre de grandeur du paramètre $\theta_{Fe \rightarrow IRP}$. Cette déduction inattendue a mené nos collaborateurs à investir expérimentalement ce résultat.

Après une revue bibliographique, il est apparu que la concentration en fer conditionnant l'activité régulatrice des IRP (et par conséquent la prolifération des cellules) n'a jamais été évaluée précédemment. Nos collaborateurs ont donc conçu une expérimentation permettant d'estimer cette quantité de fer, requise par la cellule hématopoïétique pour survivre et se diviser. Cette expérimentation a récemment fait l'objet d'une publication [79]. Dans cette publication, la notion de FIC (*Functional Iron Concentration*), correspondant à cette concentration minimale, est introduite. Pour mesurer ce FIC, à partir de cellules carencées en fer (et donc non proliférantes), du fer est introduit dans le milieu de culture des cellules et la quantité de cellules à nouveau proliférantes, ainsi que la concentration d'IRP, sont mesurées. Cette expérimentation a permis de découvrir que cette concentration minimale est de l'ordre du nano-molaire, soit le même ordre que celui de notre seuil. Cet ordre de grandeur est significativement plus faible que ce qui était classiquement considéré comme la concentration minimale nécessaire à la survie d'une cellule (de l'ordre du micro-molaire). Ces mesures ont également été opérées pour des progéniteurs leucémiques issus de patients porteurs de leucémie aiguë myéloïde, afin d'identifier, sur le long terme, des pistes permettant de mieux définir les caractéristiques associées à ces leucémies qui permettent aux blastes¹³ de proliférer.

Ce résultat montre l'importance de l'utilisation de modèles et d'outils d'analyse de ces modèles.

4.4 Comparaison de modèles

L'homéostasie du fer chez les mammifères implique des réseaux de régulation intervenant à différentes échelles, allant de l'échelle intra-cellulaire à l'échelle de l'organisme. Différents modèles, considérant des régulations à différentes échelles, ont été créés [136].

Parmi les modèles se plaçant au niveau de l'organisme entier, se trouve celui créé par Lopes *et al.* [137]. Ce modèle, basé sur un système d'équations différentielles, représente la concentration en fer au sein de 17 localisations chez la souris, ainsi que les échanges de fer entre ces localisations. En utilisant des données de concentration de fer au sein des différentes localisations dans différentes conditions expérimentales, les auteurs ont pu calculer les flux de fer entre les localisations ainsi que le temps de résidence du fer au sein de ces localisations.

13. Un blaste est une cellule indifférenciée qui, dans le cas des leucémies aiguës myéloïdes, se divise de manière incontrôlée.

Certains travaux considèrent, en plus des interactions au niveau de l'organisme, des régulations au niveau intra-cellulaire. Nous pouvons, par exemple, citer les travaux de thèse de F. Achcar [138] qui présente un modèle booléen probabiliste des interactions relatives au fer, au niveau de l'organisme humain. Ce modèle contient 235 acteurs et 420 réactions entre ces acteurs. Les acteurs appartiennent à quatre types cellulaires : les cellules de l'intestin (permettant l'entrée du fer dans l'organisme), les cellules du foie (stockant le fer), les réticulocytes (consommant le fer) et les macrophages (recyclant le fer des réticulocytes). L'application de perturbations sur le modèle (typiquement la suppression de l'expression d'un gène) permet de retrouver les phénotypes des maladies génétiques correspondantes. Ce formalisme est également utilisé pour modéliser le réseau de l'homéostasie du fer chez la levure [14]. Nous pouvons également citer les travaux de Formanowicz, Sackmann *et al.* [15, 139], qui présentent un modèle de l'homéostasie du fer au niveau de l'organisme. Leur modèle, utilisant le formalisme des réseaux de Petri, intègre autant des régulations intra-cellulaires que systémiques. Les auteurs analysent le modèle afin de déterminer les réactions provoquant une variation de la synthèse de l'hepcidine¹⁴, de TfR et de la ferritine, face à une anémie ou à une situation inflammatoire. Plus récemment, Blazewicz *et al.* [140] présentent une nouvelle version de ce modèle, utilisant le formalisme des réseaux de Petri temporisés et testent la capacité du modèle à revenir dans son état normal à partir de n'importe quel état.

D'autres travaux de modélisation se sont focalisés sur des aspects très spécifiques du réseau du fer. Parmi ceux-ci, Becker *et al.* [141] proposent trois modèles de l'interaction du récepteur de l'EPO avec l'EPO¹⁵. Le but de ces modèles, basés sur des systèmes d'EDO non linéaires, est de proposer une description détaillée de la dynamique de cette interaction. Nous pouvons également citer les travaux de Salgado *et al.* [112] qui étudient du stockage du fer au sein des ferritines lors de l'absorption intestinale du fer. Pour cela, ils présentent deux modèles, basés sur des systèmes d'équations différentielles, qui décrivent finement le mécanisme de chargement en fer des ferritines. Cette étude conclut que les ferritines agissent comme un tampon, permettant de maintenir la quantité de fer mobilisable par la cellule constante, malgré la présence de perturbations.

Il existe quelques modèles se concentrant sur l'homéostasie du fer à l'échelle intra-cellulaire, à l'instar du modèle que nous avons établi. Trois modèles, utilisant un formalisme similaire à celui que nous utilisons, considèrent des acteurs proches de ceux que nous avons considérés. Ces modèles sont ceux présentés dans les travaux de Omholt *et al.* [96], Mitchell *et al.* [104] et Chifman *et al.* [103]. Nous présentons de manière plus détaillée ces travaux dans les trois sous-sections suivantes.

4.4.1 Comparaison avec le modèle d'Omholt *et al.*

Les travaux de Omholt *et al.* [96] présentent un « méta-modèle unifié », c'est-à-dire, un modèle qui n'est pas spécifique d'un type cellulaire particulier. Leur modèle prend en compte plus précisément le mécanisme de régulation par les IRP que le nôtre. Les auteurs modélisent explicitement les deux IRP (IRP1 et IRP2), ainsi que les formes actives et non-actives de IRP1. Toutefois, ce modèle n'intègre pas la conversion de ACO1 (forme non-active d'IRP1) en IRP1, alors que cette réaction a été prouvée expérimentalement. En effet, la perte de l'agrégat fer-soufre [4Fe-4S] de ACO1 supprime l'activité aconitase et confère une activité régulatrice à la protéine. Les deux IRP partagent les mêmes cibles, au moins *in vitro*, mais les mécanismes provoquant une variation de l'activité de ces IRP sont différents. Omholt *et al.* modélisent cette différence en

14. L'hepcidine est une molécule inhibant l'export du fer par la ferroportine.

15. L'EPO est une hormone circulante qui impacte l'activité régulatrice des IRP.

considérant différentes expressions analytiques pour la régulation par le fer des IRP. Dans leur modèle, en présence de fer, d'un côté, la dégradation de IRP2 est augmentée, alors que, de l'autre côté, la forme active d'IRP1 n'est plus synthétisée et est convertie en forme inactive. Par ailleurs, contrairement à notre modèle, la ferritine et la ferroportine ne sont pas explicitement prises en compte alors que la traduction de celles-ci est régulée par les IRP et qu'elles sont directement liées à la concentration de fer intra-cellulaire. Une autre différence entre les modèles est la cinétique des réactions : dans le modèle d'Omholt *et al.*, toutes les cinétiques sont sigmoïdales. De plus, l'étude de leur modèle considère uniquement l'état stationnaire en présence de fer. Au contraire de notre méthodologie, ils ne considèrent qu'un seul jeu de paramètres : une valeur numérique est donnée à chaque paramètre, excepté à celui décrivant la raideur des sigmoïdes, et la stabilité de l'unique état stationnaire, ainsi que la concentration du fer, sont étudiées en fonction de la valeur de ce paramètre.

4.4.2 Comparaison avec le modèle de Mitchell *et al.*

Plus récemment, Mitchell *et al.* [104] ont publié un modèle systémique de la régulation du fer dans le foie chez l'humain. Leur modèle en équations différentielles représente de manière précise l'entrée en fer via les récepteurs TfR1 et TfR2, ainsi que les interactions entre ces récepteurs et la protéine HFE (H ainsi que les interactions entre ces récepteurs et la protéine HFE (HFE, pour *human haemochromatosis protein*, est une protéine qui peut se fixer sur les récepteurs de la transferrine, empêchant alors la transferrine de se fixer). Les valeurs prises pour les paramètres de leur modèle sont, soit issues de la littérature, soit estimées lorsqu'aucune donnée n'est disponible. Les auteurs étudient la réponse du modèle dans le cas d'un KO¹⁶ de HFE ou de TfR2 et trouvent que le comportement du modèle correspond qualitativement aux expériences biologiques. Finalement, ils font une analyse de sensibilité de leur modèle pour une variation de $\pm 10\%$ de la valeur des paramètres du modèle. Cette variation paraît faible par rapport aux larges incertitudes de certaines données considérées lors de la modélisation : en pratique, les incertitudes sont plus grandes que 10% . L'approche de Mitchell *et al.* est basée sur l'utilisation d'outils classiques et repose sur l'utilisation d'un seul jeu de paramètres. Comme nous l'avons expliqué dans la section 1.2, l'utilisation d'une telle approche est trop limitée puisqu'elle ne considère pas d'autres jeux de paramètres en accord avec les données. Une analyse plus orientée vers la recherche de l'ensemble des jeux de paramètres valides serait désirable.

4.4.3 Comparaison avec le modèle de Chifman *et al.*

Chifman *et al.* [103] ont publié, en 2012, un modèle de l'homéostasie du fer pour les cellules épithéliales du sein, l'objectif final étant de construire un outil permettant de comprendre le métabolisme du fer chez les cellules saines et les cellules tumorales. Dans ce modèle, la dynamique de cinq acteurs (à savoir la ferritine, la ferroportine, le récepteur de la transferrine, IRP et le fer) est décrite par un système de cinq équations différentielles contenant quinze paramètres. À ces acteurs, s'ajoutent deux entrées qui sont la saturation de la transferrine extra-cellulaire et l'hepcidine. Étant donné qu'il est difficile d'attribuer des valeurs numériques aux paramètres, les auteurs étudient les propriétés du modèle qui sont indépendantes de ces valeurs. Ils prouvent dans un premier temps l'existence d'un état stationnaire unique, puis tentent de savoir s'il est stable. Pour cela, ils génèrent un grand nombre de jeux de paramètres aléatoires et testent la stabilité de l'état stationnaire. Faisant ainsi, ils donnent une bonne idée de la stabilité de cet

16. En biologie, un KO (ou Knock-Out) consiste à inactiver un gène, empêchant ainsi la protéine correspondante d'être produite.

état stationnaire, même s'ils ne le prouvent pas formellement. Les auteurs analysent ensuite l'identifiabilité de leur modèle et montrent qu'un léger changement dans l'écriture des équations modélisant le système permet d'obtenir un modèle où tous les paramètres sont identifiables. Finalement, ils étudient les relations entre les acteurs du modèle et montrent que, à l'état stationnaire, une sur-expression de la ferroportine provoque une baisse de la quantité de fer stocké dans le LIP, une diminution de la concentration de la ferritine et une augmentation de celles de TfR1 et des IRP.

Les acteurs du modèle de Chifman *et al.* et du nôtre sont les mêmes et les graphes d'interaction sont similaires, le signe des arcs communs aux deux graphes d'interaction étant les mêmes (il est à noter que le graphe de la figure 1 du papier de Chifman *et al.* [103] ne respecte pas la définition formelle des graphes d'interaction, qui est utilisée pour obtenir le nôtre). De plus, des choix de cinétiques se basent sur deux hypothèses communes. La première est que le fer contenu dans une ferritine n'est relargué que lorsque celle-ci est dégradée et que les ferritines nouvellement créées se remplissent immédiatement en fer. La deuxième est que la quantité de fer importée dans la cellule (resp. exportée hors de la cellule) est proportionnelle à la concentration du récepteur de la transferrine (resp. de la ferroportine). Par ailleurs, l'étude des états stationnaires est proche. En effet, d'une part, en présence de fer, les deux modèles possèdent un unique état stationnaire qui est stable. Il est à noter toutefois que cette propriété n'est que fortement supposée pour le modèle de Chifman *et al.*, mais est prouvée pour notre modèle. D'autre part, en situation de carence, les deux modèles possèdent un unique état stationnaire.

Toutefois, il existe plusieurs différences entre ces deux modèles. La première principale différence réside dans l'expression analytique de la régulation des IRP sur la ferritine, la ferroportine et le récepteur de la transferrine. Dans le modèle présenté par Chifman *et al.*, la cinétique de cette régulation est hyperbolique pour les ARNm ayant un IRE en 3'-UTR (telle la ferritine), ainsi que pour les ARNm ayant cinq IRE en 5'-UTR (tel le récepteur de la transferrine). Au contraire, dans notre modèle, deux cinétiques différentes sont considérées pour prendre en compte cette différence. La deuxième principale différence entre les deux modèles réside dans le terme décrivant le chargement/déchargement de la ferritine en fer : dans le modèle de Chifman *et al.*, d'une part, il est considéré que chaque ferritine ne contient qu'un seul atome de fer (dans notre modèle, nous décrivons explicitement la quantité d'atomes de fer par ferritine grâce au paramètre n_{Ft}); d'autre part, la vitesse de production de la ferritine est dépendante du fer, ce qui implique l'ajout d'une interaction dans le graphe d'interaction de Chifman *et al.* D'autres différences entre les modèles sont listées ci-dessous :

- la modélisation proposée par Chifman *et al.* ne considère pas de consommation de fer pour les besoins internes de la cellule. Le fer intracellulaire est soit relargué, soit capturé par les ferritines. Notre modèle, au contraire, prend en compte ces besoins internes ;
- Chifman *et al.* ont fait le choix d'une cinétique hyperbolique pour la régulation des IRP par le fer. Dans notre modèle, la cinétique de cette régulation est sigmoïdale ;
- le modèle de Chifman *et al.* considère l'influence de l'hepcidine sur la dégradation de la ferroportine. L'hepcidine étant produite par un autre type cellulaire, ce peptide n'est pas intégré dans notre modèle. La quantité d'hepcidine est *a priori* constante dans les expériences auxquelles notre modèle est appliqué.

Les systèmes d'équations différentielles des deux modèles étant relativement similaires, il est légitime de se demander si le modèle de Chifman *et al.* peut vérifier les contraintes comportementales issues de la littérature. Pour cela, nous considérons le scénario utilisé pour l'étude de notre modèle, à savoir, à partir d'un état stationnaire en présence de fer, nous perturbons le système en coupant l'entrée en fer à $t=6$ heures. Nous appliquons ce scénario au modèle de Chifman *et al.* et étudions sa dynamique.

Nous présentons, ci-dessous, la modélisation à base d'équations différentielles du système qui sera étudiée. Ensuite, nous présentons la méthodologie édifée pour rechercher des concentrations initiales telles que le système soit à l'état stationnaire. Finalement, nous présentons les résultats obtenus pour de la recherche d'un jeu de paramètres tel que le système vérifie les contraintes sur la dynamique.

Système d'équations

Le système d'équations non-linéaires du modèle que nous étudions est décrit par les équations (4.4) à (4.8) où la variable x_1 représente la concentration en fer, x_2 celle du récepteur de la transferrine, x_3 celle de la ferroportine, x_4 celle de ferritine et x_5 l'activité des IRP. Les variables Hep et Fe_{ex} sont les entrées du système et représentent, respectivement, la concentration en hepcidine et le taux de saturation de la transferrine.

$$\frac{dx_1}{dt} = \alpha_1 \cdot Fe_{ex} \cdot x_2 + \gamma_4 \cdot x_4 - \alpha_6 \cdot x_1 \cdot x_3 - \alpha_4 \cdot x_1 \cdot \frac{k_{54}}{k_{54} + x_5} \quad (4.4)$$

$$\frac{dx_2}{dt} = \alpha_2 \cdot \frac{x_5}{k_{52} + x_5} - \gamma_2 \cdot x_2 \quad (4.5)$$

$$\frac{dx_3}{dt} = \alpha_3 \cdot \frac{k_{53}}{k_{53} + x_5} - (\gamma_3 + \gamma_h \cdot Hep) \cdot x_3 \quad (4.6)$$

$$\frac{dx_4}{dt} = \alpha_4 \cdot x_1 \cdot \frac{k_{54}}{k_{54} + x_5} - \gamma_4 \cdot x_4 \quad (4.7)$$

$$\frac{dx_5}{dt} = \alpha_5 \cdot \frac{k_{15}}{k_{15} + x_1} - \gamma_5 \cdot x_5 \quad (4.8)$$

État stationnaire

Notre scénario d'étude commence par une situation où le système est à l'état stationnaire. Le système d'équations différentielles se transforme alors en un système d'équations algébriques qui lie les variables et les paramètres. Afin que ces équations soient respectées dès le début des simulations, nous recherchons des expressions analytiques permettant de calculer les conditions initiales en fonctions des paramètres. La résolution directe du système d'équations algébriques dérivé des équations différentielles (4.4) à (4.8) fournit des expressions analytiques, pour chacune des variables, qui sont trop longues pour être exploitables directement. Nous avons donc reformulé le système afin d'exprimer les variables x_2 à x_5 en fonction de x_1 , selon les équations (4.9).

$$\begin{aligned} x_2 &= \frac{\alpha_2}{\gamma_2 \cdot \left(1 + \frac{k_{52} \cdot \gamma_5 \cdot (k_{15} + x_1)}{\alpha_5 \cdot k_{15}}\right)} & x_3 &= \frac{\alpha_3 \cdot k_{53}}{(\gamma_3 + \gamma_h \cdot Hep) \cdot \left(k_{53} + \frac{\alpha_5 \cdot k_{15}}{\gamma_5 \cdot (k_{15} + x_1)}\right)} \\ x_4 &= \frac{\alpha_4 \cdot k_{54} \cdot x_1}{\gamma_4 \cdot \left(k_{54} + \frac{\alpha_5 \cdot k_{15}}{\gamma_5 \cdot (k_{15} + x_1)}\right)} & x_5 &= \frac{\alpha_5 \cdot k_{15}}{\gamma_5 \cdot (k_{15} + x_1)} \end{aligned} \quad (4.9)$$

Après remplacement des variables x_2 à x_5 , nous obtenons la formule (4.10), qui est du troisième degré selon x_1 , et que nous résolvons en utilisant le solveur symbolique Sage [102].

$$\alpha_1 \cdot Fe_{ex} \cdot \left(\frac{\alpha_2}{\gamma_2 \cdot \left(1 + \frac{k_{52} \cdot \gamma_5 \cdot (k_{15} + x_1)}{\alpha_5 \cdot k_{15}}\right)} \right) - \alpha_6 \cdot x_1 \cdot \left(\frac{\alpha_3 \cdot k_{53}}{(\gamma_3 + \gamma_h \cdot Hep) \cdot \left(k_{53} + \frac{\alpha_5 \cdot k_{15}}{\gamma_5 \cdot (k_{15} + x_1)}\right)} \right) = 0 \quad (4.10)$$

À partir d'une re-formulation de cette équation, Chifman *et al.* ont démontré que parmi les trois solutions de l'équation (4.10), il n'en existe qu'une seule qui soit réelle positive. L'obtention de cette solution nécessite de faire face à deux difficultés majeures :

- il n'est pas possible de déterminer analytiquement laquelle des trois sera réelle positive à cause de la complexité de ces solutions ;
- le calcul numérique des trois solutions, pour un jeu de paramètres donné, peut ne renvoyer que des solutions complexes ou négatives, à cause d'erreurs numériques dues au calcul.

Nous avons donc développé une méthodologie répondant au double objectif de ne pas nécessiter un temps d'exécution trop important et de fournir, pour un jeu de paramètres donné, la condition initiale de x_1 qui soit réelle positive si elle existe et gérant les conditions initiales complexes sinon. Pour cela, nous classifions une condition initiale dans une des trois catégories suivantes :

- « bonne » si elle est réelle positive ;
- « mauvaise » si sa partie réelle est significativement négative ;
- « peut-être » dans les autres cas.

La méthode que nous utilisons est alors la suivante : dès qu'une « bonne » solution est trouvée, elle est gardée. Si aucune « bonne » solution n'est trouvée, mais qu'une « peut-être » est trouvée, la partie réelle de cette dernière est gardée. Si aucune « bonne » solution n'est trouvée et que plusieurs « peut-être » sont trouvées, la solution « peut-être » pour laquelle le membre de gauche de l'équation (4.10) est le plus proche de 0 est gardée. Nous considérons qu'au moins une solution sera « bonne » ou « peut-être », puisque Chifman *et al.* ont démontré qu'il existe toujours une solution positive réelle.

Dynamique

Une fois que nous avons trouvé une méthode permettant d'obtenir un état stationnaire pour n'importe quel jeu de paramètres, nous pouvons étudier la dynamique de ce système suite à une carence en fer. L'objectif de cette étude est de déterminer l'existence d'un jeu de paramètres tel que le système d'équations vérifie les contraintes comportementales. Nous avons utilisé deux approches différentes pour rechercher un tel jeu de paramètres.

La première approche consiste à générer aléatoirement des valeurs pour les paramètres et à vérifier si les simulations correspondantes ont le comportement attendu. Pour générer les valeurs des paramètres, nous avons utilisé les deux méthodes décrites par la section 3.3 du papier de Chifman *et al.* [103], à savoir, soit générées aléatoirement selon une distribution exponentielle de moyenne $\mu = 1\ 000$, soit générées selon une distribution uniforme continue entre 0 et 10 000. Dans les deux cas, la concentration de l'hepcidine a été générée selon une distribution uniforme standard entre 0 et 1, et la saturation de la transferrine a été fixée à 0.3, comme dans notre modélisation. Comme la plupart de ces valeurs de paramètres n'appartiennent pas aux intervalles obtenus à partir des données issues de la littérature, nous ne testons pas les contraintes relatives à l'état stationnaire, mais seulement celles relatives au comportement après la coupure de l'entrée du fer. Nous avons généré 1 000 000 de jeux de paramètres selon chaque méthode de génération et aucune de ces simulations ne vérifie les formules décrivant le comportement. Pour examiner si cette absence de jeu de paramètres valide est due au fait que les intervalles utilisés n'ont pas de signification biologique, nous avons généré aléatoirement les valeurs des paramètres selon a) une distribution continue uniforme et b) une distribution continue uniforme logarithmique à l'intérieur des intervalles calculés à partir des données de la littérature et décrits dans la table 4.1. Nous avons généré 1 000 000 de jeux de paramètres selon chaque méthode, mais nous n'avons pas non plus obtenu de jeu de paramètres vérifiant les formules décrivant le comportement après la mise en carence.

Nom	Intervalle	Unité
x_1	[0.0, 2.0e-6]	mol·L ⁻¹
x_2	[1.0e-8, 1.0e-7]	mol·L ⁻¹
x_3	[1.0e-13, 1.0e-5]	mol·L ⁻¹
x_4	[1.0e-13, 1.0e-5]	mol·L ⁻¹
x_5	[4.5e-9, 1.55e-8]	mol·L ⁻¹
Fe_{ex}	0.3	-
Hep	[1.0e-13, 1.0e-5]	mol·L ⁻¹
α_1	[2.0e-2, 3.9e-2]	s ⁻¹
α_2	[1.0e-12, 2.0e-12]	mol·(L·s) ⁻¹
α_3	[1.0e-15, 1.0e-9]	mol·(L·s) ⁻¹
α_4	[1.0e-9, 1.0e4]	s ⁻¹
α_5	[1.0e-14, 1.0e-12]	mol·(L·s) ⁻¹
α_6	[0.0, 1.0e5]	L·(mol·s) ⁻¹
k_{15}	[0.0, 1.0e-5]	mol·L ⁻¹
k_{52}	[3.3e-8, 1.13e-7]	mol·L ⁻¹
k_{53}	[1.0e-11, 1.0e-4]	mol·L ⁻¹
k_{54}	[1.0e-11, 1.0e-4]	mol·L ⁻¹
γ_2	[2.0e-5, 3.0e-5]	s ⁻¹
γ_3	[0.0, 9.6e-6]	s ⁻¹
γ_4	[1.0e-9, 1.0e-2]	s ⁻¹
γ_5	[1.28e-5, 1.84e-5]	s ⁻¹
γ_h	[1.0e-9, 1.0e-2]	L·(mol·s) ⁻¹

TABLE 4.1 – Ce tableau fournit les intervalles des conditions initiales et des paramètres qui ont été utilisés pour la recherche d’un jeu de paramètres valide.

La deuxième approche consiste à utiliser l’algorithme d’optimisation que nous avons décrit plus haut afin de trouver automatiquement un jeu de paramètres valide. Nous avons tiré 400 jeux de paramètres initiaux selon une distribution uniforme logarithmique et fixé le nombre maximal d’itérations de l’algorithme de recherche à 100. Cette approche n’a pas non plus trouvé de jeu de paramètres valide. Ces résultats n’invalident pas formellement le modèle de Chifman *et al.* [103], mais amènent à avoir quelques doutes concernant sa pertinence. Il semble en tout cas prématuré de qualifier ce modèle de « valide ». Cette méthode, basée sur l’analyse d’un polynôme, sera d’autre part probablement difficile à généraliser à des systèmes de plus grande dimension.

Chapitre 5

Modèle intégrant les ARNm

Le troisième modèle développé est créé pour permettre l'intégration de données supplémentaires obtenues par nos collaborateurs. Pour cela, nous créons un modèle intégrant explicitement les ARN messagers et détaillant finement le mécanisme de régulation par les IRP. Pour étudier ce modèle, nous utilisons la méthodologie présentée précédemment. Lors de son application sur ce modèle, la recherche d'un jeu de paramètres valide initial, en utilisant l'algorithme de recherche proposé par Nelder-Mead, a échoué et nous avons donc dû améliorer cette partie de la méthodologie. Nous développons ici une méthode *ad hoc* permettant de calculer un ensemble de jeux de paramètres respectant les contraintes à l'état stationnaire et d'utiliser ces jeux pour rechercher un jeu de paramètres valide.

Après une introduction du contexte dans lequel le modèle est créé, nous présentons, dans la section 5.2, le système d'équations différentielles utilisé. Nous introduisons ensuite, dans la section 5.3, les données utilisées ainsi que l'utilisation que nous en faisons. Les sections 5.4 et 5.5 présentent l'application de notre méthodologie pour l'étude de l'état stationnaire et la recherche d'une région valide. Ces sections présentent également les problèmes rencontrés lors de l'application de notre méthodologie, ainsi que les solutions apportées.

5.1 Introduction

Dans le modèle exposé précédemment, l'action de la protéine IRP sur les protéines cibles est modélisée empiriquement. Nous avons expliqué, dans le chapitre 2, que la protéine IRP se lie aux ARNm portant un *Iron Responsive Element* (IRE) dans une des régions non codantes de l'ARNm, et que l'effet de cette complexation dépend de la localisation de l'IRE sur l'ARNm. Si l'IRE est dans la région 5'-UTR (*UnTranslated Region*), région située en amont de la région traduite en protéine, la présence d'IRP inhibe la traduction, ayant donc un effet négatif sur la quantité de la protéine correspondante. Si, au contraire, l'IRE est dans la région 3'-UTR, située en aval de la région traduite en protéine, la présence d'IRP stabilise l'ARNm, ayant alors un effet positif sur la quantité de la protéine correspondante.

Nous présentons ici un modèle plus élaboré, décrivant plus finement ces mécanismes de régulation. Trois motivations principales nous ont amenés à la création de ce modèle. La première est d'obtenir un modèle plus réaliste, au prix d'un modèle de plus grande taille, donc computationnellement plus lourd à traiter. La deuxième motivation est que le fait d'introduire explicitement des concentrations d'ARNm en tant que variables dynamiques nous permet d'exploiter des données obtenues par nos collaborateurs (Jean-Marc Moulis, Fiona Louis, Emmanuel Pourcelot). Dans les expériences réalisées, la concentration relative de plusieurs ARNm est mesurée en si-

tuation carencée et en situation contrôle. Seule une partie de ces données pourra être exploitée avec ce nouveau modèle, mais nous pensons que l'augmentation de taille (en nombre de variables dynamiques et nombre de paramètres) sera compensée par le plus grand nombre de contraintes expérimentales utilisables. La troisième motivation est que ce nouveau modèle permettra d'aborder une question portant sur la concentration d'IRP par rapport à la concentration d'ARNm à réguler. En effet, trois situations sont possibles : soit la concentration d'IRP est largement supérieure à la concentration totale d'ARNm à réguler (c'est-à-dire, ceux portant un ou plusieurs IRE), cette hypothèse étant soutenue par des expériences montrant qu'il faut diminuer de manière très importante la concentration d'IRP pour observer un effet [142]. La deuxième possibilité est que la concentration d'IRP est inférieure à la concentration d'ARNm, et, dans ce cas, il n'est pas intuitif de comprendre comment les IRP régulent de manière efficace. La troisième situation est une situation intermédiaire où les concentrations d'IRP et d'ARNm sont du même ordre de grandeur.

De même que pour le modèle précédent, nous étudions ce nouveau modèle en considérant le scénario d'étude décrit dans la sous-section 2.7.1 et modélisé tel qu'indiqué dans la sous-section 3.2.1, à savoir, commencer dans une situation où l'entrée en fer n'est pas nulle et attendre pendant un temps permettant au système de se stabiliser à l'état stationnaire, puis à $t_{cutoff} = 6$ h, mettre à zéro l'entrée Tf_{sat} .

5.2 Présentation du modèle

Nous présentons les hypothèses de modélisation puis les conventions de notation utilisées. Ensuite, nous introduisons le modèle, basé sur un système d'équations différentielles.

5.2.1 Hypothèses de modélisation

Nous présentons, dans cette sous-section, l'ensemble des hypothèses de modélisation qui sont considérées ainsi que des alternatives possibles à ces hypothèses.

- L'ARN messenger du récepteur TfR1 contient cinq IRE en 3'-UTR, donc cinq sites de fixation pour l'IRP. Afin de ne pas multiplier le nombre de variables, nous considérons qu'il n'y a qu'un seul site IRE.
- Les IRE des différents ARNm diffèrent par leur séquence nucléotidique, ce qui implique que la constante de dissociation (nommée kd) des IRP avec les IRE varie d'un ARNm à l'autre. Par ailleurs, la vitesse de dissociation du complexe IRP:IRE dépend de la concentration en fer [143]. Pour considérer précisément ces deux points, une solution consiste à décrire la constante de dissociation avec une expression de la forme $kd(Fe) = kd_0 + kd_1 \cdot Fe$ pour chaque ARN messenger. Toutefois, le paramètre kd_1 est mal connu. De plus, les mesures publiées pour les constantes de dissociation entre IRP1 et différents IRE isolés sont du même ordre de grandeur [144]. Nous choisissons donc de ne pas faire dépendre la constante de dissociation de la concentration en fer et de la considérer unique pour tous les IRE.
- Nous considérons également que la constante d'association (nommée ka) des IRP avec les IRE est unique pour tous les ARN messagers. Cette hypothèse, combinée avec la précédente, implique que la constante d'équilibre $K_d = kd/ka$ est constante.
- Nous considérons que, lorsqu'une protéine IRP est fixée sur un IRE situé en 5'-UTR d'un ARN messenger, la traduction de cet ARN en protéine est complètement bloquée.
- Nous considérons que, pour les ARNm ayant un IRE en 5'-UTR, la constante de vitesse de dégradation spontanée est la même qu'un IRP soit fixé, ou non, sur l'IRE. Cette hypothèse

est soutenue par le fait que la stabilité de l'ARNm de la ferritine n'est pas impactée par l'élévation de la concentration en fer [145].

- Lorsqu'un ARN messenger, sur lequel est fixé un IRP, est dégradé, le devenir de l'IRP n'est pas connu. Soit celui-ci est dégradé, soit il est relâché intact et retourne dans le pool d'IRP non complexés. Il nous semble plus probable qu'il ne soit pas dégradé, mais retourne dans le pool d'IRP non complexés, que l'IRE soit en 5'-UTR ou en 3'-UTR. Une variante pourrait être étudiée dans laquelle l'IRP n'est pas relâché. Il est, en effet, possible que les IRP (ou une partie d'entre eux) restent attachés à des fragments, sans être vraiment dégradés.

Ces choix ne présentent pas de contradiction avec les hypothèses simplificatrices qui fondent le modèle précédent (endocytose non décrite, compartiment mitochondrial non décrit, synthèse de l'hème non représentée, regroupement des deux IRP, etc). Ces hypothèses sont reconduites dans le présent modèle et il ne serait pas justifié de modéliser en détail certains processus particuliers alors que des simplifications importantes sont faites par ailleurs.

5.2.2 Conventions de notation

Dans ce modèle étendu, nous décrivons explicitement les réactions de complexation et dé-complexation de l'IRP et des IRE portés par divers ARN messagers. Par conséquent, nous introduisons des variables décrivant la concentration des ARN libres et complexés de chaque espèce. Nous présentons, dans cette sous-section, les notations utilisées pour représenter ces différents acteurs. Les conventions utilisées pour nommer les paramètres cinétiques sont également décrites :

- Les variables indicées par « $_p$ » représentent les concentrations des **protéines** ;
- Les variables indicées par « $_f$ » représentent les concentrations d'ARNm non liés à un IRP (« **free** » en anglais) ;
- Les variables indicées par « $_b$ » représentent les concentrations d'ARNm complexés avec un IRP (« **bound** » en anglais) ;
- Les vitesses de **production** des ARNm sont notées $p_{...}$, où « ... » est remplacé par le nom de l'ARN messenger ;
- Les constantes de vitesse de la **traduction** des ARNm en protéines sont notées $t_{...}$, où « ... » est remplacé par le nom de l'ARN messenger ;
- Les constantes de vitesse de **dégradation** spontanée des ARNm sont notées $dr_{...}$, où « ... » est remplacé par le nom de l'ARN messenger ;
- Les constantes de vitesse de **dégradation** des ARNm, ayant un IRE en 3'-UTR, stabilisés par des IRP, sont notées $drs_{...}$, où « ... » est remplacé par le nom de l'ARN messenger ;
- Les constantes de vitesse de **dégradation** des **protéines** sont notées $dp_{...}$, où « ... » est remplacé par le nom de la protéine ;
- Les paramètres n_{Ft} , k_{Fe_input} , k_{Fe_export} , k_{Fe_cons} , $k_{Fe \rightarrow IRP}$, $\theta_{Fe \rightarrow IRP}$ et k_{IRP_prod} ont la même signification que dans les modèles précédents ;
- Les paramètres ka et kd représentent, respectivement, les constantes d'association et de dissociation d'un IRP avec un IRE ;
- Lors de calculs, des variables indicées par « $_r$ » apparaissent. Ces variables représentent les sommes des concentrations des ARN libres et complexés avec un IRP.

Le variable Fe a la même signification que la variable ayant le même nom dans le modèle précédent. Les variables Ft_p , $FPN1a_p$ et $TfR1_p$ ont la même signification que, respectivement, les variables Ft , $FPN1a$ et $TfR1$ des deux modèles précédents. La variable IRP de ce modèle représente la concentration d'IRP non complexé et la quantité totale d'IRP est représentée par

la somme des variables Ft_b , $FPN1a_b$, $TfR1_b$, $IRE3_b$, $IRE5_b$ et IRP de ce modèle. Par contre, dans le modèle précédent, la quantité totale d'IRP est simplement représentée par la variable IRP .

5.2.3 Système d'équations différentielles

Le modèle dans son ensemble est donné ici. Dans ces équations, pour chaque terme, une explication est fournie, permettant de comprendre précisément ce que modélisent ces termes. La figure 5.1 représente de manière schématique le réseau qui est modélisé.

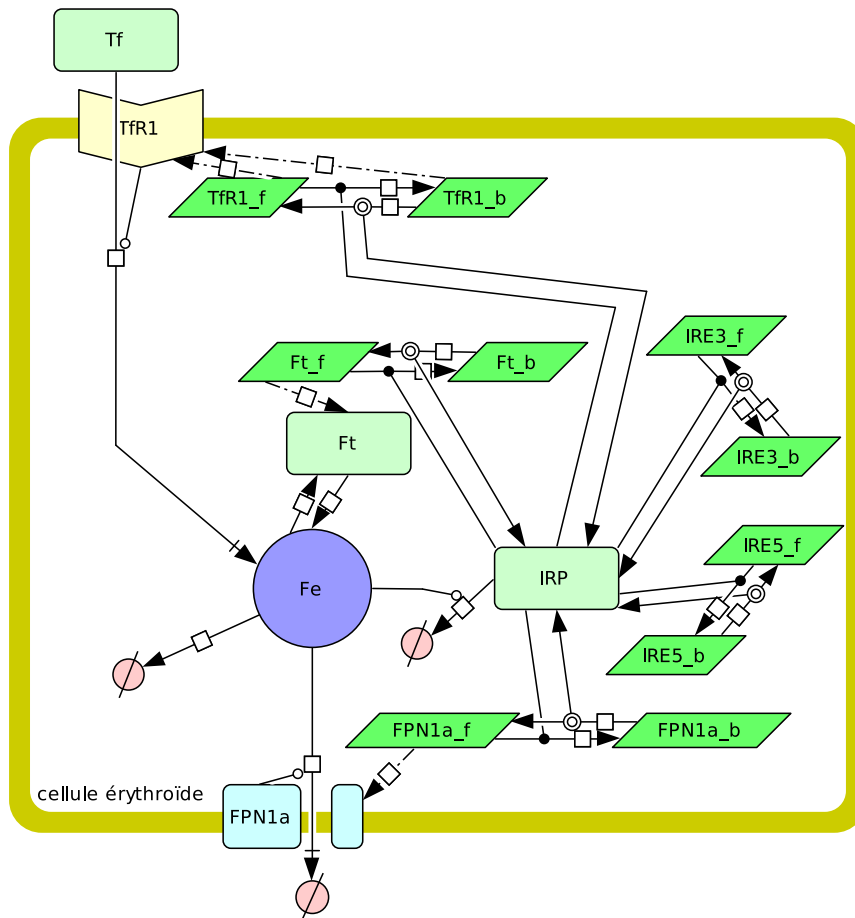


FIGURE 5.1 – Représentation schématique du réseau du fer modélisé dans le chapitre 5. Les parallélogrammes représentent les ARNm messagers. Ceux dont le nom se termine par « $_f$ » représentent les ARNm non complexés avec un IRP, ceux dont le nom se termine par « $_b$ » représentent les ARNm complexés avec un IRP. Les rectangles aux bords arrondis représentent des protéines. L'hexagone concave représente le récepteur de la transferrine. Les flèches contenant un point noir représentent une complexation, celles contenant deux ronds vides imbriqués représentent une décomplexation. Les flèches se terminant par un trait perpendiculaire ainsi qu'une flèche représentent un transport du fer. Les flèches en pointillés représentent la traduction des ARNm messagers. La figure est produite avec le logiciel CellDesigner [55].

Équations de la ferritine et la ferroportine

Nous commençons par présenter les six équations liées à la ferritine et à la ferroportine. Ces deux protéines sont composées de plusieurs sous-unités, dont les ARN messagers contiennent un IRE en 5'-UTR.

Les équations (5.1) et (5.2) décrivent les concentrations de l'ARN messenger des sous-unités de la ferritine, non complexés et complexés avec un IRP. Le premier terme de l'équation (5.1) traduit la synthèse de l'ARN messenger. Les deux termes suivants représentent la complexation et la décomplexation des IRP sur l'IRE de l'ARNm. Ces deux termes se retrouvent, avec un signe contraire, dans l'équation (5.2). Le dernier terme traduit la dégradation spontanée de l'ARNm, proportionnelle à la concentration d'ARNm. Un terme similaire est présent dans l'équation (5.2). La troisième équation (équation (5.3)) décrit la concentration de la protéine ferritine. Le premier terme de cette équation décrit la traduction de l'ARN messenger (non complexé avec un IRP) en protéine. Il contient un facteur $1/24$ puisque la protéine est composée de 24 sous-unités. Le deuxième terme de l'équation traduit la dégradation spontanée de la protéine.

$$\begin{aligned} \frac{d Ft_f}{dt} &= p_{Ft} && \text{transcription} && (5.1) \\ &- ka \cdot Ft_f \cdot IRP && \text{complexation} \\ &+ kd \cdot Ft_b && \text{décomplexation} \\ &- dr_{Ft} \cdot Ft_f && \text{dégradation de l'ARNm} \end{aligned}$$

$$\begin{aligned} \frac{d Ft_b}{dt} &= ka \cdot Ft_f \cdot IRP && \text{complexation} && (5.2) \\ &- kd \cdot Ft_b && \text{décomplexation} \\ &- dr_{Ft} \cdot Ft_b && \text{dégradation de l'ARNm} \end{aligned}$$

$$\begin{aligned} \frac{d Ft_p}{dt} &= \frac{t_{Ft}}{24} \cdot Ft_f && \text{traduction et assemblage} && (5.3) \\ &- dp_{Ft} \cdot Ft_p && \text{dégradation de la protéine} \end{aligned}$$

Les trois équations suivantes sont celles de la ferroportine, de l'ARN messenger de la ferroportine non complexé avec un IRP et de l'ARN messenger de la ferroportine, complexé avec un IRP. Ces trois équations sont similaires aux équations (5.1), (5.2) et (5.3), excepté pour le terme représentant la traduction en protéine qui contient un facteur $1/2$ puisque la ferroportine est un dimère.

$$\begin{aligned} \frac{d FPN1a_f}{dt} &= p_{FPN1a} && \text{transcription} && (5.4) \\ &- ka \cdot FPN1a_f \cdot IRP && \text{complexation} && \\ &+ kd \cdot FPN1a_b && \text{décomplexation} && \\ &- dr_{FPN1a} \cdot FPN1a_f && \text{dégradation de l'ARNm} && \end{aligned}$$

$$\begin{aligned} \frac{d FPN1a_b}{dt} &= ka \cdot FPN1a_f \cdot IRP && \text{complexation} && (5.5) \\ &- kd \cdot FPN1a_b && \text{décomplexation} && \\ &- dr_{FPN1a} \cdot FPN1a_b && \text{dégradation de l'ARNm} && \end{aligned}$$

$$\begin{aligned} \frac{d FPN1a_p}{dt} &= \frac{t_{FPN1a}}{2} \cdot FPN1a_f && \text{traduction et assemblage} && (5.6) \\ &- dp_{FPN1a} \cdot FPN1a_p && \text{dégradation de la protéine} && \end{aligned}$$

Équations du récepteur de la transferrine

Nous présentons ensuite les équations reliées au récepteur de la transferrine. L'ARN messager de ce récepteur comporte cinq IRE situés en 3'-UTR. Comme indiqué dans la sous-section 5.2.1, nous considérons, ici, que ces cinq IRE se comportent comme s'il n'y en avait qu'un seul.

L'équation (5.7), décrivant la concentration de l'ARN messager de Tfr1, non complexé avec un IRP, est similaire aux équations (5.1) et (5.4). L'équation (5.8), décrivant la concentration de l'ARNm de Tfr1, complexé avec un IRP, contient trois termes. Les deux premiers décrivent la complexation et la décomplexation de l'IRE de l'ARNm avec un IRP. Le troisième terme décrit la dégradation de l'ARNm et considère une constante de vitesse de dégradation spécifique. La troisième équation (équation (5.9)) décrit la concentration du récepteur et est composée d'un terme décrivant la synthèse du dimère à partir des ARN messagers (complexés ou non complexés avec un IRP) et d'un terme de dégradation spontanée. Nous supposons que la vitesse de traduction est la même, que l'ARNm soit complexé ou non avec un IRP.

$$\begin{aligned} \frac{d Tfr1_f}{dt} &= p_{Tfr1} && \text{transcription} && (5.7) \\ &- ka \cdot Tfr1_f \cdot IRP && \text{complexation} && \\ &+ kd \cdot Tfr1_b && \text{décomplexation} && \\ &- dr_{Tfr1} \cdot Tfr1_f && \text{dégradation de l'ARNm} && \end{aligned}$$

$$\begin{aligned} \frac{d Tfr1_b}{dt} &= ka \cdot Tfr1_f \cdot IRP && \text{complexation} && (5.8) \\ &- kd \cdot Tfr1_b && \text{décomplexation} && \\ &- drs_{Tfr1} \cdot Tfr1_b && \text{dégradation de l'ARN stabilisé} && \end{aligned}$$

$$\begin{aligned} \frac{d Tfr1_p}{dt} &= \frac{t_{Tfr1}}{2} \cdot (Tfr1_f + Tfr1_b) && \text{traduction et assemblage} && (5.9) \\ &- dp_{Tfr1} \cdot Tfr1_p && \text{dégradation de la protéine} && \end{aligned}$$

Équations des autres ARN messagers

Puisque nous décrivons explicitement la complexation de la protéine IRP avec les IRE, nous devons prendre en compte tous les ARN messagers ayant un IRE puisque ceux-ci captent une partie des IRP. Pour représenter ces ARN messagers, nous introduisons pour cela deux paires de variables : ($IRE5_f$ et $IRE5_b$) et ($IRE3_f$ et $IRE3_b$). La première paire représente collectivement tous les ARNm portant un IRE en 5'-UTR autre que ceux de la ferritine et de la ferroportine. De même, la deuxième paire représente les ARNm portant un IRE en 3'-UTR autre que celui de TfR1. Les deux équations (5.10) et (5.11) sont identiques à celles des ARNm de la ferritine et de la ferroportine, et les équations (5.12) et (5.13), à celles de l'ARNm de TfR1. Nous n'avons pas besoin d'introduire les équations correspondantes aux protéines codées par ces ARNm, ces protéines ne jouant aucun rôle dans le système d'intérêt.

$$\begin{aligned} \frac{d IRE5_f}{dt} &= p_{IRE5} && \text{transcription} && (5.10) \\ &- k_a \cdot IRP \cdot IRE5_f && \text{complexation} \\ &+ k_d \cdot IRE5_b && \text{décomplexation} \\ &- dr_{IRE5} \cdot IRE5_f && \text{dégradation de l'ARNm} \end{aligned}$$

$$\begin{aligned} \frac{d IRE5_b}{dt} &= k_a \cdot IRP \cdot IRE5_f && \text{complexation} && (5.11) \\ &- k_d \cdot IRE5_b && \text{décomplexation} \\ &- dr_{IRE5} \cdot IRE5_b && \text{dégradation de l'ARNm} \end{aligned}$$

$$\begin{aligned} \frac{d IRE3_f}{dt} &= p_{IRE3} && \text{transcription} && (5.12) \\ &- k_a \cdot IRE3_f \cdot IRP && \text{complexation} \\ &+ k_d \cdot IRE3_b && \text{décomplexation} \\ &- dr_{IRE3} \cdot IRE3_f && \text{dégradation de l'ARNm} \end{aligned}$$

$$\begin{aligned} \frac{d IRE3_b}{dt} &= k_a \cdot IRE3_f \cdot IRP && \text{complexation} && (5.13) \\ &- k_d \cdot IRE3_b && \text{décomplexation} \\ &- dr_{s_{IRE3}} \cdot IRE3_b && \text{dégradation de l'ARNm stabilisé} \end{aligned}$$

Équation des IRP

L'équation différentielle (5.14) décrit la dynamique du pool d'IRP non complexés avec un IRE. La première et les deux dernières lignes traduisent la production et la dégradation, spontanée et régulée par le fer, des IRP. Le deuxième ligne traduit la fixation d'IRP sur les IRE des différents ARN messagers. La troisième ligne traduit l'apport d'IRP par la décomplexation des complexes IRE:IRP. Les quatrième et cinquième lignes traduisent la libération d'IRP suite à la dégradation des ARNm complexés avec un IRP.

$$\begin{aligned}
\frac{d IRP}{dt} = & k_{IRP_prod} && \text{production} && (5.14) \\
& - ka \cdot (Ft_{-f} + FPN1a_{-f} + TfR1_{-f} + IRE3_{-f} + IRE5_{-f}) \cdot IRP && \text{complexation} \\
& + kd \cdot (Ft_{-b} + FPN1a_{-b} + TfR1_{-b} + IRE3_{-b} + IRE5_{-b}) && \text{décomplexation} \\
& + dr_{-Ft} \cdot Ft_{-b} + dr_{-FPN1a} \cdot FPN1a_{-b} + dr_{-IRE5} \cdot IRE5_{-b} && \text{dégradation des} \\
& + drs_{-TfR1} \cdot TfR1_{-b} + drs_{-IRE3} \cdot IRE3_{-b} && \text{ARNm complexés} \\
& - k_{Fe \rightarrow IRP} \cdot \text{sig}^+(Fe, \theta_{Fe \rightarrow IRP}) \cdot IRP && \text{régulation par le fer} \\
& - dp_{-IRP} \cdot IRP && \text{dégradation}
\end{aligned}$$

Équation du fer

L'équation différentielle régissant la concentration de fer comporte, classiquement, un terme traduisant l'entrée en fer au sein de la cellule, un terme traduisant l'utilisation et le relargage de fer par les ferritines, un terme traduisant l'export de fer par les ferroportines et un terme traduisant la consommation pour les besoins internes de la cellule.

$$\begin{aligned}
\frac{d Fe}{dt} = & k_{Fe_input} \cdot TfR1_{-p} \cdot Tf_{sat} && \text{entrée du fer} && (5.15) \\
& - n_{Ft} \cdot \left(\frac{d Ft_{-p}}{dt} \right) && \text{variation due à } Ft_{-p} \\
& - k_{Fe_export} \cdot Fe \cdot FPN1a_{-p} && \text{export du fer} \\
& - k_{Fe_cons} \cdot Fe && \text{consommation interne}
\end{aligned}$$

5.3 Modélisation des données

Dans ce modèle, la protéine IRP a les trois mêmes cibles que dans le modèle précédent. Toutefois, la régulation de la traduction des ARNm ayant un IRE en 5'-UTR n'est plus décrite par une sigmoïde, comme dans le modèle précédent, mais est décrite par un changement du rapport entre ARNm non complexé et ARNm complexé avec un IRP. Ceci concerne les protéines Ft (ferritine) et FPN1a (ferroportine). De même, la production de TfR1 ne dépend plus linéairement de la concentration en IRP. Le modèle que nous présentons ne contient plus qu'une seule sigmoïde, qui est associée à l'effet du fer sur IRP. Nous faisons alors une revue des données disponibles utilisables avec ce nouveau modèle.

Comme pour le modèle précédent, nous pouvons distinguer les données statiques, où le temps n'intervient pas, et les données temporelles (ou comportementales), liées ici à la réponse des cellules à la perturbation qu'est la mise en carence en fer. Dans la première catégorie, nous avons, d'une part, les intervalles de définition des paramètres et des concentrations à l'état stationnaire, d'autre part, les inégalités faisant intervenir des expressions ne contenant que des paramètres et des concentrations à l'état stationnaire. Dans la deuxième catégorie, nous avons quelques données expérimentales provenant de mesures effectuées 24 heures et 48 heures après la mise en carence.

Nous donnons plus de détails sur les intervalles et sur les données comportementales dans les sous-sections 5.3.1 et 5.3.2. L'ensemble des contraintes portant sur l'état stationnaire sont présentées dans la section A.3. Nous présentons ensuite trois réécritures du système d'équations

différentielles permettant d'intégrer plus facilement des données (sous-section 5.3.3). Deux de ces réécritures seront appliquées, la troisième n'étant pas conservée car elle complique trop le modèle. Une réécriture des contraintes à l'état stationnaire est donnée dans la section A.4.

5.3.1 Intervalles

Ce nouveau modèle comporte des paramètres et des variables qui n'apparaissent pas dans le modèle précédent. Ces paramètres sont les constantes de vitesse d'association et de dissociation associées à la complexation/décomplexation des IRP avec les IRE, les constantes de vitesses de transcription et de dégradation des ARNm, ainsi que les paramètres décrivant la traduction.

Il est donc nécessaire d'utiliser de nouvelles données pour déterminer des intervalles pour ces paramètres. Par exemple, l'article de Sharova *et al.* [146] présente les taux de dégradation de nombreux ARNm dans les cellules souches de la souris. Nous pouvons y trouver les taux de dégradation des ARNm de la ferritine et du récepteur de la transferrine. Nous pouvons aussi citer la constante de vitesse de traduction de l'ARNm de FPN1a, pour laquelle nous utilisons les résultats du modèle, décrivant la vitesse de traduction, proposée par Reuveni *et al.* [147]. Au vu de l'hétérogénéité des données, nous imposons une contrainte sur la taille des intervalles : le rapport de la borne supérieure de l'intervalle sur la borne inférieure de l'intervalle doit être supérieur ou égal à 10. L'ensemble des données utilisées ainsi que les intervalles déduits pour les paramètres et les conditions initiales sont présentés dans les sections A.1 et A.2.

Après l'utilisation des données, nous observons que les intervalles des paramètres et des conditions initiales peuvent se classer en trois catégories, selon la taille des intervalles. Neuf intervalles s'étalent sur un ordre de grandeur et sont considérés comme étroits, quatorze intervalles s'étalent sur 2 à 5 ordres de grandeur et dix-neuf intervalles s'étalent sur 6 à 8 ordres de grandeur. Finalement, deux intervalles s'étendent sur 16 ordres de grandeur.

5.3.2 Contraintes sur la dynamique

Des données expérimentales, relatives à la concentration du fer après la mise en carence, ont été obtenues par nos collaborateurs. En conséquence, nous n'utilisons plus l'hypothèse qui consiste à imposer l'existence d'un palier sur le fer pendant une certaine durée.

Les données dont nous disposons, sur le comportement du système après la mise en carence en fer, sont les suivantes :

- Chez des cellules K562, pendant 48 heures après une mise en carence, la quantité totale de fer reste supérieure à 16 % de la quantité avant la mise en carence. Étant donné que nos simulations durent 48 heures, nous traduisons cette donnée par la formule STL φ_{B1} imposant qu'après l'application de la perturbation, la quantité totale de fer (correspondant à la quantité de fer décrite par la variable Fe à laquelle s'ajoute la quantité de fer contenu dans les ferritines) soit toujours supérieure à 16 % de la quantité totale de fer à $t=4$ h¹⁷ :

$$\varphi_{B1} = \text{alw_}[6*3600, \text{inf}] (\\ (\text{Fe}[t] + n_Ft * Ft_p[t]) > 0.16 * (\text{Fe}[4*3600] + n_Ft * Ft_p[4*3600]))$$

- Chez des cellules K562, 24 heures après la mise en carence, la quantité totale du fer est d'environ 25 % de la quantité avant la mise en carence. Cette donnée se traduit par la formule STL φ_{B2} :

$$\varphi_{B2} = \text{Fe}[(6+24)*3600] + n_Ft * Ft_p[(6+24)*3600] \\ < 0.25 * (\text{Fe}[4*3600] + n_Ft * Ft_p[4*3600])$$

17. La perturbation étant appliquée à $t=6$ h, nous considérons que l'instant $t=4$ h est représentatif de la valeur à l'état stationnaire.

- 24 heures après la coupure de l'entrée en fer, la concentration de l'ARNm de TfR1 est multipliée par un facteur compris entre 3 et 4 [148, 149]. Ceci revient à considérer un intervalle pour le ratio concentration d'ARNm de TfR1 à l'état stationnaire sur concentration d'ARNm de TfR1 24 heures après la mise en carence. Afin d'éviter un intervalle trop petit, et donc trop difficile à respecter, nous considérons un rapport dix entre la borne supérieure et la borne inférieure de l'intervalle. Les formules STL $\varphi_{B3.1}$ et $\varphi_{B3.2}$ traduisent cette donnée :

$$\varphi_{B3.1} = \frac{(\text{TfR1_f}[(6+24)*3600] + \text{TfR1_b}[(6+24)*3600])}{(\text{TfR1_f}[4*3600] + \text{TfR1_b}[4*3600])} > 1.2$$

$$\varphi_{B3.2} = \frac{(\text{TfR1_f}[(6+24)*3600] + \text{TfR1_b}[(6+24)*3600])}{(\text{TfR1_f}[4*3600] + \text{TfR1_b}[4*3600])} < 12$$

- Chez des cellules de la lignée cellulaire KG1, 24 heures après la mise en carence en fer, la ferritine est en concentration au moins trois fois plus faible qu'à l'état stationnaire. Nous traduisons cette donnée par la formule STL φ_{B4} :

$$\varphi_{B4} = \frac{\text{Ft}[(6+24)*3600]}{\text{Ft}[4*3600]} < 3.0$$

5.3.3 Réécriture

Il peut arriver qu'aucune donnée précise sur deux paramètres donnés soit disponible, mais qu'une telle donnée existe sur le rapport de ces deux paramètres. Cette situation est alors modélisée par un intervalle étroit sur le rapport de ces deux paramètres et deux intervalles larges pour les deux paramètres. Lors de la recherche d'un jeu de paramètres valide, deux manières d'utiliser cette information sont possibles. La première consiste à choisir une valeur pour les deux paramètres et vérifier que le ratio de ces deux paramètres est dans l'intervalle étroit. La deuxième consiste à choisir une valeur pour un paramètre ainsi qu'une valeur de l'intervalle du rapport des deux paramètres et à vérifier que le produit de ces deux valeurs est dans l'intervalle large du second paramètre. Cette deuxième manière est plus efficace, mais nécessite une réécriture des équations. Nous présentons ici trois réécritures permettant de prendre en compte de telles situations.

Très peu de données sont disponibles sur les constantes d'association (ka) et de dissociation (kd) du complexe IRE:IRP. Toutefois, une mesure de la constante $K_d = kd/ka$ est disponible. Pour intégrer aisément cette donnée, nous introduisons cette constante et remplaçons le paramètre kd par l'expression ($ka \cdot K_d$) dans les équations concernées.

Les expériences de nos collaborateurs indiquent, qu'en régime stationnaire en présence de fer, la concentration de ARNm de la ferritine est 20 à 200 fois supérieure à celle de l'ARNm de la ferroportine. À l'état stationnaire, la somme des équations (5.1) et (5.2) permet d'obtenir que la concentration de ferritine à l'état stationnaire est égale à $\frac{p_{-Ft}}{dr_{-Ft}}$. De même, en sommant les équations (5.4) et (5.5), nous obtenons que la concentration de la ferroportine à l'état stationnaire est égale à $\frac{p_{-FPN1a}}{dr_{-FPN1a}}$. Par conséquent, nous avons la relation $\frac{p_{-Ft}}{dr_{-Ft}} = [20, 200] \cdot \frac{p_{-FPN1a}}{dr_{-FPN1a}}$. Pour intégrer explicitement cette donnée dans les équations, nous remplaçons le paramètre p_{-FPN1a} par l'expression $\frac{p_{-Ft} \cdot dr_{-FPN1a}}{dr_{-Ft} \cdot v_{-FPN1a}}$, où le nouveau paramètre v_{-FPN1a} a une valeur comprise entre 20 et 200. L'équation (5.4) de la ferroportine se réécrit alors :

$$\frac{d \text{FPN1a}_f}{dt} = \frac{p_{-Ft} \cdot dr_{-FPN1a}}{dr_{-Ft} \cdot v_{-FPN1a}} - ka \cdot \text{FPN1a}_f \cdot \text{IRP} + (ka \cdot K_d) \cdot \text{FPN1a}_b - dr_{-FPN1a} \cdot \text{FPN1a}_f \quad (5.16)$$

Trois autres mesures de concentrations relatives d'ARN messenger, à l'équilibre, sont connues. Ces mesures fournissent les ratios $Ft_r/TfR1_r$, $FPN1a_r/TfR1_r$ et $TfR1_b/TfR1_f$. Toutefois,

nous ne pouvons pas appliquer la même méthode pour exprimer ces données. En effet, il n'est pas possible d'obtenir une expression décrivant $TfR1_{-r}$ à l'équilibre en fonction des paramètres de $TfR1$. Une méthode permettant de dépasser ce problème est de créer des nouvelles variables qui représentent ces rapports. Par exemple, l'équation décrivant le rapport $Ft_{-r}/TfR1_{-r}$ est donnée dans l'équation (5.17). Cette nouvelle équation implique de supprimer l'équation (5.2) et de remplacer la variable Ft_{-b} par l'expression « $x_{-Ft_{-b}} \cdot (TfR1_{-b} + TfR1_{-f}) - Ft_{-f}$ » dans les équations (5.1) et (5.14).

$$\frac{d x_{-Ft_{-b}}}{d t} = \frac{p_{-Ft} - dr_{-Ft} \cdot Ft_{-f} - x_{-Ft_{-b}} \cdot (p_{-TfR1} - dr_{-TfR1} \cdot TfR1_{-f} - drs_{-TfR1} \cdot TfR1_{-b})}{TfR1_{-f} + TfR1_{-b}} \quad (5.17)$$

Les équations de ces nouvelles variables sont complexes et l'utilisation de ces nouvelles variables rend la compréhension du modèle plus difficile. Nous décidons de ne pas appliquer cette réécriture.

5.4 Étude de l'état stationnaire

Pour étudier ce modèle à l'état stationnaire, nous suivons la méthodologie que nous avons développée. Celle-ci consiste à procéder à une simplification des équations, puis à exprimer les concentrations à l'état stationnaire en fonction des paramètres du système d'équations, et à tester l'existence d'un unique état stationnaire stable en présence de fer. Nous utilisons ensuite le solveur sur intervalle Realpaver pour réduire l'espace de recherche.

5.4.1 Réduction de modèle

Notre scénario d'étude du modèle commence par une situation où du fer est présent. Cette situation est modélisée par un état stationnaire où la concentration en fer est supérieure au seuil de la régulation du fer sur l'expression des IRP. Nous pouvons donc remplacer la sigmoïde de l'équation (5.14) par la valeur 1, simplifiant ainsi l'équation.

Il est aisé de remarquer qu'en faisant la somme des deux équations différentielles associées aux ARNm libres et liés de la ferritine, nous obtenons une équation affine indépendante de IRP, les termes de complexation et décomplexation s'éliminant (cette réflexion s'applique également aux équations relatives à la ferroportine et à *IRE5*). La somme de ces deux concentrations converge vers un état stationnaire stable indépendant de IRP. Une fois cet état stationnaire atteint, la coupure de l'entrée en fer ne provoque pas de changement sur la somme $Ft_{-b} + Ft_{-f} = Ft_{-r}$, qui reste égale à $Ft_{-r}^{eq} = p_{-Ft}/dr_{-Ft}$. Il est alors tentant d'éliminer une variable, par exemple Ft_{-b} , en la remplaçant par l'expression $Ft_{-b}(t) = Ft_{-r}^{eq} - Ft_{-f}(t)$. Cela présente cependant un risque car les équations algébriques que nous utilisons pour calculer les concentrations à l'état stationnaire sont basées sur l'approximation des sigmoïdes à 0 ou 1. Cela n'étant pas strictement vrai, nous laissons du temps en début de simulation au système pour converger vers l'état stationnaire. En d'autres termes, la relation ci-dessus donnant $Ft_{-b}(t)$ n'est pas valide en début de simulation, et cela est d'autant plus marqué que l'exposant n dans la sigmoïde est petit (proche de 1). Pour éviter ce biais, nous ne faisons pas cette simplification.

5.4.2 Calcul analytique des concentrations

Avant de faire des simulations pour étudier la dynamique, nous désirons obtenir les concentrations à l'état stationnaire en fonction des paramètres. Pour cela, nous exprimons les concen-

trations de chaque entité en fonction de la concentration des IRP, ceux-ci étant centraux dans le système. Nous utilisons le solveur symbolique Sage [102] pour effectuer les calculs.

Nous exprimons, dans un premier temps, les concentrations des ARNm ayant un IRE en 5'-UTR. Les équations relatives aux ARNm de la ferritine (équations (5.1) et (5.2)) et de *IRE5* (équations (5.10) et (5.11)) permettent de calculer les expressions suivantes :

$$Ft_{-f}^{eq} = \frac{p_{-Ft} \cdot (K_d \cdot ka + dr_{-Ft})}{dr_{-Ft} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-Ft})} \quad (5.18)$$

$$Ft_{-b}^{eq} = \frac{p_{-Ft} \cdot IRP^{eq} \cdot ka}{dr_{-Ft} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-Ft})} \quad (5.19)$$

$$IRE5_{-f}^{eq} = \frac{p_{-IRE5} \cdot (K_d \cdot ka + dr_{-IRE5})}{dr_{-IRE5} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-IRE5})} \quad (5.20)$$

$$IRE5_{-b}^{eq} = \frac{p_{-IRE5} \cdot IRP^{eq} \cdot ka}{dr_{-IRE5} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-IRE5})} \quad (5.21)$$

De la même manière, nous pouvons calculer deux expressions donnant la concentration, à l'état stationnaire, des ARNm de la ferroportine, à partir des équations (5.16) et (5.5). Ces expressions diffèrent peu de celles de *Ft* et *IRE5* :

$$FPN1a_{-f}^{eq} = \frac{p_{-Ft} \cdot (K_d \cdot ka + dr_{-FPN1a})}{dr_{-Ft} \cdot v_{-FPN1a} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-FPN1a})} \quad (5.22)$$

$$FPN1a_{-b}^{eq} = \frac{p_{-Ft} \cdot IRP^{eq} \cdot ka}{dr_{-Ft} \cdot v_{-FPN1a} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-FPN1a})} \quad (5.23)$$

Nous calculons ensuite les expressions décrivant les concentrations des ARNm ayant un IRE situé dans la région 3'-UTR. En utilisant les équations décrivant la dynamique de ces variables, nous obtenons les expressions suivantes :

$$TfR1_{-f}^{eq} = \frac{p_{-TfR1} \cdot (K_d \cdot ka + dr_{-TfR1})}{IRP^{eq} \cdot ka \cdot dr_{-TfR1} + dr_{-TfR1} \cdot (K_d \cdot ka + dr_{-TfR1})} \quad (5.24)$$

$$TfR1_{-b}^{eq} = \frac{IRP^{eq} \cdot ka \cdot p_{-TfR1}}{IRP^{eq} \cdot ka \cdot dr_{-TfR1} + dr_{-TfR1} \cdot (K_d \cdot ka + dr_{-TfR1})} \quad (5.25)$$

$$IRE3_{-f}^{eq} = \frac{p_{-IRE3} \cdot (K_d \cdot ka + dr_{-IRE3})}{IRP^{eq} \cdot ka \cdot dr_{-IRE3} + dr_{-IRE3} \cdot (K_d \cdot ka + dr_{-IRE3})} \quad (5.26)$$

$$IRE3_{-b}^{eq} = \frac{IRP^{eq} \cdot ka \cdot p_{-IRE3}}{IRP^{eq} \cdot ka \cdot dr_{-IRE3} + dr_{-IRE3} \cdot (K_d \cdot ka + dr_{-IRE3})} \quad (5.27)$$

Nous pouvons alors calculer les expressions décrivant la concentration de la ferritine (resp. ferroportine) à l'état stationnaire en remplaçant, dans l'équation (5.3) (resp. (5.6)), l'expression de Ft_{-f} (reps. $FPN1a_{-f}$) décrite dans l'équation (5.18) (resp. (5.22)).

$$Ft_{-p}^{eq} = \frac{p_{-Ft} \cdot t_{-Ft} \cdot (K_d \cdot ka + dr_{-Ft})}{24 \cdot dp_{-Ft} \cdot dr_{-Ft} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-Ft})} \quad (5.28)$$

$$FPN1a_{-f}^{eq} = \frac{p_{-Ft} \cdot t_{-FPN1a} \cdot (K_d \cdot ka + dr_{-FPN1a})}{2 \cdot dp_{-FPN1a} \cdot dr_{-Ft} \cdot v_{-FPN1a} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{-FPN1a})} \quad (5.29)$$

Pour obtenir l'expression de la variable *TfR1* à l'état stationnaire en fonction de la variable *IRP*, nous remplaçons, dans l'équation algébrique dérivée de l'équation (5.9), les termes $TfR1_{-f}$

et $TfR1_b$ par les expressions décrites dans les équations (5.24) et (5.25). Nous obtenons alors l'expression :

$$TfR1_p^{eq} = \frac{p_{TfR1} \cdot t_{TfR1} \cdot (ka \cdot (IRP^{eq} + K_d) + drs_{TfR1})}{2 \cdot dp_{TfR1} \cdot (IRP^{eq} \cdot drs_{TfR1} \cdot ka + dr_{TfR1} \cdot (K_d \cdot ka + drs_{TfR1}))} \quad (5.30)$$

Nous pouvons alors utiliser l'équation algébrique dérivée de l'équation (5.15) pour obtenir une expression décrivant la variable Fe à l'état stationnaire :

$$Fe^{eq} = \frac{num_{Fe}}{denom_{Fe}} \quad (5.31)$$

avec

$$num_{Fe} = Tf_{sat}^{eq} \cdot dp_{FPN1a} \cdot dr_{Ft} \cdot k_{Fe_input} \cdot p_{TfR1} \cdot t_{TfR1} \cdot v_{FPN1a} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{FPN1a}) \cdot (ka \cdot (IRP^{eq} + K_d) + drs_{TfR1})$$

et

$$denom_{Fe} = dp_{TfR1} \cdot (IRP^{eq} \cdot drs_{TfR1} \cdot ka + dr_{TfR1} \cdot (K_d \cdot ka + drs_{TfR1})) \cdot (2 \cdot dp_{FPN1a} \cdot dr_{Ft} \cdot k_{Fe_cons} \cdot v_{FPN1a} \cdot (ka \cdot (IRP^{eq} + K_d) + dr_{FPN1a}) + k_{Fe_export} \cdot p_{Ft} \cdot t_{FPN1a} \cdot (ka \cdot K_d + dr_{FPN1a}))$$

Nous avons donc une expression pour toutes les variables, à l'état stationnaire, en fonction de la variable IRP . En remarquant que celle-ci peut se réécrire comme indiqué par l'équation (5.32), et en considérant la simplification de la sigmoïde comme indiqué dans la sous-section 5.4.1, nous obtenons l'expression décrivant la concentration de IRP à l'état stationnaire, donnée par l'équation (5.33).

$$\frac{d IRP}{dt} = k_{IRP_prod} - (k_{Fe \rightarrow IRP} \cdot \text{sig}^+(Fe, \theta_{Fe \rightarrow IRP}) + dp_{IRP}) \cdot IRP - \frac{d Ft_b}{dt} - \frac{d FPN1a_b}{dt} - \frac{d TfR1_b}{dt} - \frac{d IRE3_b}{dt} - \frac{d IRE5_b}{dt} \quad (5.32)$$

$$IRP^{eq} = \frac{k_{IRP_prod}}{dp_{IRP} + k_{Fe \rightarrow IRP}} \quad (5.33)$$

Nous avons donc une expression analytique décrivant les concentrations de l'ensemble des acteurs en fonction de paramètres du système. Nous avons donc prouvé l'existence d'un unique état stationnaire, quels que soient les paramètres.

5.4.3 Stabilité de l'état stationnaire

L'étape suivante consiste à tester la stabilité de l'état stationnaire obtenu. Pour cela, nous calculons la matrice Jacobienne de notre système à l'état stationnaire et tentons de tester si les parties réelles des valeurs propres de cette matrice sont négatives. Étant donné que la matrice Jacobienne est de taille 15, le solveur Sage n'arrive pas à calculer, de manière symbolique, les valeurs propres. De plus, pour ce modèle, la Jacobienne n'est, ni triangulaire, ni diagonale, ce qui nous empêche de connaître le signe de la partie réelle des valeurs propres. Il n'est donc pas possible de déterminer analytiquement la stabilité de l'état stationnaire. Lors de la recherche d'un premier jeu de paramètres valide, il sera alors nécessaire de vérifier, pour chaque jeu de paramètres testé, la stabilité de l'état stationnaire.

5.4.4 Réduction des intervalles

En considérant le système à l'équilibre, nous obtenons un ensemble d'équations algébriques. Nous fournissons ces équations au solveur sur intervalles RealPaver [106], ainsi que les relations liant les paramètres et variables à l'état stationnaire. De plus, nous fournissons les intervalles pour chacun des paramètres ainsi que pour les variables à l'état stationnaire. En demandant à Realpaver de calculer l'enveloppe convexe d'un pavage de 1 048 576 boîtes du problème ainsi défini, nous obtenons les réductions d'intervalles décrites dans la table 5.1.

Déduction	Intervalle initial
$FPN1a_p^{eq} > 5.2e-13$	[1.0e-13, 1.0e-5] mol·L ⁻¹
$TfR1_f^{eq} < 9.099e-10$	[1.0e-13, 9.1.0e-7] mol·L ⁻¹
$FPN1a_f^{eq} < 1.18e-9$	[1.0e-13, 1.2e-9] mol·L ⁻¹
$TfR1_b^{eq} < 9.099e-10$	[1.0e-13, 9.1e-10] mol·L ⁻¹
$Fe^{eq} > 1.258e-10$	[1.0e-10, 2.0e-6] mol·L ⁻¹
$FPN1a_b^{eq} < 1.18e-9$	[1.0e-13, 1.2e-9] mol·L ⁻¹
$k_{Fe_export} < 3.072e13$	[2.0e-2, 1.6e14] L·(mol·s) ⁻¹
$\theta_{Fe \rightarrow IRP} < 1.59e-6$	[1.0e-10, 2.0e-6] mol·L ⁻¹

TABLE 5.1 – Déductions obtenues avec Realpaver [106] pour le modèle intégrant les ARNm. Ces résultats sont obtenus en 156 secondes, en demandant au solveur de calculer l'enveloppe convexe d'un pavage de 1 048 576 boîtes. L'ensemble des intervalles et des contraintes fournies au solveur est indiqué dans l'annexe A.

5.5 Recherche d'une région valide

Une fois que les données relatives à l'état stationnaire ont été utilisées, l'étape suivante de la méthodologie consiste à chercher un jeu de paramètres satisfaisant l'ensemble des contraintes (statiques et temporelles). À partir de ce jeu de paramètres valide, l'algorithme de Zamora-Sillero *et al.* [119] procède à une exploration conduisant à l'identification d'une région valide.

Nous utilisons l'algorithme d'optimisation, proposé par Nelder-Mead [118], pour rechercher un jeu de paramètres valide. Comme pour le modèle précédent, nous divisons la formule STL décrivant l'ensemble des contraintes en plusieurs sous-formules. Pour ce modèle, nous considérons deux sous-formules. La première, nommée φ_{Sall} , décrit l'ensemble des contraintes, relatives à l'état stationnaire, présentées dans les sections A.2, A.3 et A.4. La seconde, nommée φ_{Ball} , décrit les contraintes, relatives à la dynamique, présentées dans la sous-section 5.3.2. Nous considérons ensuite un ensemble de jeux de paramètres, échantillonnés uniformément sur une échelle logarithmique, nommé $\mathcal{P}_{U_1}, \dots, \mathcal{P}_{U_N}$. Pour chacun de ces jeux, nous utilisons l'algorithme d'optimisation afin d'optimiser l'évaluation de φ_{Sall} (resp. φ_{Ball}), et obtenons un ensemble de jeux de paramètres $\mathcal{P}'_{U_{1,1}}, \dots, \mathcal{P}'_{U_{N,1}}$ (resp. $\mathcal{P}'_{U_{1,2}}, \dots, \mathcal{P}'_{U_{N,2}}$). Nous utilisons alors les jeux de paramètres $\mathcal{P}'_{U_{1,1}}, \dots, \mathcal{P}'_{U_{N,1}}, \mathcal{P}'_{U_{1,2}}, \dots, \mathcal{P}'_{U_{N,2}}$ comme jeux de paramètres initiaux pour la recherche d'un jeu de paramètres vérifiant la formule (φ_{Sall}) and (φ_{Ball}).

Cette approche n'aboutit pas à l'obtention d'un jeu de paramètres valide, même en considérant un grand nombre de jeux de paramètres initiaux et un grand nombre d'itérations. Un jeu initial étant indispensable pour utiliser l'algorithme de Zamora-Sillero *et al.*, nous devons explorer d'autres méthodes pour obtenir un tel jeu de paramètres. Pour cela, nous cherchons quelle est la sous-formule la plus difficile à vérifier. Cette sous-formule est φ_{Sall} puisqu'aucun

jeu de paramètres ne la respecte.

Nous présentons, dans la suite de cette section, les trois pistes explorées pour obtenir un ou plusieurs jeux de paramètres respectant les contraintes à l'état stationnaire. Les deux premières n'aboutissent pas, mais la troisième permet de résoudre le problème. À partir de ces jeux de paramètres, nous recherchons un jeu de paramètres valide, satisfaisant donc l'ensemble des contraintes.

5.5.1 Recherche de jeux de paramètres valides à l'état stationnaire

Utilisation de RealPaver

La première piste explorée repose sur l'utilisation du solveur sur intervalles Realpaver. Comme indiqué dans la première partie de la sous-section 4.3.1, il est possible d'indiquer au solveur la taille maximale des boîtes calculées. Nous utilisons alors le solveur en lui demandant de calculer une seule boîte, mais d'une précision très élevée. Du fait de la petite taille de cette boîte, celle-ci s'apparente à un point, nous fournissant alors un jeu de paramètres valide. Cette piste, bien que prometteuse, n'est pas exploitable puisque le solveur ne fournit pas de solution en un temps raisonnable.

Calcul analytique de l'état stationnaire

La deuxième piste consiste à chercher des expressions analytiques permettant d'exprimer les paramètres et les conditions initiales ayant un intervalle très large en fonction des paramètres et conditions initiales ayant un intervalle étroit.

Nous échantillons ensuite, selon une distribution logarithmique uniforme, les paramètres et conditions initiales utilisés pour définir les paramètres et conditions initiales ayant un intervalle large. Pour chaque jeu de paramètres de cet échantillonnage, nous testons, d'une part, si les valeurs des paramètres ou conditions initiales définis par une expression sont dans leurs intervalles respectifs et, d'autre part, si les contraintes relatives à l'état stationnaire sont respectées. Malgré un échantillonnage intensif, aucun jeu de paramètres, tel que tous les paramètres et conditions initiales soient dans leurs intervalles, n'est trouvé.

Nouvelle méthode levant ce verrou

Nous sommes face à un verrou qui empêche toute progression. Il apparaît que l'étape qui bloque est l'identification d'un jeu de paramètres (point dans l'espace des paramètres) satisfaisant les contraintes statiques. Nous nous focalisons donc sur ce sous-problème qui est une manifestation de ce que l'on appelle le fléau de la dimensionnalité (*curse of dimensionality*).

Description Pour bien caractériser le problème et suggérer une solution, prenons un exemple simple. Considérons deux inconnues k_1 et k_2 appartenant au rectangle $[0, M1] \times [0, M2]$. Ce rectangle définit l'espace de recherche. Supposons que ces deux inconnues soient soumises aux inégalités $k_2 - k_1 > -\epsilon$ et $k_1 - k_2 > -\epsilon$, où $\epsilon > 0$. Ces deux inégalités définissent une bande centrée sur la diagonale $k_1 = k_2$ du domaine rectangulaire. Cette bande est d'autant plus fine que ϵ est petit par rapport à k_1 et k_2 . On se rend compte facilement que la méthode consistant à échantillonner aléatoirement le rectangle puis tester les inégalités est très inefficace lorsque la bande est très fine. La probabilité de tomber par hasard dans la bande est égale au rapport de la surface de la bande par celle du rectangle. Cette difficulté de toucher une petite région par échantillonnage aléatoire est amplifiée exponentiellement lorsque la dimension de l'espace

de recherche augmente (manifestation de la *curse of dimensionality*). On s'aperçoit qu'il serait beaucoup plus efficace de tirer aléatoirement une valeur, par exemple k_1 , puis propager cette information dans le système d'inégalités. On en déduit, dans cet exemple, ϵ ayant une valeur donnée, que k_2 doit appartenir à l'intervalle $[v(k_1) - \epsilon, v(k_1) + \epsilon]$, où $v(k_1)$ est la valeur de k_1 tirée aléatoirement. Il suffit alors de tirer aléatoirement k_2 dans cet intervalle.

Pour lever le verrou, nous développons une nouvelle approche en généralisant cette idée. Nous considérons les paramètres du modèle et les concentrations à l'état stationnaire comme un ensemble d'inconnues. Nous en avons $28 + 15 = 43$. Ces inconnues sont liées par 15 équations algébriques. Nous n'essayons pas d'exprimer les concentrations en fonction des paramètres, mais nous nous laissons au contraire toute liberté dans le choix du sous-ensemble d'inconnues qui seront exprimées en fonction des inconnues restantes (en exploitant les équations algébriques). L'idée est d'échantillonner une ou plusieurs inconnues et de propager l'information ainsi obtenue en exploitant les équations algébriques et les inégalités, ce qui réduit l'espace de recherche pour les inconnues restantes. On itère ces opérations jusqu'à tomber sur une contradiction ou jusqu'à ce que toutes les inconnues soient instanciées. Dans ce dernier cas, on a obtenu un jeu de paramètres respectant les contraintes de l'état stationnaire. L'algorithme va donc se déployer en un arbre de recherche.

Pour appliquer cette méthode, il faut déterminer un ordre d'échantillonnage des inconnues (autrement dit, l'ordre dans lequel les inconnues vont être instanciées dans l'arbre de recherche).

On donne la priorité aux variables ayant l'intervalle de définition le plus étroit (variables dites « étroites »), permettant ainsi de déduire des informations sur des inconnues « larges » (c'est-à-dire, celles dont l'intervalle de définition est large). De manière générale le principe appliqué pour le choix d'un ordre d'échantillonnage optimal est qu'une variable est prioritaire si elle permet d'appliquer le plus grand nombre de contraintes (inégalités) les plus strictes. Le but est d'utiliser les équations algébriques et les inégalités les plus contraignantes en premier, de façon à tomber sur une contradiction le plus tôt possible le long d'une branche de recherche, ou au moins réduire l'espace de recherche pour les inconnues restantes. De ce fait, à mesure de l'avancée le long d'une branche, les inconnues les plus étroites, les équations algébriques et les inégalités les plus contraignantes sont utilisées. Ceci permet de détecter les échecs tôt dans la recherche (élagage précoce de l'arbre). Lorsque, le long d'une branche, toutes les équations algébriques ont été utilisées ainsi que les inégalités les plus contraignantes, les inégalités non encore utilisées sont par conséquent moins contraignantes et la probabilité d'un échec diminue.

Pour renforcer cet effet d'élagage précoce, nous introduisons des contraintes redondantes. Par exemple, si nous avons $A + B < C$ (avec A , B et C positifs), nous ajoutons au système les contraintes $A < C$ et $B < C$. De cette façon, l'une des inégalités ajoutées pourra être testée plus tôt que l'inégalité d'origine (qui reste de toute façon dans le système) et conduire ainsi éventuellement à une contradiction. Dans le même ordre d'idée, si nous avons une expression du type $D = (A - B)/C$, (avec A , B , C et D positifs), nous introduisons $A - B > 0$.

Une idée supplémentaire consiste à visualiser la structure du système d'équations et d'inéquations sous forme de graphe de dépendance entre inconnues. Pour cela, nous utilisons un graphe biparti où le premier type de nœud représente les inconnues et le deuxième représente les relations entre inconnues. Lors du choix de l'ordre d'instanciation des inconnues, nous privilégions celles qui conduisent à un découplage du système en clusters indépendants d'inconnues. Ce critère est combiné au critère précédent. Le graphe de dépendance associé au modèle (figure 5.2) qui nous occupe montre que l'instanciation des quatre inconnues (Ft_{-f} , $FPN1a_{-f}$, $TfR1_{-f}$ et $TfR1_{-b}$) permet d'obtenir trois clusters déconnectés. Ceci permet de factoriser le problème initial en sous-problèmes indépendants et réduit considérablement le coût de la recherche.

inconnues sont testées. Si la contrainte est respectée, alors le paramètre p_{Ft} est échantillonné et la valeur de Ft_b est calculée en utilisant la somme des équations algébriques dérivées des équations (5.1) et (5.2). Ceci permet alors de vérifier que la contrainte $Ft_f + Ft_b \in [1.3e-8, 3.1e-7]$ est respectée. Les inconnues dont les valeurs sont calculées sont ensuite : ka , $TfR1_b$, $k_{IRP_{prod}}$, drs_{TfR1} , p_{TfR1} , $IRE3_f$, p_{IRE3} , $FPN1a_b$, $FPN1a_f$, p_{IRE5} , dp_{FPN1a} et Fe . Nous pouvons observer qu'aucune de ces inconnues n'a un intervalle étroit.

Une fois l'ordre des instanciations fixé, les calculs effectués à chaque étape sont extrêmement simples et l'algorithme nous donne une liste de points valides, pour les contraintes statiques, en quelques minutes.

5.5.2 Recherche d'un jeu de paramètres valide

Une fois que nous avons obtenu un ensemble de jeux de paramètres respectant les contraintes de l'état stationnaire, nous voulons obtenir un jeu de paramètres respectant également les contraintes sur la dynamique du système. Pour cela, chaque jeu de paramètres est simulé et nous testons si la simulation respecte la formule STL φ_{Ball} , afin de ne conserver que les jeux de paramètres valides.

Nous trouvons qu'aucun point de la liste initiale ne satisfaisait les contraintes temporelles. Pour résoudre ce problème, nous choisissons d'appliquer une approche dans laquelle la région valide est approchée progressivement. Pour cela, nous affaiblissons le système de contraintes de façon à ce qu'un des jeux de paramètres obtenus précédemment soit valide. Nous utilisons alors l'algorithme d'expansion de Zamora-Sillero *et al.* [119], à partir de ce jeu de paramètres, ce qui donne une région valide, selon les contraintes affaiblies, que nous nommons $R1$. Ensuite, nous utilisons l'échantillonnage de $R1$ fourni par l'algorithme de Zamora *et al.* et, pour chaque point (jeu de paramètres), nous testons les contraintes d'origine (non affaiblies), à la recherche d'un jeu de paramètres satisfaisant toutes les contraintes. Lorsqu'un tel jeu initial a été trouvé, nous exécutons une deuxième fois l'algorithme de Zamora-Sillero *et al.* à partir de celui-ci pour obtenir la région valide souhaitée $R2$. L'intérêt de procéder de la sorte est de résoudre deux problèmes plus faciles que le problème initial. La région $R1$ est plus facile à identifier car son volume est plus important que celui de la région $R2$. Toutefois, son volume est extrêmement réduit par rapport à l'espace de recherche initial, rendant ensuite la recherche de la région $R2$ plus aisée. La région $R2$, si elle existe, est en principe à l'intérieur de $R1$. Pour la trouver, on peut se permettre d'échantillonner $R1$ beaucoup plus finement que ce qu'on aurait pu faire avec l'hyper-rectangle initial.

Pour appliquer cette méthode, il est nécessaire de choisir des contraintes à affaiblir dans le système de contraintes initial. Pour cela, nous faisons une statistique sur les contraintes qui sont le plus souvent non satisfaites. Dans notre modèle, pour avoir un jeu de paramètres valide permettant d'obtenir la région $R1$, nous relaxons 1) la contrainte φ_{B1} en remplaçant le facteur 0.16 par un facteur 0.1, 2) la contrainte φ_{B2} en remplaçant le facteur 0.25 par 0.35 et 3) la contrainte $\varphi_{B3.2}$ en la supprimant. L'algorithme d'expansion calcule alors cette région $R1$. Au sein de cette région, nous avons trouvé un jeu de paramètres respectant les contraintes non affaiblies, permettant d'obtenir la région $R2$.

5.6 Résultats

Nous présentons les résultats obtenus lors de la recherche d'une région valide. Nous avons caractérisé une région valide par une union de 26 ellipsoïdes. Cette représentation géométrique

de la région valide est bien adaptée aux hautes dimensions : le ratio du volume de la région valide sur le volume de l'hyper-rectangle englobant l'union d'ellipsoïdes vaut $1.2\text{e-}11 \pm 3.0\text{e-}13$.

La figure 5.3 présente les trajectoires des centres des ellipsoïdes. Nous considérons, dans un premier temps, qu'elles sont représentatives de l'ensemble des trajectoires correspondant aux jeux de paramètres valides trouvés. Ce résultat devra cependant être amélioré, de façon à visualiser plus complètement l'ensemble des comportements compatibles avec les contraintes. Nous observons que, comme pour le modèle précédent, la concentration de la ferroportine ne varie pas au cours du temps. L'observation de ces courbes permet de répondre à la question initialement posée : « est-ce que la quantité d'IRP actif est supérieure à la quantité d'IRE ou est-ce l'inverse ? » Nous pouvons constater que, après la mise en carence, la concentration d'IRP est aux alentours des $1.0\text{e-}7 \text{ mol}\cdot\text{L}^{-1}$, alors que la somme des concentration des ARNm est au plus aux alentours des $1.0\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$. Nous concluons que la quantité d'IRP actifs est significativement plus élevée que celle des IRE.

Nous montrons, dans la figure 5.4, pour chaque paramètre et condition initiale, la répartition des valeurs de cette inconnue au sein de l'intervalle possible des valeurs. Nous observons que la valeur du seuil $\theta_{Fe \rightarrow IRP}$ est légèrement inférieure à $1.0\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$. Ce résultat est en accord avec les résultats obtenus pour le modèle précédent. Dans une publication récente [79], nos collaborateurs ont estimé, à partir de résultats expérimentaux, la valeur de ce seuil à $6.1\text{e-}8 \text{ mol}\cdot\text{L}^{-1}$. Cet accord entre nos résultats et l'estimation obtenue à partir de données expérimentales est un résultat majeur de la présente thèse.

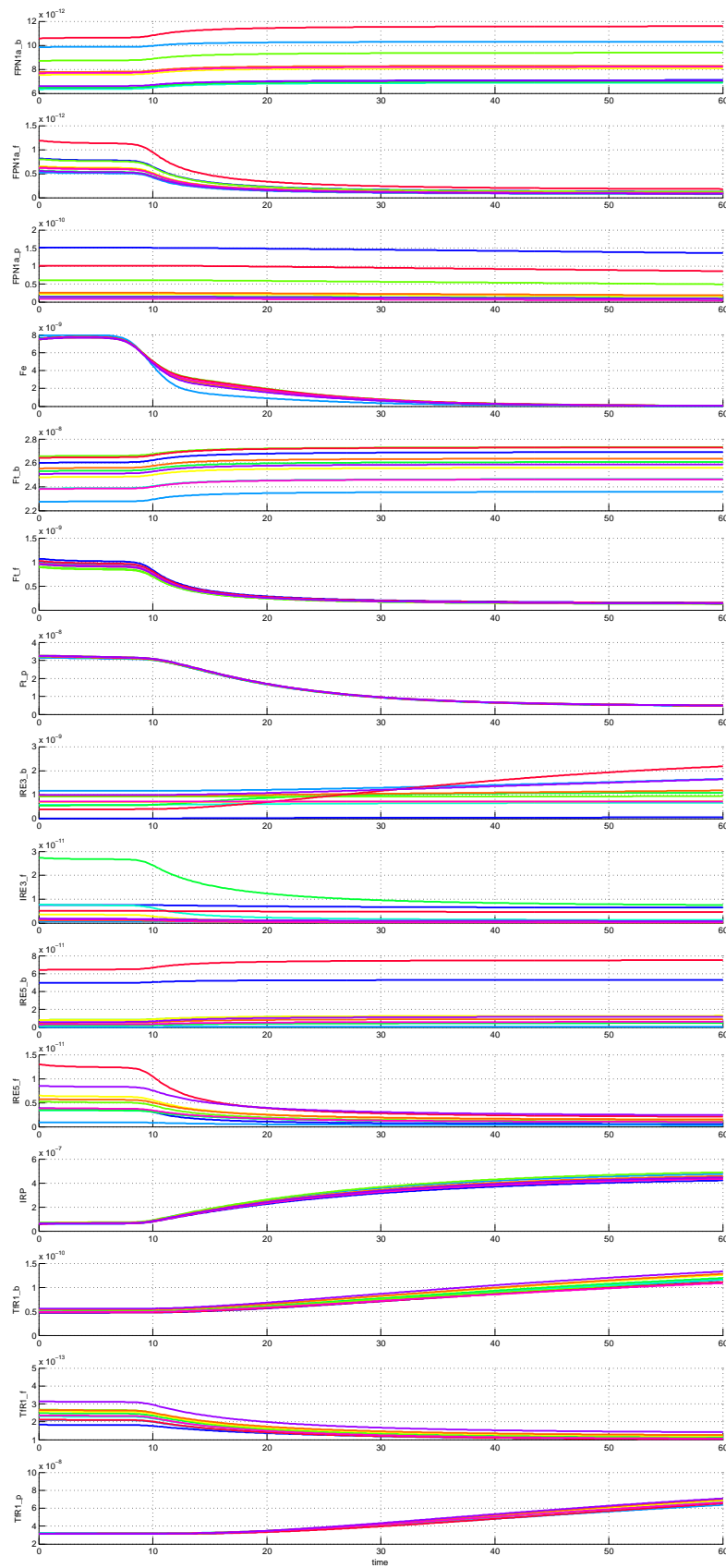


FIGURE 5.3 – Simulations des jeux de paramètres qui sont les centres des ellipsoïdes définissant la région valide. L'application de la mise en carence à lieu à $t=6$ heure. En abscisse, le temps en heure, en ordonnée, les concentrations en $\text{mol}\cdot\text{L}^{-1}$.

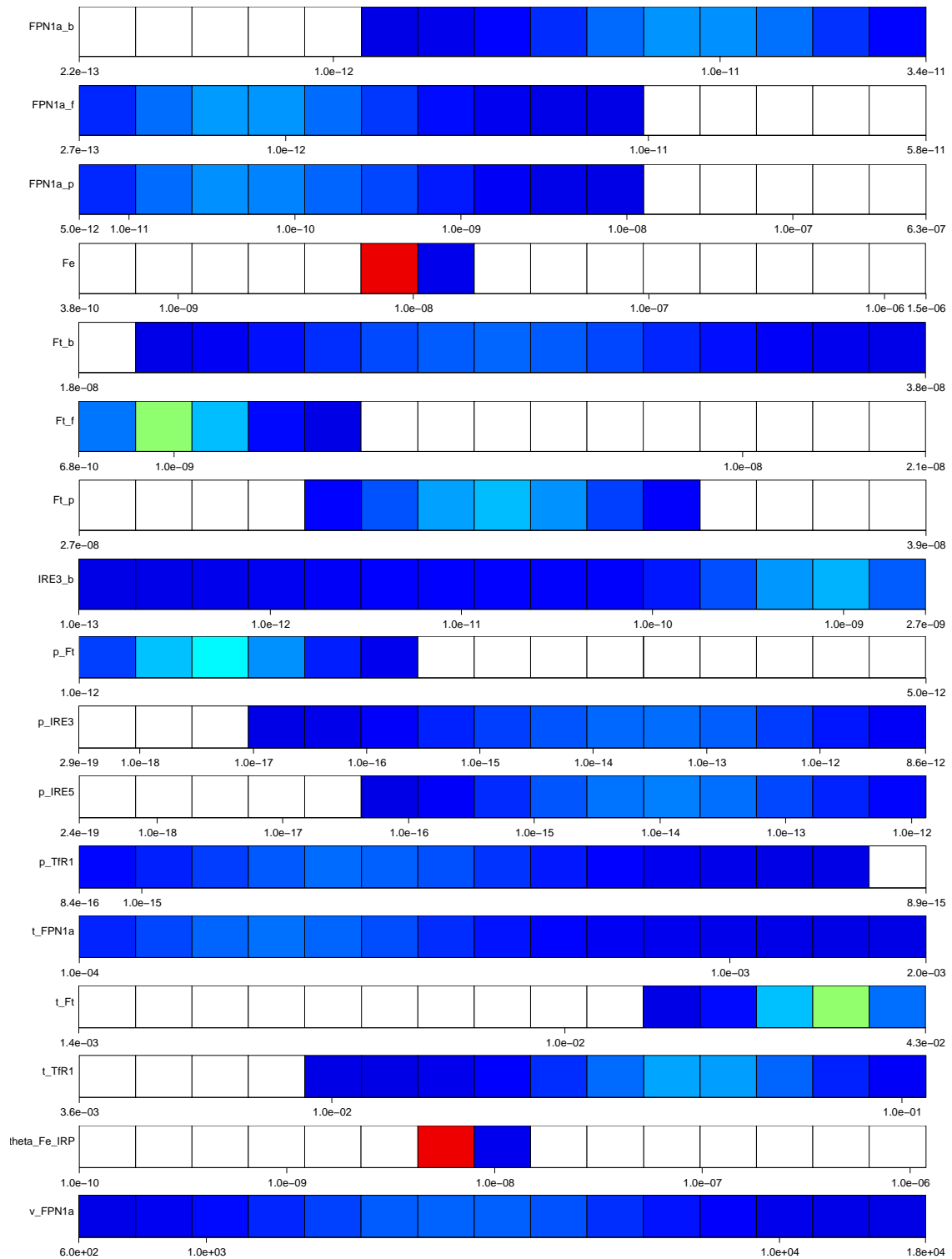


FIGURE 5.4 – Distribution des valeurs de certains paramètres et conditions initiales. La couleur des rectangles représente la proportion de jeux de paramètres pour lesquels la valeur du paramètre est dans l'intervalle. La proportion est indiquée dans la figure 5 de l'article décrivant le modèle précédent.

Chapitre 6

Conclusion

Les progrès en biologie expérimentale, tant au niveau moléculaire que systémique, ont amélioré notre connaissance des réseaux biologiques et des processus cellulaires. L'utilisation d'outils basés sur des modèles quantitatifs détaillés est devenue nécessaire pour pouvoir intégrer l'ensemble des données disponibles et étudier ces réseaux. L'utilisation de tels modèles permet d'apporter un autre éclairage sur les systèmes biologiques complexes et ainsi d'aider le biologiste dans le choix des expériences à effectuer.

Nous avons créé une méthodologie pour l'étude de systèmes biologiques qui consiste à traduire l'ensemble des données disponibles en intervalles, relations à l'état stationnaires et en formule STL. Ensuite, elle se concentre sur l'étude de l'état stationnaire, puis recherche un jeu de paramètres valide utilisé pour trouver une région valide. Cette méthodologie présente les avantages de permettre l'intégration de toutes les données disponibles, de générer l'ensemble des modèles respectant ces données, de déduire des propriétés de cet ensemble de modèles et d'être automatique. Pour cela, cette méthodologie intègre différents outils : formels, analytiques et numériques. Le solveur sur intervalles, Realpaver, permet d'utiliser les données disponibles sur l'état stationnaire pour réduire les intervalles des paramètres et des conditions initiales. Le solveur symbolique Sage permet de calculer, d'une part, des expressions donnant les conditions initiales en fonction des paramètres afin que le système soit dans un état stationnaire, et d'autre part, la Jacobienne permettant de vérifier la stabilité de cet état stationnaire. Ensuite, l'outil Breach, intégrant le solveur d'équations différentielles CVODES, est utilisé pour simuler le système d'équations différentielles et calculer l'évaluation quantitative des formules STL traduisant les données comportementales. L'algorithme d'optimisation proposé par Nelder-Mead permet d'obtenir un premier jeu de paramètres valide (dans le cas où cet algorithme n'est pas efficace, nous avons utilisé une méthode *ad hoc* pour obtenir un tel jeu de paramètres). Enfin, l'algorithme développé par Zamora *et al.*, basé sur l'utilisation d'ellipsoïdes, permet d'obtenir une région valide, à partir de ce jeu de paramètres valide. En dernier lieu, des outils statistiques, tels le calcul de corrélation ou de régression linéaire, permettent d'analyser la région valide obtenue.

Nous avons créé deux modèles du réseau de l'homéostasie du fer, auxquels nous avons appliqué cette méthodologie. De ces études, il est ressorti que l'ordre de grandeur pour le seuil $\theta_{Fe \rightarrow IRP}$ est inférieur à $1e-8 \text{ mol} \cdot \text{L}^{-1}$. Ce résultat de nos études a guidé nos collaborateurs dans le choix de leurs expériences. Les résultats de ces expériences confirment le résultat que nous avons obtenus en utilisant notre méthodologie.

L'étude du deuxième modèle de l'homéostasie du fer a permis de découvrir une limitation de notre méthodologie. Nous avons alors proposé une méthode semi-manuelle permettant de s'affranchir de cette limitation. Dans un objectif d'automatisation de la méthodologie, il est

alors nécessaire d'automatiser cette nouvelle étape. Par ailleurs, il peut être impossible de calculer analytiquement la stabilité de l'état stationnaire. Il est alors possible d'utiliser d'autres méthodes, plus globales, pour tester la stabilité, comme, par exemple, celles liées à la décomposition potentielle-hamiltonienne [151]. Finalement, une optimisation possible consiste à paralléliser les calculs de simulation et d'évaluation des formules STL, réduisant ainsi les temps de calcul.

Diverses améliorations peuvent également être apportées aux modèles. Par exemple, il est possible de s'inspirer du modèle de l'entrée du fer dans les ferritines proposé par Salgado *et al.* [112] pour modéliser plus finement ce processus. Il est également possible de coupler le réseau de l'homéostasie du fer avec d'autres réseaux, tel le réseau du stress oxydatif, ces deux réseaux étant en interaction.

Annexes

Annexe A

Données et intervalles pour le modèle intégrant les ARNm

Dans cette annexe, nous présentons les données utilisées pour le modèle intégrant les ARN messagers, décrit dans le chapitre 5. Nous commençons par présenter les intervalles des paramètres et des conditions initiales, ainsi que les données utilisées pour les déterminer. Ensuite, nous présentons, dans la section A.3, l'ensemble des contraintes s'appliquant à l'état stationnaire, puis, dans la section A.4, une réécriture de ces contraintes permettant de mieux les exploiter.

De manière générale, afin d'éviter des intervalles trop étroits, nous imposons que le rapport de la borne supérieure sur la borne inférieure d'un intervalle soit au moins égal à 10. Lorsque nous parlons des concentrations à l'état stationnaire, nous ajoutons « eq » en indice des variables.

A.1 Paramètres

$p_{Ft} \in [6.5e-14, 1.6e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ En considérant l'état stationnaire en présence de fer, en additionnant les équations de Ft_b et Ft_f , nous avons l'égalité $p_{Ft} = dr_{Ft} \cdot Ft_r^{eq}$. Nous savons que dr_{Ft} est compris dans $[5.0e-6, 5.0e-5] \text{ s}^{-1}$ et Ft_r^{eq} appartient à l'intervalle $[1.3e-8, 3.1e-7] \text{ mol} \cdot \text{L}^{-1}$ (cf partie *Contraintes du système* : $Ft_f^{eq} + Ft_b^{eq}$). Nous déduisons donc que p_{Ft} appartient à $[6.5e-14, 1.6e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$.

$t_{Ft} \in [4.8e-5, 1.8e2] \text{ s}^{-1}$ 1/ Nous savons que, à l'équilibre, la vitesse de traduction des sous-unités de la ferritine est comprise dans $[1.5e-11, 1.8e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$. Cette vitesse de traduction est décrite par $t_{Ft} \cdot Ft_f^{eq}$. 2/ La concentration de l'ARNm non complexé (des sous-unités) de la ferritine est compris dans $[1.0e-13, 3.1e-7] \text{ mol} \cdot \text{L}^{-1}$. 3/ Nous en déduisons que t_{Ft} est compris dans l'intervalle $[4.8e-5, 1.8e2] \text{ s}^{-1}$.

$dr_{Ft} \in [5.0e-6, 5.0e-5] \text{ s}^{-1}$ Le papier de Sharova *et al.* [146] décrit la constante de vitesse de dégradation des deux sous-unités de la ferritine pour des cellules souches chez la souris. La constante de vitesse de dégradation de l'ARNm de Fth1 est de 0.0717 h^{-1} et celle de Ft12 est de 0.0807 h^{-1} . L'erreur est d'environ 0.01 h^{-1} . Nous obtenons donc que la constante de vitesse de dégradation de l'ARNm de la ferritine est incluse dans $[1.7e-5, 2.3e-5] \text{ s}^{-1}$. Afin d'éviter un intervalle trop étroit, nous élargissons l'intervalle.

$dp_{Ft} \in [3.8e-6, 3.8e-5] \text{ s}^{-1}$ Chez des K562, la demi-vie de la ferritine est de 11 heures [152], ce qui correspond à une constante de vitesse de dégradation de $1.75e-5 \text{ s}^{-1}$. Il nous paraît raisonnable de considérer que la demi-vie de la ferritine est comprise entre 10 et 25 heures. Nous obtenons alors une constante de vitesse de dégradation comprise dans

$[7.7\text{e-}6, 1.9\text{e-}5] \text{ s}^{-1}$. Afin d'éviter un intervalle trop étroit, nous élargissons l'intervalle obtenu.

$k_{Fe_input} \in [8.5\text{e-}3, 8.5\text{e-}2] \text{ s}^{-1}$ Chez des cellules K562, en situation non carencée, la vitesse de l'entrée du fer est de $7.2\text{e}4 \pm 2.4\text{e}4$ atomes de Fe/(cellule·min) [80]. Le volume d'une cellule K562 étant compris dans $[1.7\text{e-}12, 3.4\text{e-}12]$ litre, nous obtenons que la vitesse d'entrée du fer est dans $[3.9\text{e-}10, 1.6\text{e-}9] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$. De plus, nous savons que Tf_{sat} est égal à 0.3 et que la concentration de $TfR1_{-p}^{eq}$ est comprise dans $[6.8\text{e-}8, 1.4\text{e-}7] \text{ mol} \cdot \text{L}^{-1}$. Nous obtenons donc que k_{Fe_input} est compris dans $[9.3\text{e-}3, 7.8\text{e-}2] \text{ s}^{-1}$. Afin d'éviter un intervalle trop étroit, nous élargissons l'intervalle.

$n_{Ft} \in [50, 2000]$ - Chez les mammifères, une ferritine contient jusqu'à 4500 atomes de fer [111]. Cette donnée est utilisée dans le modèle sans ARN. L'intervalle décrit semble précautionneux.

$k_{Fe_cons} \in [2.0\text{e-}7, 16] \text{ s}^{-1}$ En raisonnant en terme de flux chez des cellules K562, à l'état stationnaire en présence de fer, la vitesse d'entrée du fer est comprise dans l'intervalle $[3.9\text{e-}10, 1.6\text{e-}9] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ (cf. *intervalle pour k_{Fe_input}*). La concentration en fer à l'équilibre est comprise dans l'intervalle $[1.0\text{e-}10, 2.0\text{e-}6] \text{ mol} \cdot \text{L}^{-1}$. Si nous considérons que ce flux est le seul flux de sortie du fer, nous trouvons que k_{Fe_cons} est compris dans $[2.0\text{e-}4, 16] \text{ s}^{-1}$. Nous avons donc la borne maximale de l'intervalle décrit. En supposant que ce flux représente au moins un millième de l'utilisation du fer, nous obtenons la borne inférieure.

$k_{Fe_export} \in [2.0\text{e-}2, 1.6\text{e}14] \text{ L} \cdot (\text{mol} \cdot \text{s})^{-1}$ En raisonnant en terme de flux chez des cellules K562, à l'état stationnaire, la vitesse d'entrée du fer est comprise dans l'intervalle $[3.9\text{e-}10, 1.6\text{e-}9] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ (cf. *intervalle pour k_{Fe_input}*). Nous savons par ailleurs que l'intervalle pour $FPN1a_{-p}^{eq}$ est $[1.0\text{e-}13, 1.0\text{e-}5] \text{ mol} \cdot \text{L}^{-1}$. La concentration en fer à l'état stationnaire étant comprise dans $[1.0\text{e-}10, 2.0\text{e-}6] \text{ mol} \cdot \text{L}^{-1}$, si nous supposons que ce flux est le seul flux de sortie du fer, nous trouvons que k_{Fe_export} est compris dans $[2.0\text{e}1, 1.6\text{e}14] \text{ L} \cdot (\text{mol} \cdot \text{s})^{-1}$. Nous en déduisons la borne maximale de l'intervalle. En supposant que ce flux représente au moins un millième de l'utilisation du fer, nous obtenons la borne inférieure.

$\theta_{Fe \rightarrow IRP} \in [1.0\text{e-}10, 2.0\text{e-}6] \text{ mol} \cdot \text{L}^{-1}$ La concentration du fer est comprise dans cet intervalle. Le seuil correspond à une concentration possible du fer.

$k_{IRP_prod} \in [1.1\text{e-}17, 1.0\text{e-}10] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ 1/ Dans leur article, Omholt *et al.* [96] estiment que la vitesse de production de IRP est de $1.0\text{e-}13 \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$. Cette donnée n'étant pas issue d'expériences, par précaution, nous supposons que la vitesse de production de IRP n'est pas 1000 fois plus élevée, nous obtenons alors la borne supérieure. 2/ Dans l'article de Pantopoulos *et al.* [153], les auteurs carencent en fer des cellules fibroblastes Ltk⁻. Une heure après la mise en carence, la vitesse de production de IRP1 (actif) est de $0.4 u$ par heure, où u correspond à la concentration de IRP à l'état stationnaire avant carence. Nous savons que l'intervalle décrivant la concentration de IRP à l'état stationnaire est $[1.0\text{e-}13, 1.0\text{e-}6] \text{ mol} \cdot \text{L}^{-1}$. Étant donné que nous ne considérons, a minima, pas la dégradation de IRP, nous n'obtenons qu'une borne inférieure. Nous obtenons donc que la vitesse de production de IRP est au moins de $1.1\text{e-}17 \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$.

$k_{Fe \rightarrow IRP} \in [1.0\text{e-}5, 1.0\text{e-}4] \text{ s}^{-1}$ L'ajout de fer dans le milieu de culture de cellules H1299 réduit la demi-vie de IRP2 à 2 heures [129]. La constante de vitesse de dégradation due à la dégradation spontanée et à la régulation par le fer est donc de $9.6\text{e-}5 \text{ s}^{-1}$. En considérant la valeur obtenue pour dp_{IRP} dans les conditions expérimentales similaires avant l'ajout en fer (cf. *paramètre dp_{IRP}*), soit $3.9\text{e-}5 \text{ s}^{-1}$, nous trouvons que $k_{Fe \rightarrow IRP}$ est égal à

$5.7\text{e-}5 \text{ s}^{-1}$. Par ailleurs, Samaniego *et al.* [56] indiquent qu'en présence de fer, la demi-vie de IRP2 descend à 6 heures, correspondant à une constante de vitesse de dégradation égale à $3.2\text{e-}5 \text{ s}^{-1}$. La demi-vie dans des conditions similaires, exceptée pour la présence de fer, est supérieure à 24 heures (*cf. paramètre* dp_{IRP}), soit une constante de vitesse de dégradation inférieure à $8.0\text{e-}6 \text{ s}^{-1}$. Nous obtenons donc que la constante de vitesse de dégradation due au fer est supérieure ou égale à $2.4\text{e-}5 \text{ s}^{-1}$. Afin d'éviter un intervalle trop étroit, nous considérons un intervalle autour de ces valeurs.

$dp_{IRP} \in [5.0\text{e-}6, 5.0\text{e-}5] \text{ s}^{-1}$ Chez des H1299, la demi-vie de IRP2 est d'environ 5 heures [129]. Nous obtenons ainsi que dp_{IRP} est égal à $3.9\text{e-}5 \text{ s}^{-1}$. Par ailleurs, Samaniego *et al.* [56] indiquent une demi-vie de IRP2 supérieure à 24 heures. Nous obtenons donc que la constante de vitesse de dégradation due à la présence de fer est inférieure à $8.0\text{e-}6 \text{ s}^{-1}$. Afin d'éviter un intervalle trop étroit, nous considérons un intervalle autour de ces valeurs.

$v_{FPN1a} \in [2.0\text{e}1, 1.6\text{e}6]$ - Ce paramètre, décrivant le rapport $Ft_{-r}^{eq}/FPN1a_{-r}^{eq}$ chez des cellules K562, est compris entre 20 et 200 [154]. Il permet de déduire la vitesse de production de l'ARNm de la ferroportine. Nous ne considérons que la valeur minimale fournie (soit 20), et la prenons comme une borne inférieure. Nous savons que, à l'état stationnaire, la concentration maximale de l'ARNm de la ferritine est de $3.1\text{e-}7 \text{ mol}\cdot\text{L}^{-1}$ (*cf. contrainte* $Ft_{-f}^{eq} + Ft_{-b}^{eq}$) et que la concentration de la ferroportine est supérieure à $2.0\text{e-}13 \text{ mol}\cdot\text{L}^{-1}$ (*cf. intervalle de* $FPN1a_{-f}^{eq}$ *et de* $FPN1a_{-b}^{eq}$). Nous obtenons donc que la valeur maximale de ce paramètre est $1.6\text{e}6$.

$t_{FPN1a} \in [1.0\text{e-}4, 2.0\text{e-}3] \text{ s}^{-1}$ Le papier de Sun *et al.* [155] présente un modèle du système P53 ayant des constantes de vitesse de traduction comprises entre $1.7\text{e-}4$ et $1.7\text{e-}3 \text{ s}^{-1}$. Par ailleurs, le papier de Reuveni *et al.* [147] présente une méthode pour calculer la constante de vitesse de traduction d'un ARN, et leur exemple montre des constantes de vitesse de traduction de l'ordre de $1.0\text{e-}4$ à $2.0\text{e-}4 \text{ s}^{-1}$. Nous considérons donc que la constante de vitesse de traduction de l'ARNm de FPN1a est comprise dans $[1.0\text{e-}4, 2.0\text{e-}3] \text{ s}^{-1}$.

$dr_{FPN1a} \in [4.0\text{e-}5, 4.0\text{e-}4] \text{ s}^{-1}$ La demi-vie de l'ARNm de FPN1a, chez des cellules K562, est de 90 minutes [70], ce qui correspond à une constante de vitesse de dégradation de $1.28\text{e-}4 \text{ s}^{-1}$. Afin d'éviter les intervalles trop étroit, nous considérons l'intervalle décrit.

$dp_{FPN1a} \in [2.2\text{e-}8, 9.6\text{e-}6] \text{ s}^{-1}$ Chez des macrophages chez la souris, la demi-vie de FPN1a est supérieure à 20 heures [128], ce qui correspond à une constante de vitesse de dégradation inférieure à $9.6\text{e-}6 \text{ s}^{-1}$. Nous supposons que la demi-vie de FPN1a n'est pas supérieure à 1 an. Ceci correspond à une borne inférieure de $2.2\text{e-}8 \text{ s}^{-1}$.

$p_{TfR1} \in [3.4\text{e-}18, 6.4\text{e-}14] \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$ 1/ Dans l'article de Seiser *et al.* [62], les auteurs indiquent que, chez des cellules Ltk⁻, la demi-vie de l'ARNm de TfR1 en présence de fer est de 3.5 heures. Cela correspond à une constante de vitesse de dégradation de $5.5\text{e-}5 \text{ s}^{-1}$. De plus, dans Sharova *et al.* [146], les auteurs indiquent que, chez des cellules souches de la souris, la constante de vitesse de dégradation de TfRc (autre nom de TfR1), en présence de fer, est de $0.08\pm 0.015 \text{ h}^{-1}$ (écart-type estimé à partir des mesures présentées), soit $[1.8\text{e-}5, 2.6\text{e-}5] \text{ s}^{-1}$. Nous élargissons l'intervalle et considérons alors une constante de vitesse de dégradation de $TfR1_{-r}$ comprise dans $[7.0\text{e-}6, 7.0\text{e-}5] \text{ s}^{-1}$. 2/ Par ailleurs, à l'état stationnaire, la concentration en ARNm du récepteur est comprise dans $[4.9\text{e-}13, 9.1\text{e-}10] \text{ mol}\cdot\text{L}^{-1}$ (*cf. partie Contraintes à l'état stationnaire*). 3/ En prenant la somme des équations de $TfR1_{-f}$ et $TfR1_{-b}$ à l'équilibre, nous obtenons la relation $0 = p_{TfR1} - [7.0\text{e-}6, 7.0\text{e-}5] \cdot TfR1_{-r}^{eq}$. Nous obtenons alors que p_{TfR1} appartient à $[3.4\text{e-}18, 6.4\text{e-}14] \text{ mol}\cdot(\text{L}\cdot\text{s})^{-1}$.

$t_{_TfR1} \in [3.6e-3, 1.4e1] \text{ s}^{-1}$ En considérant l'équation de $TfR1_p$ à l'équilibre, nous obtenons

$t_{_TfR1} = \frac{2 \cdot dp_{_TfR1} \cdot TfR1_p^{eq}}{TfR1_f^{eq} + TfR1_b^{eq}}$. Nous savons que la valeur mesurée pour $dp_{_TfR1}$ est de $2.4e-5 \text{ s}^{-1}$, celle pour $TfR1_p^{eq}$ est dans $[6.8e-8, 1.4e-7] \text{ mol} \cdot \text{L}^{-1}$ et que l'intervalle pour $TfR1_r^{eq}$ est dans $[4.9e-13, 9.1e-10] \text{ mol} \cdot \text{L}^{-1}$ (cf. *partie Contrainte à l'état stationnaire : TfR1_f + TfR1_b*). Nous obtenons donc que $t_{_TfR1}$ est compris dans l'intervalle $[3.6e-3, 1.4e1] \text{ s}^{-1}$.

$dr_{_TfR1} \in [2.2e-8, 7.0e-2] \text{ s}^{-1}$ Nous savons que le rapport $\frac{TfR1_b}{TfR1_f}$ est inférieur à 1.0e3 (cf.

partie Contraintes à l'état stationnaire) et que la dégradation totale de l'ARNm de TfR1 (libre+complexé) est comprise dans $[7.0e-6, 7.0e-5] \text{ s}^{-1}$ (cf. *intervalle pour p_{_TfR1}*). Comme montré dans la partie « ré-écriture des contraintes », nous en déduisons que $dr_{_TfR1}$ est inférieur à $7.0e-2 \text{ s}^{-1}$. Nous supposons que la demi-vie de $TfR1_f$ n'est pas supérieure à un an. Nous obtenons donc que la borne inférieure est égale à $2.2e-8 \text{ s}^{-1}$.

$drs_{_TfR1} \in [2.2e-8, 7.0e-2] \text{ s}^{-1}$ La constante de vitesse de dégradation de l'ARNm de TfR1 stabilisé est inférieure à celle de la dégradation de l'ARNm non stabilisé. Nous obtenons donc la borne supérieure de l'intervalle. Nous supposons que la demi-vie de $TfR1_f$ n'est pas supérieure à un an. Nous obtenons donc que la borne inférieure est égale à $2.2e-8 \text{ s}^{-1}$.

$dp_{_TfR1} \in [7.6e-6, 7.6e-5] \text{ s}^{-1}$ La demi-vie de TfR1 chez des cellule K562 est de 8 heures [130], ce qui correspond à $dp_{_TfR1} = 2.4e-5 \text{ s}^{-1}$. Nous avons considéré un intervalle autour de cette valeur.

$ka \in [2.1e-1, 1.6e15] \text{ L} \cdot (\text{mol} \cdot \text{s})^{-1}$ Supposons, sans justification particulière, que la quantité d'ARNm complexée, à l'état stationnaire, est de l'ordre de grandeur de la quantité d'ARNm produit. Nous considérons, arbitrairement, cette hypothèse pour la ferritine. Nous avons donc la relation $p_{_Ft} = ka \cdot Ft_f^{eq} \cdot IRP^{eq}$ d'où nous tirons que $ka = \frac{p_{_Ft}}{Ft_f^{eq} \cdot IRP^{eq}}$. La concentration de Ft_f^{eq} est comprise dans $[1.0e-13, 3.1e-7] \text{ mol} \cdot \text{L}^{-1}$, $p_{_Ft}$ est compris dans $[6.5e-14, 1.6e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$, et IRP^{eq} est dans $[1.0e-13, 1.0e-6] \text{ mol} \cdot \text{L}^{-1}$. Nous en déduisons que ka appartient à $[2.1e-1, 1.6e15] \text{ L} \cdot (\text{mol} \cdot \text{s})^{-1}$.

$K_d \in [1.0e-11, 1.0e-8] \text{ mol} \cdot \text{L}^{-1}$ Les mesures de la constante thermodynamique K_d (correspondant au rapport constante de vitesse de dissociation sur constante de vitesse d'association, i.e. : kd/ka) faites par Goforth *et al.* [144], comme celles faites par nos collaborateurs, indiquent que cette constante varie peu selon les différents IRE considérés. Nous considérons donc que cette constante thermodynamique est la même pour tous les IRE. Néanmoins, les valeurs mesurées par Goforth *et al.* sont comprises dans l'intervalle $[1.0e-11, 1.3e-10] \text{ mol} \cdot \text{L}^{-1}$ [144], alors que celle obtenues par nos collaborateurs se situent dans l'intervalle $[1.0e-9, 1.0e-8] \text{ mol} \cdot \text{L}^{-1}$. Nous considérons donc que K_d peut être compris dans l'intervalle $[1.0e-11, 1.0e-8] \text{ mol} \cdot \text{L}^{-1}$.

$p_{_IRE3} \in [1.0e-19, 1.0e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ Par analogie avec les autres paramètres de production d'ARNm, nous considérons l'intervalle décrit.

$dr_{_IRE3} \in [2.2e-8, 7.0e-1] \text{ s}^{-1}$ Nous considérons l'intervalle de $dr_{_TfR1}$ auquel nous rajoutons un ordre de grandeur à la borne supérieure. Par ailleurs, nous considérons que la demi-vie de $IRE3_f$ n'est pas supérieure à un an. Nous obtenons donc une borne inférieure égale à $2.2e-8 \text{ s}^{-1}$.

$drs_{_IRE3} \in [2.2e-8, 7.0e-1] \text{ s}^{-1}$ Nous considérons l'intervalle de $drs_{_TfR1}$ auquel nous rajoutons un ordre de grandeur à la borne supérieure. Par ailleurs, nous considérons que la

demi-vie de $IRE3_b$ n'est pas supérieure à un an. Nous obtenons donc une borne inférieure égale à $2.2e-8 \text{ s}^{-1}$.

$p_{IRE5} \in [1.0e-19, 1.0e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ Par analogie avec les autres paramètres de production d'ARNm, nous considérons l'intervalle décrit.

$dr_{IRE5} \in [2.2e-8, 1.0e-2] \text{ s}^{-1}$ Les constantes de vitesse de dégradation des ARNm habituellement obtenues (*cf. intervalles de dr_{Ft} et de dr_{FPN1a}*) sont de l'ordre de $1.0e-5 \text{ s}^{-1}$. Nous considérons que la constante de vitesse de dégradation n'est pas 1000 fois plus élevée. Nous obtenons alors la borne supérieure. Par ailleurs, nous considérons que la demi-vie de $IRE5_r$ n'est pas supérieure à un an. Nous obtenons donc une borne inférieure égale à $2.2e-8 \text{ s}^{-1}$.

A.2 Concentrations à l'état stationnaire

$Ft_p^{eq} \in [3.6e-11, 6.7e-7] \text{ mol} \cdot \text{L}^{-1}$ 1/ Dans leur article, Brazzolotto *et al.* [156] indiquent que la concentration en ferritine chez des K562 en situation contrôle est dans $[1.0e-9 ; 1.1e-9] \text{ g}/\mu\text{g}$ de protéine cytoplasmique. Étant donné qu'il y a $1.0e-11$ à $5.0e-10$ gramme de protéine cytoplasmique par cellule [157], il y a donc entre $1.0e-14$ et $5.5e-13$ gramme de ferritine par cellule. De plus, le volume d'une cellule K562 étant de $[1.7e-12, 3.4e-12]$ litre, il y a $[2.9e-3, 3.2e-1]$ gramme de ferritin par litre. Finalement, la masse de la ferritine étant comprise entre $24 \times 20\,020$ (sous-unité légère) et $24 \times 21\,226$ (sous-unité lourde) Dalton, sa masse molaire est comprise dans l'intervalle $[480480, 509424] \text{ g/mol}$. La concentration de la ferritine est donc comprise dans $[5.8e-9, 6.7e-7] \text{ mol} \cdot \text{L}^{-1}$. Nous obtenons donc la borne supérieure. 2/ Les expériences de nos collaborateurs montrent que la concentration en ferritine chez des cellules K562 est comprise dans $[1.5e-7, 1.1e-6]$ gramme de sous-unité de ferritine par milligramme de protéine cytoplasmique, ce qui correspond à une concentration en ferritine comprise dans $[3.6e-11, 2.8e-8] \text{ mol/L}$. Nous obtenons donc la borne inférieure.

$Ft_f^{eq} \in [1.0e-13, 3.1e-7] \text{ mol} \cdot \text{L}^{-1}$ Nous avons une contrainte imposant que $Ft_f^{eq} + Ft_b^{eq} < 3.1e-7 \text{ mol} \cdot \text{L}^{-1}$ (*cf. partie Contraintes du système : $Ft_f^{eq} + Ft_b^{eq}$*). La concentration en ARNm non complexé ne peut donc pas être supérieure à $3.1e-7 \text{ mol} \cdot \text{L}^{-1}$. Par ailleurs, une concentration inférieure à $1.0e-13 \text{ mol} \cdot \text{L}^{-1}$ représente moins d'un ARNm par cellule. Nous utilisons cette valeur comme borne inférieure de l'intervalle.

$Ft_b^{eq} \in [1.0e-13, 3.1e-7] \text{ mol} \cdot \text{L}^{-1}$ Voir Ft_f^{eq} .

$Fe^{eq} \in [1.0e-10, 2.0e-6] \text{ mol} \cdot \text{L}^{-1}$ La concentration de fer mobilisable par un chélateur tel que la calcéine dans les cellules K562 est de l'ordre de 2 micro molaire [132]. Nous supposons que la concentration en fer à l'état stationnaire n'est pas inférieure à $1.0e-10 \text{ mol} \cdot \text{L}^{-1}$.

$IRP^{eq} \in [1.0e-13, 1.0e-6] \text{ mol} \cdot \text{L}^{-1}$ Nous avons une contrainte imposant que la concentration d'IRP actif est comprise dans $[1.0e-11, 1.0e-6] \text{ mol} \cdot \text{L}^{-1}$ (*cf. partie Contraintes à l'état stationnaire*). La quantité d'IRP actif correspondant à la somme des variables IRP^{eq} , Ft_b^{eq} , $FPN1a_b^{eq}$, $TfR1_b^{eq}$, $IRE3_b^{eq}$ et $IRE5_b^{eq}$, nous n'obtenons qu'une borne supérieure. Une concentration inférieure à $1.0e-13 \text{ mol} \cdot \text{L}^{-1}$ correspondant à moins d'une molécule par cellule, nous considérons cette valeur comme la borne inférieure.

$FPN1a_p^{eq} \in [1.0e-13, 1.0e-5] \text{ mol} \cdot \text{L}^{-1}$ En l'absence d'information, nous considérons que la concentration d'exporteur est comprise entre $1.0e-13$ et $1.0e-5 \text{ mol} \cdot \text{L}^{-1}$.

$FPN1a_f^{eq} \in [1.0e-13, 1.2e-9] \text{ mol} \cdot \text{L}^{-1}$ Nous avons une contrainte imposant que la concentration en ARNm total de la ferroportine (donc libre+complexé) soit inférieure à

$1.2\text{e-}9 \text{ mol}\cdot\text{L}^{-1}$ (cf. partie Contraintes du système : $FPN1a_f^{eq} + FPN1a_b^{eq}$). La concentration de l'ARNm non complexé avec un IRP ne peut donc pas être supérieure à cette valeur. Nous obtenons la borne supérieure de l'intervalle. Une concentration inférieure à $1.0\text{e-}13 \text{ mol}\cdot\text{L}^{-1}$ correspond à moins d'un ARNm par cellule. Nous obtenons donc la borne inférieure.

$FPN1a_b^{eq} \in [1.0\text{e-}13, 1.2\text{e-}9] \text{ mol}\cdot\text{L}^{-1}$ Voir $FPN1a_f^{eq}$.

$TfR1_p^{eq} \in [3.0\text{e-}8, 3.0\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$ Chez des cellules K562, il y a 140000 récepteurs de transferrine [134]. En considérant que le volume d'une cellule K562 est compris dans $[1.7\text{e-}12, 3.4\text{e-}12]$ litre, nous obtenons que la concentration du récepteur est comprise dans $[6.8\text{e-}8, 1.4\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$. Nous considérons un intervalle autour de cette valeur.

$TfR1_f^{eq} \in [1.0\text{e-}13, 9.1\text{e-}10] \text{ mol}\cdot\text{L}^{-1}$ Nous avons une contrainte imposant que la concentration de $TfR1_r^{eq}$ est comprise dans $[4.9\text{e-}13, 9.1\text{e-}10] \text{ mol}\cdot\text{L}^{-1}$ (cf. partie Contraintes à l'état stationnaire). Nous obtenons donc la borne supérieure de l'intervalle de $TfR1_f^{eq}$. Une concentration inférieure à $1.0\text{e-}13 \text{ mol}\cdot\text{L}^{-1}$ représentant moins d'un ARNm par cellule, nous considérons cette valeur comme borne inférieure.

$TfR1_b^{eq} \in [1.0\text{e-}13, 9.1\text{e-}10] \text{ mol}\cdot\text{L}^{-1}$ Voir $TfR1_f^{eq}$.

$IRE3_f^{eq} \in [1.0\text{e-}13, 3.1\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$ Nous avons une contrainte imposant que $IRE3_f^{eq} + IRE3_b^{eq}$ soit inférieure à $3.1\text{e-}7 \text{ mol}\cdot\text{L}^{-1}$ (cf. partie Ré-écriture de contraintes). La concentration de $IRE3_f$ ne peut donc pas être supérieure à cette valeur. Nous obtenons alors la borne supérieure. Une concentration inférieure à $1.0\text{e-}13 \text{ mol}\cdot\text{L}^{-1}$ correspond à moins d'un ARNm par cellule. Nous prenons cette valeur comme borne inférieure.

$IRE3_b^{eq} \in [1.0\text{e-}13, 3.1\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$ Voir $IRE3_f^{eq}$.

$IRE5_f^{eq} \in [1.0\text{e-}13, 3.1\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$ Nous avons une contrainte imposant que $IRE5_f^{eq} + IRE5_b^{eq}$ soit inférieure à $3.1\text{e-}7 \text{ mol}\cdot\text{L}^{-1}$ (cf. partie Ré-écriture de contraintes). La concentration de $IRE5_f$ ne peut donc pas être supérieure à cette valeur. Nous obtenons alors la borne supérieure. Une concentration inférieure à $1.0\text{e-}13 \text{ mol}\cdot\text{L}^{-1}$ correspond à moins d'un ARNm par cellule. Nous prenons cette valeur comme borne inférieure.

$IRE5_b^{eq} \in [1.0\text{e-}13, 3.1\text{e-}7] \text{ mol}\cdot\text{L}^{-1}$ Voir $IRE5_f^{eq}$.

$Tf_{sat} = 0.3$ Le taux de saturation de la transferrine chez l'humain est compris entre 25 et 45% [131]. Nous prenons 30%.

A.3 Contraintes à l'état stationnaire

$drs_{TfR1} < dr_{TfR1}$ La fixation d'un IRP sur l'ARNm de TfR1 le stabilise. La constante de vitesse de dégradation de l'ARNm complexé est donc plus faible que celle de l'ARNm non complexé.

$drs_{IRE3} < dr_{IRE3}$ La fixation d'un IRP sur un IRE en 3'-UTR stabilise l'ARNm. La constante de vitesse de dégradation de cet ARNm complexé est donc plus faible que celle d'un ARNm non complexé.

$k_{Fe_{cons}} \geq k_{Fe_{export}} \cdot FPN1a_p^{eq}$ Chez les cellules érythroïdes, la moitié des transcrits de FPN1 contient un IRE [70]. L'export de fer par ces protéines étant inclus dans le terme $k_{Fe_{cons}} \cdot Fe$, ce terme doit être au moins aussi grand que celui de l'export de fer par les FPN1a possédant un IRE. En simplifiant les deux termes par Fe , nous obtenons la relation décrite.

$Fe^{eq} > \theta_{Fe \rightarrow IRP}$ La concentration en fer à l'état stationnaire est supérieure au seuil à partir duquel le mécanisme IRP intervient.

$$IRP^{eq} + Ft_{-b}^{eq} + FPN1a_{-b}^{eq} + TfR1_{-b}^{eq} + IRE3_{-b}^{eq} + IRE5_{-b}^{eq} \in [1.0e-11, 1.0e-6] \text{ mol}\cdot\text{L}^{-1}$$

Dans le foie du rat, la quantité d'IRP actif est autour de 0.16 pmol/mg de protéine cytoplasmique [133]. En considérant qu'il y a [1.0e-11, 5.0e-10] g de protéine cytoplasmique par cellule et que le volume d'une cellule moyenne est de 4.2e-12 litre, nous obtenons donc que la concentration d'IRP actif est comprise dans [3.8e-10, 1.9e-8] mol·L⁻¹. Néanmoins, comme la mesure est faite chez le rat, nous élargissons l'intervalle et considérons que la concentration d'IRP actif est comprise entre 1.0e-11 et 1.0e-6 mol·L⁻¹.

$Ft_{-f}^{eq} + Ft_{-b}^{eq} \in [1.3e-8, 3.1e-7] \text{ mol}\cdot\text{L}^{-1}$ Selon les expériences faites par nos collaborateurs, chez des cellules K562, la concentration de l'ARNm de la ferritine est comprise dans [2.7e4, 3.2e5] transcrits par cellule. Le volume d'une cellule K562 étant compris dans [1.7e-12, 3.4e-12] litre, la concentration en ARNm de ferritine est alors comprise dans [1.3e-8, 3.1e-7] mol·L⁻¹.

$FPN1a_{-f}^{eq} + FPN1a_{-b}^{eq} < 1.2e-9 \text{ mol}\cdot\text{L}^{-1}$ Selon les expériences faites par nos collaborateurs, chez des cellules K562, la quantité de transcrit de FPN1a est au plus de 1200 transcrits par cellule. Le volume d'une cellule K562 étant compris dans [1.7e-12, 3.4e-12] litre, nous obtenons que la concentration en ARNm de ferroportine est inférieure à 1.2e-9 mol·L⁻¹.

$TfR1_{-f}^{eq} + TfR1_{-b}^{eq} \in [4.9e-13, 9.1e-10] \text{ mol}\cdot\text{L}^{-1}$ Selon les expériences faites par nos collaborateurs, chez des cellules K562, il y a entre 1 et 950 transcrits d'ARNm du récepteur de la transferrine par cellule. Le volume d'une cellule K562 étant compris dans [1.7e-12, 3.4e-12] litre, nous obtenons que la concentration en ARNm de TfR1 chez des K562 est compris dans [4.9e-13, 9.1e-10] mol·L⁻¹.

$\frac{FPN1a_{-f}^{eq} + FPN1a_{-b}^{eq}}{TfR1_{-f}^{eq} + TfR1_{-b}^{eq}} < 6.5e-1$ 1/ Selon les données de nos collaborateurs, chez des cellules K562, le rapport de la concentration d'ARNm de FPN1a sur celle de l'ARNm de TfR1 est autour de 0.2. En agrandissant l'intervalle, nous obtenons que ce rapport est compris dans [6.5e-2, 6.5e-1]. 2/ Des mesures faites sur d'autres lignées cellulaires indiquent que, selon la lignée, la concentration de l'ARNm de FPN1a peut être largement plus faible. Nous ne considérons alors que la borne supérieure de l'intervalle. Nous obtenons donc la relation décrite.

$\frac{Ft_{-f}^{eq} + Ft_{-b}^{eq}}{TfR1_{-f}^{eq} + TfR1_{-b}^{eq}} \in [100, 1000]$ Selon les données de nos collaborateurs, chez des cellules K562, en situation contrôle, la concentration de l'ARNm de la ferritine est 100 à 1000 fois plus importante que celle de TfR1.

$Ft_{-f}^{eq} + Ft_{-b}^{eq} > IRE3_{-f}^{eq} + IRE3_{-b}^{eq}$ Selon les données de nos collaborateurs, la quantité de transcrit de la ferritine est largement supérieure à celle de autres transcrits.

$Ft_{-f}^{eq} + Ft_{-b}^{eq} > IRE5_{-f}^{eq} + IRE5_{-b}^{eq}$ Selon les données de nos collaborateurs, la quantité de transcrit de la ferritine est largement supérieure à celle de autres transcrits.

$\frac{TfR1_{-b}^{eq}}{TfR1_{-f}^{eq}} < 1000$ Selon Chan *et al.* [148], pour les ARNm de TfR1, le rapport complexé/non complexé semble très confortable à 1000. Nous considérons donc une borne supérieure pour ce rapport.

$\frac{dr_{-TfR1} \cdot TfR1_{-f}^{eq} + drs_{-TfR1} \cdot TfR1_{-b}^{eq}}{TfR1_{-f}^{eq} + TfR1_{-b}^{eq}} \in [7.0e-6, 7.0e-5] \text{ s}^{-1}$ 1/ La constante de vitesse de dégradation de l'ARNm total de TfR1 est comprise dans [7.0e-6, 7.0e-5] s⁻¹ (*cf. intervalle de p_{-TfR1}*). 2/ Cette dégradation correspond à la somme des dégradations de

$TfR1-f$ et de $TfR1-b$. Nous avons donc la relation :

$$[7.0e-6, 7.0e-5] \cdot (TfR1-f^{eq} + TfR1-b^{eq}) = dr_{TfR1} \cdot TfR1-f^{eq} + drs_{TfR1} \cdot TfR1-b^{eq}$$

d'où nous tirons la relation décrite.

$t_{Ft} \cdot Ft-f^{eq} \in [5.5e-12, 5.5e-11] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ Dans l'article de Chu *et al.* [158], chez des cellules HeLa, la vitesse de traduction de sous-unité de la ferritine est estimée à [1.1e-11, 1.4e-11] mol de sous-unité par heure pour 5.0e7 cellules Hela. Le volume d'une cellule HeLa étant de 4.2e-12 litre, nous obtenons que la vitesse de traduction de sous-unité de la ferritine est comprise dans [1.5e-11, 1.8e-11] mol·(L·s)⁻¹. Nous considérons un intervalle autour de cet intervalle.

A.4 Ré-écriture de contraintes

Cette section présente des contraintes qui ont été combinées. Cela permet d'une part de réduire les intervalles de certaines variables, et d'autre part d'obtenir de nouvelles contraintes à rajouter au système.

$\frac{p_{Ft} \cdot dr_{FPN1a}}{dr_{Ft} \cdot v_{FPN1a}} \in [8.0e-18, 7.6e-13] \text{ mol} \cdot (\text{L} \cdot \text{s})^{-1}$ 1/ Si nous n'avions pas réécrit l'équation

de $FPN1a-f$ afin d'intégrer directement une donnée expérimentale, nous aurions la relation

$p_{FPN1a} = dr_{FPN1a} \cdot FPN1a-r^{eq}$ (obtenue en additionnant les équations à l'équilibre de $FPN1a-f$ et $FPN1a-b$). De plus, nous savons que, à l'état stationnaire, $FPN1a-r^{eq}$ est compris

dans [2.0e-13, 1.2e-9] mol·L⁻¹ (*cf. partie Contraintes du système : FPN1a-f + FPN1a-b*) et que la constante de vitesse de dégradation dr_{FPN1a} est comprise dans [4.0e-5, 4.0e-4] s⁻¹.

Nous obtenons donc que p_{FPN1a} est compris dans [8.0e-18, 7.6e-13] mol·(L·s)⁻¹. 2/ Nous avons remplacé le paramètre cinétique de traduction de l'ARNm de FPN1a (p_{FPN1a}) par

l'expression : $\frac{p_{Ft} \cdot dr_{FPN1a}}{dr_{Ft} \cdot v_{FPN1a}}$. Nous en déduisons donc cette relation.

$dr_{TfR1} + [0.0, 1.0e3] \cdot drs_{TfR1} \in [7.0e-6, 7.0e-2] \text{ s}^{-1}$ La constante de vitesse de dégradation de $TfR1-r$ comprise dans [7.0e-6, 7.0e-5] s⁻¹ (*cf. intervalle pour p_{TfR1}*). À l'état stationnaire en présence de fer, nous avons donc la relation :

$$[7.0e-6, 7.0e-5] \cdot TfR1-r^{eq} = dr_{TfR1} \cdot TfR1-f^{eq} + drs_{TfR1} \cdot TfR1-b^{eq}$$

De plus, nous avons la relation $\frac{TfR1-b^{eq}}{TfR1-f^{eq}} < 1.0e3$ (*cf. Partie contrainte à l'état stationnaire*).

Nous en déduisons les deux relations $TfR1-r^{eq} = [1, 1001] \cdot TfR1-f^{eq}$ et $TfR1-b^{eq} = [0, 1000] \cdot TfR1-f^{eq}$. En remplaçant dans l'équation précédente, nous obtenons la relation :

$$[7.0e-6, 7.0e-5] \cdot [1, 1001] \cdot TfR1-f^{eq} = dr_{TfR1} \cdot TfR1-f^{eq} + drs_{TfR1} \cdot [0.0, 1.0e3] \cdot TfR1-f^{eq}$$

$TfR1-f^{eq}$ n'étant pas nul, nous divisons les deux membres de la relation par $TfR1-f^{eq}$ et nous en déduisons les deux relations suivantes :

- $dr_{TfR1} \in [7.0e-6 - 1.0e3 \cdot drs_{TfR1}, 7.0e-2] \text{ s}^{-1}$

- $drs_{TfR1} > 7.0e-9 - \frac{dr_{TfR1}}{1.0e3} \text{ s}^{-1}$

$IRE3-f^{eq} + IRE3-b^{eq} < 3.1e-7 \text{ mol} \cdot \text{L}^{-1}$ 1/ Nous avons une contrainte imposant que $IRE3-f^{eq} + IRE3-b^{eq}$ soit inférieur à $Ft-f^{eq} + Ft-b^{eq}$ (*cf. partie Contraintes à l'état stationnaire*). 2/ Nous avons une contrainte imposant que $Ft-f^{eq} + Ft-b^{eq}$ soit inférieur à 3.1e-7 mol/L (*cf. partie Contraintes à l'état stationnaire*). Nous en déduisons donc que $IRE3-f^{eq} + IRE3-b^{eq}$ est inférieur à 3.1e-7 mol·L⁻¹.

$IRE5_f^{eq} + IRE5_b^{eq} < 3.1e-7 \text{ mol}\cdot\text{L}^{-1}$ En utilisant un raisonnement similaire à celui utilisé pour obtenir la contrainte $IRE3_f^{eq} + IRE3_b^{eq} < 3.1e-7 \text{ mol}\cdot\text{L}^{-1}$ (cf. partie Ré-écriture de contraintes), nous obtenons que $IRE5_f^{eq} + IRE5_b^{eq}$ est inférieur à $3.1e-7 \text{ mol}\cdot\text{L}^{-1}$.

Annexe B

Modélisation en ASP des réseaux de Thomas

B.1 Introduction

Dans la section 1.3.1, nous avons introduit le formalisme des réseaux discrets dits *de Thomas*, et donné un aperçu des limitations de ce formalisme. Nous avons en particulier mentionné qu'il ne s'applique pas aux réseaux biochimiques car les lois de conservation ne peuvent pas être prises en compte. Il reste cependant très utile pour une première analyse qualitative de réseaux de régulation transcriptionnelle et est aujourd'hui largement utilisé, non seulement pour les réseaux transcriptionnels, mais aussi pour des voies de phosphorylation. Il est en effet possible d'appliquer des techniques dites de *model-checking* pour une étude exhaustive de l'espace des paramètres (ceux-ci appartenant à un domaine entier fini). L'esprit est le même que celui de l'approche STL, à savoir formaliser les observations expérimentales par des formules temporelles et ensuite utiliser ces dernières pour sélectionner les jeux de valeurs de paramètres correspondant à des comportements qui satisfont les contraintes (*model-checking*). Avec le formalisme de Thomas, nous pouvons, en plus du *model-checking*, utiliser des techniques de programmation par contraintes, qui sont a priori plus puissantes. En effet, elles mettent en œuvre des techniques de déduction qui sont susceptibles d'éliminer des parties importantes de l'espace de recherche, sans avoir à effectuer de simulation pour chaque jeu de paramètres. Avec ce type de techniques, des fonctionnalités assez sophistiquées peuvent être développées, telles que l'inférence de paramètres, l'inférence de propriétés valides pour un ensemble de solutions, ou même la révision automatique de modèle [30]. Ces techniques de déduction sont très difficiles à mettre en œuvre sur des systèmes dynamiques différentiels, cependant, des tentatives existent [159]. Pour ces raisons, et du fait qu'un des deux principaux objectifs de cette thèse est le développement de nouvelles méthodes et nouveaux outils pour la biologie systémique, nous avons étudié une approche déclarative, basée sur l'utilisation de contraintes, pour l'étude qualitative des réseaux de régulation génique, dans le formalisme des réseaux de Thomas.

Cette approche propose une modélisation de ces réseaux, utilisant le paradigme logique ASP (*Answer Set Programming*). Ce paradigme, basé sur une logique *non monotone*¹⁸, fournit les modèles¹⁹ *minimaux*²⁰ d'un programme. Un programme ASP est un ensemble de règles de la

18. Une logique est dite non-monotone lorsque, à partir d'un ensemble de données, l'ajout de nouvelles données ne permet plus d'obtenir toutes les déductions obtenues avant cet ajout.

19. Dans cette annexe, le terme « modèle » représente un ensemble d'atomes vérifiant les règles d'un programme ASP.

20. Un modèle n'est pas minimal si la suppression d'un ou plusieurs atomes de ce modèle constitue un autre

forme $a_0 \rightarrow a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n$ où $0 \leq m \leq n$ et $\forall i | 0 \leq i \leq n, a_i$ est un atome. Les règles pour lesquelles m est différent de n sont des règles contenant des *défauts*.

Cette annexe est issue d'un article récemment accepté pour la conférence IWBBIO [160], que nous avons amélioré pour présenter les dernières évolutions. Dans cette annexe, nous montrons les avantages de l'utilisation d'ASP. Par exemple, l'utilisation de règles *par défaut* permet d'intégrer aisément des connaissances partielles sur le modèle. Cet article commence par présenter la modélisation formelle des réseaux de Thomas et des données biologiques, comme, par exemple, des données comportementales. Puis nous présentons trois exemples de systèmes biologiques pour lesquels cette approche déclarative est utilisée. Ensuite, nous introduisons une modélisation plus optimisée, permettant, d'une part, de prendre en compte des propriétés *généralement* vraies et, d'autre part, de proposer une gestion automatique des incohérences. Pour cela, nous créons une méthode pour construire des *conjonctions de défauts*, et nous l'appliquons à la modélisation des réseaux de Thomas. En dernier lieu, nous donnons une courte description des travaux en cours sur l'implémentation de formules CTL en ASP et de compositions d'interactions de gènes (nommées multiplexes).

B.2 Présentation des travaux

modèle. En d'autres termes, on ne peut enlever aucun atome d'un modèle si celui-ci est minimal.

Logical Modeling and Analysis of Regulatory Genetic Networks in a Non Monotonic Framework

N. Mobilia¹, A. Rocca¹, S. Chorlton², É. Fanchon¹, and L. Trilling¹

¹ Laboratoire TIMC-IMAG, Université de Grenoble, France

² Department of Medicine, McMaster University, Hamilton, Canada

Abstract. We present a constraint based declarative approach for analyzing qualitatively genetic regulatory networks (GRNs) with the discrete formalism of R. Thomas. For this purpose, we use the logic programming technology ASP (Answer Set Programming) whose related logic is non monotonic. Our aim is twofold. First, we give a formal modeling of both Thomas' GRNs and biological data like experimental behaviors and gene interactions and we evaluate the declarative approach on three real biological applications. Secondly, for taking into account both gene interaction properties which are only **generally** true and automatic inconsistency repairing, we introduce an optimized modeling which leads us to exhibit new logical expressions for the conjunction of defaults and to show that they can be applied safely to Thomas' GRNs. Moreover, we give a short description on related ongoing works on the ASP implementations of general temporal CTL formulas and of a strong composition of gene interactions called *multiplex*.

Keywords: computational systems biology · gene networks · discrete modeling · AI-oriented declarative approach · non monotonic logic · Answer Set Programming

1 Introduction

Mathematical modeling and simulation tools may help to understand how complex genetic regulatory networks (GRNs), composed of many genes and their intertwined interactions, control the functioning of living systems. They provide a framework to unambiguously describe the network structure and to infer predictions of the dynamical behavior of the system.

The typical model building cycle starts with gathering existing knowledge on a biological system and formulating working hypotheses, on the basis of which a model formalism is chosen and the structure of the GRN is defined. The development of the dynamical model and its parametrization lead to an initial model, whose predictions are confronted with experimental data. This often reveals inconsistencies, and calls for a revision of the structure of the GRN and/or the parameter values of the model. The process is repeated iteratively until the validation step is considered satisfactory. The generate-and-test approach underlying the above-mentioned method demands a large number of simulations to be carried out and usually leads to the formulation of a unique model consistent with biological data. In this paper, we adopt an alternative method for the systematic construction and analysis of models of GRNs by means of an artificial intelligence oriented *declarative* approach. The models are developed using the

formalism of R. Thomas [9], which offers an appropriate discrete description of GRNs, as most available data on regulatory interactions are qualitative. Instead of instantiating the models as in classical modeling approaches, all possible knowledge on the network structure and its dynamics (e.g., existence of cycles or stationary states, response of the network to environmental or genetic perturbations) is expressed *intensionally* as constraints, i.e. logical formulas. Without resorting to numerous simulations and manipulations, first, the compatibility of the network structure with the biological constraints is determined and, second, in case all the constraints are satisfied, all properties, expressed in a predefined language, that are common to all consistent models can be automatically deduced [12], for the profit of biologists. This intensional representation is well suited for biological data which are often incomplete: typically, kinetic parameters are unknown. Furthermore, in case of inconsistency, an automatic repairing can be applied.

We use for that purpose the logic programming technology ASP [161] which is based on a non monotonic logic defined with *stable* models. The aim of this paper is to show the benefits provided by ASP for the declarative approach of Thomas' GRNs. We pay special attention to modeling methods to take advantages of the ASP non monotonic feature, for example for tackling potential inconsistencies and for expressing gene interaction rules that are **generally** true.

This paper is organized as follows. ASP is introduced in Section 2 and the logical specifications of Thomas' GRNs together with biological data description (including behaviors, interactions between genes and mutant models) and typical queries for analyzing GRN, in Section 3. Then, three illustrating applications are presented in Section 4. An optimized ASP modeling of Thomas' GRNs exemplifying the non monotonic aspects of ASP is described in Section 5. This section introduces the notion of conjunction of defaults. Original logical formulas to express them and a methodology to construct them is developed and applied in Section 6. Finally, we conclude in Section 7. We describe shortly, in Section 7.2, ongoing works on the ASP implementation of all logical temporal CTL formulas (needed for expressing general behaviors) and on general composition of interactions of genes (involving conjunctions of interactions called *multiplexes*).

2 Answer Set Programming (ASP)

Here is a short presentation based on Gebser *et al.* [162] which proposes the *gringo* language for expressing ASP programs.

A logical ASP program is a finite set of rules:

$$a_0 \leftarrow a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n$$

where $0 \leq m \leq n$ and $\forall i \mid 0 \leq i \leq n, a_i$ is an atom. For any rule r , $\text{head}(r) = a_0$ is the head of the rule, and $\text{body}(r) = \{a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n\}$ is the body of the rule. If $\text{head}(r)$ is empty, r is called an *integrity constraint*. If $\text{body}(r)$ is empty, r is a *fact*.

Let $A = \{a_i \mid i \in [0, n]\}$ be the set of atoms, $\text{body}^+(r) = \{a \in A \mid a \in \text{body}(r)\}$ and $\text{body}^-(r) = \{a \in A \mid \text{not } a \in \text{body}(r)\}$. A set $X \subseteq A$ is an *answer set (AS)* or stable

model of a program P if X is the minimal model³ of the following program, called *reduct*:

$$P^X = \{head(r) \leftarrow body^+(r) \mid r \in P, body^-(r) \cap X = \emptyset\}$$

For example, let E be the following ASP program where “ \leftarrow ” is represented by “ $:-$ ”:

```
a :- not b, c.
c.
```

Then, let the set $X = \{a, c\}$. The corresponding reduct is $E^X = \{a \leftarrow c, c\}$ and its minimal model is $\{a, c\}$. Then X is an answer set of E . Let an other set $X' = \{a, b, c\}$. The corresponding reduct is $E^{X'} = \{c\}$, and its minimal model is $\{c\}$. Then X' is not an answer set of E .

A rule r such that $m < n$ (i.e. $body^-(r) \neq \emptyset$) is called a *default* rule [163]. The first rule of this example is typical of this kind of rule. It expresses that generally (in the absence of knowledge on b) a is implied by c . But if b holds because of additional knowledge, this rule is no longer applicable. For instance, by addition of the fact b . to this example, we get the answer set $\{b, c\}$ exemplifying the non monotonic character of ASP: a does not hold any more, without leading to an inconsistency.

Rules in the *gringo* language [162] are extended for accepting heads which are disjunctions of literals (exclusive unless both literals are proven elsewhere) using the operator “ $|$ ”. Furthermore, *gringo* provides logical variables and functional terms, in a limited way (so that the program can be transformed in an equivalent finite propositional one). It also provides cardinality constraints on the number of true literals. If we impose the constraint $u\{l_1, \dots, l_n\}v$ we obtain only models such that the number of true literals l_i is bigger than (or equal to) u (0 by default) and smaller than (or equal to) v (n by default). Moreover, this formalism allows the expression of conditional enumerations of literals through the symbol “ $:$ ”, as conjunctions (resp. disjunctions) in the body (resp. head) of a rule. For example, in the following program:

```
dom(0). dom(1).
all_true :- p(X) : dom(X).
q(X) : dom(X) :- one_of.
at_least_one_true :- 1{p(X) : dom(X)}.
```

the two first lines express that `all_true` is deduced if $p(0)$ and $p(1)$ hold. These lines are, therefore, equivalent to the rule: `all_true :- p(0), p(1)`. and the third line equivalent to `q(0) | q(1) :- one_of`. The fourth line expresses that `at_least_one_true` is deduced if a least one among $p(0)$ and $p(1)$ holds.

Para-logic operators are also provided for maximizing or minimizing the number of atoms true among a set of atoms. Asserting `#maximize{f_1, \dots, f_n}` will produce only models with the highest number of f_i true.

The solver [162] we use proceeds in two steps to compute the answer sets of a program P . First, a *grounder* substitutes the variables of the program by terms without free variables, and consequently produces a propositional program \mathbf{P} corresponding to P . In the second step, a solver computes the answer sets of \mathbf{P} . This motivates the

³ A set of Horn clauses that does not contain negative literals always has a unique minimal model (a logical model is minimal when removing one or more atoms from it cannot provide a model). The reduct P^X is a such set, consequently, answer sets are minimal (see Section 5.2).

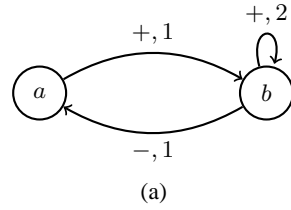
programmer to reduce as far as possible the number of resulting Boolean variables and rules subject to a big expansion (see Section 5.2).

3 Thomas' GRNs and the Declarative Approach

In Section 3.1, we specify Thomas' GRNs. Biological constraints and typical queries for constructing and analyzing models of GRNs are described in Section 3.2.

3.1 Thomas' GRN Specification

A common representation of a GRN is to view it as an *interaction graph*, where nodes represent genes and arrows represent interactions between genes. An arrow $j \rightarrow i$ is labeled with the sign of the regulatory influence (to indicate whether the gene i is activated or inhibited by the product J of gene j), and with the index r of the threshold concentration θ_j^r related to the control, by the protein J , of the expression of gene i . For a given gene i , an *interaction* from the gene j to the gene i , labeled by the threshold θ_j^r , is written (j, θ_j^r) . A simple example of interaction graph for a two-gene network is shown in Fig. 1(a).



$$\begin{aligned}
 X_a &= K_a * s^-(x_b, \theta_b^1) + \\
 &\quad K_a^b * s^+(x_b, \theta_b^1) \\
 X_b &= K_b * s^-(x_a, \theta_a^1) s^-(x_b, \theta_b^2) + \\
 &\quad K_b^a * s^+(x_a, \theta_a^1) s^-(x_b, \theta_b^2) + \\
 &\quad K_b^b * s^-(x_a, \theta_a^1) s^+(x_b, \theta_b^2) + \\
 &\quad K_b^{ab} * s^+(x_a, \theta_a^1) s^+(x_b, \theta_b^2)
 \end{aligned}$$

(b)

Fig. 1. (a) Interaction graph corresponding to a GRN of two genes. The product of gene a stimulates the expression of gene b , while the product of gene b inhibits expression of a . In addition, b activates its own expression. The label “ $-$, 1” from gene b to a means that b inhibits a expression when b is above its threshold θ_b^1 . (b) Focal equations relating a state (x_a, x_b) and its focal state (X_a, X_b) .

The protein concentrations are modeled by variables taking discrete values. For a protein, each discrete value represents an interval between two consecutive thresholds. For instance, $x_j = 0$ indicates a concentration of protein J lower than (the concentration modeled by the discrete value of) the lowest threshold of J , say θ_j^m , whereas $x_j = 1$ means that the concentration of J is greater than or equal to (the concentration modeled by the discrete value of) θ_j^m and lower than $\theta_j^m + 1$. Thus, a *state* S of a network of n genes is represented by a vector of discrete protein concentrations: $S = (x_1, \dots, x_n)$. Then, the dynamical behavior of a GRN is represented in terms of an oriented graph called *state transition graph*, where each node represents a specific state of the system and the arrows represent *transitions* between a state and its possible successors.

A specific attractor value called *focal state*, (X_1, \dots, X_n) , is associated to a given state S . It represents the expression levels toward which the genes tend to evolve (see *focal equation* (1) for a precise definition). A successor $S' = (x'_1, \dots, x'_n)$ of S in the graph is deduced from S by comparing the value of each variable x_i with that of its focal state. The transition of S to S' is assumed to be asynchronous, in the sense that at most one variable x_i is updated at a time. If the variable x_i is updated, the formal relationship between these states is expressed as follows: $x'_i = x_i + 1$ if $X_i > x_i$ and $x'_i = x_i - 1$ if $X_i < x_i$. If no logical variable x_i is updated then the focal state of S is equal to S and S is its own successor: in that case S is said to be *steady* (or *stationary*).

The focal state value X_i of gene i depends on the state S of the network and in particular, on a set of conditions regarding the presence or absence of activators and inhibitors of gene i . For the simple example in Fig. 1, the focal state value X_a of gene a depends on the influence of B (the product of b) on a , that is, if the concentration of B is below ($x_b = 0$) or above its first threshold value. Such interactions are expressed by means of products of step functions of the form: $s^+(x_j, \theta_j^r) = 1$ if $x_j \geq \theta_j^r$ else 0, $s^-(x_j, \theta_j^r) = 1$ if $x_j < \theta_j^r$ else 0.

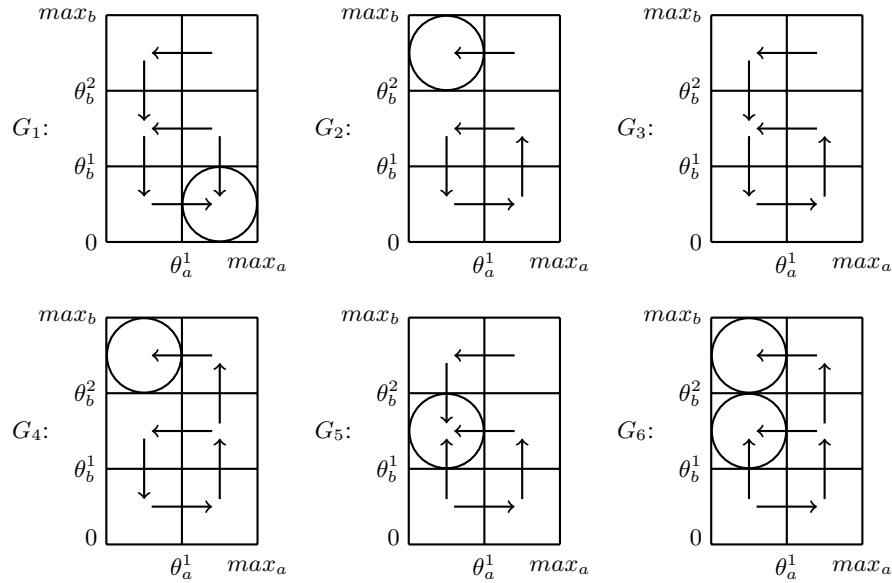


Fig. 2. Transition graphs G_1 to G_6 corresponding to the eight sets of instantiated kinetic parameters satisfying all the observability and additivity constraints associated to the example in Fig. 1, with $\theta_b^1 < \theta_b^2$. Arrows represent possible transitions between states represented by boxes. A circle in a state means that this state is steady. Each graph corresponds to a specific set of instantiated kinetic parameters, except G_1 and G_3 which correspond to two sets of instantiated kinetic parameters.

We will call *cellular context* of gene i any set of states which are equivalent with respect to i for regulation purpose. For example, if i is influenced by only one gene j associated to the threshold θ_j^1 , there are only two possible cellular contexts for i , depending on whether x_j is below or above θ_j^1 . If i is the target of two interactions, four cellular contexts have to be distinguished for i . More formally, let $[(j_1, \theta_{j_1}^{r_1}), \dots, (j_p, \theta_{j_p}^{r_p})]$ be the ordered list of interactions acting on gene i . A cellular context for i is represented as a product $c_i(\sigma) = s^{\sigma_1}(x_{j_1}, \theta_{j_1}^{r_1}) * \dots * s^{\sigma_p}(x_{j_p}, \theta_{j_p}^{r_p})$ defining a set of conditions, where $\sigma \in E^i$ with $E^i = \{(\sigma_1, \dots, \sigma_p) | k \in [1, p], \sigma_k \in \{+, -\}\}$. In other words, a cellular context is identified by a p -tuple σ of signs, which specify the position of the set of states with respect to the p thresholds belonging to the interaction set. From the above definition it follows that the set of cellular contexts with respect to any gene i constitutes a partition of the state space. For the example of Fig. 1, let $[(a, \theta_a^1), (b, \theta_b^2)]$ be the list of interactions acting on b , then the cellular context $c_b((+, -))$ is $s^+(a, \theta_a^1) * s^-(b, \theta_b^2)$.

Each cellular context $c_i(\sigma)$ is associated to a *logical kinetic parameter* (or *parameter*) defining the focal value of gene i when the network state belongs to that context. We will denote such parameters by $K_i^{l(\sigma)}$, where $l(\sigma)$ is the set $\{j_k | k \in [1, p], \sigma_k = +\}$. In this way, the cellular context associated to a logical kinetic parameter appears in its notation. For the simple example in Fig. 1, the parameters of gene b are: K_b , K_b^b , K_b^a and K_b^{ab} . They respectively are the focal value of b : when the concentrations of A and B are under their thresholds, when the concentration of B only is above its threshold, when the concentration of A only is above its threshold, and when both concentrations are above their thresholds.

Given the cellular contexts defined above and their associated logical kinetic parameters, the focal value X_i is specified by the following focal equation:

$$X_i = \sum_{\sigma \in E^i} \left(K_i^{l(\sigma)} \cdot c_i(\sigma) \right) \quad (1)$$

In the declarative approach, some parameters may not be instantiated (i.e., unknown). When all these parameters are instantiated, the focal equations define a unique instantiated model. The focal equations corresponding to the simple network example are shown in Fig. 1(b). For the following values of parameters: $K_a = 1$, $K_a^b = 0$, $K_b = 0$, $K_b^b = 1$, $K_b^a = 0$, and $K_b^{ab} = 2$, we obtain the transition graph G_1 in Fig. 2. The state $[1, 1]$ belongs to the cellular context $c_a((+))$ and to the cellular context $c_b((+, -))$. The focal state of $[1, 1]$ is therefore $[K_a^b, K_b^a] = [0, 0]$. It follows that the successors of $[1, 1]$ are $[1, 0]$ and $[0, 1]$.

The ASP modeling of successor relationship and focal equations is described in Section 5.

3.2 Biological Data Modeling and Analysis of GRN

We focus here on interactions between genes and observed experimental behaviors of the network. We also describe the modeling of other data like those obtained from *mutant* networks and the way to get predictions and specific networks by optimization criteria.

Interaction Signs It is mandatory in a logical framework to define rigorously the meaning of the signs “+” or “-” on edges of an interaction graph (they may be loosely interpreted in the literature). We propose here a rather general and intelligible definition in the form of conditions called *observability constraints* and *additivity constraints* (not to be confused with the ASP integrity constraints).

A “+” (resp. “-”) sign on an edge targeting a gene is understood as implying the existence of a couple of states (S_1, S_2) , with S_1 just below the edge threshold, such that 1) S_2 differs from S_1 only by a +1 change in the value of the source gene, and 2) S_2 has a greater (resp. lower) focal value for the target gene than S_1 . One may see why the transition graph G_4 (Fig. 2) respects the “+” label associated with the edge $a \rightarrow b$ (Fig. 1). The state $[0, 1]$ is such that the value of the source node a is lower than the threshold θ_a^1 of this edge. This state has a neighboring state $[1, 1]$, which differs only in the value of a by a change of +1. Furthermore, the focal value of b for this neighbor is strictly higher ($K_b^a = 2$) than the one of the state $[0, 1]$ ($k_b = 0$).

As all states of a cellular context have the same focal state, the existence of states (S_1, S_2) is equivalent to the existence of cellular contexts $(c1, c2)$ of the target node which have the following properties for a “+”(resp. “-”) sign: 1) for all states in $c1$, the value of the source gene is below the edge threshold and 2) for all state S in $c2$, it exists a state S' in $c1$ which differs from S only by value of the source gene greater or equal than the edge threshold and 3) the focal value of the target gene in the context $c2$ has a greater (resp. lower) value than in context $c1$. In the transition graph G_4 , considering again the positive interaction $a \rightarrow b$, such a couple of cellular contexts of b is for $c1$ the cellular context where $a < \theta_a^1 \wedge b < \theta_b^2$ holds and for $c2$ where $a \geq \theta_a^1 \wedge b < \theta_b^2$ holds.

Then, *observability constraints* for an interaction are expressed by a union of strict inequalities between kinetic parameters of the target of this interaction, just differing by one gene. For example, the observability constraint associated to the positive interaction $a \rightarrow b$ is $(K_b < K_b^a) \vee (K_b^b < K_b^{ab})$.

Additivity constraints are considered to indicate that **generally** no inhibition (resp. activation) can exist in case of a positive (resp. negative) interaction. For example, this means that in the general case for the positive interaction $a \rightarrow b$ where there is no proven inhibition (e.g. $(K_b > K_b^a) \vee (K_b^b > K_b^{ab})$ does not hold), then the negation of this inhibition is true (e.g. $(K_b \leq K_b^a) \wedge (K_b^b \leq K_b^{ab})$ holds).

The ASP modeling of observability and additivity constraints is presented in Section 5. Their definition is given in a more general framework in Section 7.2 and in the Annex 8.

Behaviors Experimental behavioral data are expressed using constraints on *paths* which are successions of states. This is the case for modeling observed steady states, cycles or repairing behavior due to stress. The declarative approach presents a decisive advantage as information on these behaviors is usually incomplete; for example, there could exist a cycle for which only some concentrations of proteins are known throughout the cycle. Despite the lack of information, this approach may provide biologically meaningful properties regarding the kinetic parameters. Constraint modeling of paths is described with the predicate `species(N, V, I, P)` meaning that V is the expression level of the gene N at step I of the path P . For example, defining a steady state can be

done through the predicate `statpath(P)` (the two states of the path `P` of length 2 are equal):

```
statpath(P) :- path(P), length(2,P), succegl(N,P):node(N).
where succegl(N,P) is true if, at the two first steps of the path P, the concentrations
of the species N are equal, length(L,P) means that the path P has L steps and
path(P) means that P is a path. Enforcing the existence of the steady state ss can
then be performed with the two facts and the integrity constraint which follow:
```

```
path(ss). length(2,ss).
:- not statpath(ss).
```

Applied to the example of Fig. 1, this gives the models corresponding to the transition graphs $\{G_1, G_2, G_4, G_5, G_6\}$ of Fig. 2.

This expressive power provides significant benefits over well-known temporal logics like *Computational Tree Logic* (CTL), which have been proposed to check instantiations of Thomas networks [22]. For example, a query asking for the existence of a model admitting three different steady states (without knowing them beforehand), is easy to formulate as an extension of the above rules, but cannot be expressed in CTL.

Nevertheless, CTL is useful to express biological observations, typically with $EF\varphi$ formulas, when applied to a state, meaning that there exists at least a path originating from this state leading to a state with property φ . In the declarative framework, one can easily enforce such formulas. For the example in Fig. 1, enforcing the existence of a path respecting the CTL formula $(a = 0 \wedge b = 0) \wedge EF(a = 0 \wedge b = 2)$ (there exists a path beginning with a state respecting $a = 0$ and $b = 0$ and reaching a state where $a = 0$ and $b = 2$) is achieved with the following rules⁴:

```
path(p). length(5,p).
:- not species(a,0,1,p).
:- not species(b,0,1,p).
exist_path :- species(a,0,I,p), species(b,2,I,p).
:- not exist_path.
```

The only models satisfying this formula are G_4 and G_6 (Fig. 2).

Enforcing universal CTL properties like $AF\varphi$, meaning that from the state to which it is applied all paths lead to a state with property φ , is not easily handled (see Section 7.2). However, such formula are not appropriate for transcribing a biological experiment in CTL because the experiments usually include only a few trials. An AF formula would be too strong: some valid models could be unduly eliminated. However, if we change the context, and search for constructing robust networks in a synthetic biological perspective, it becomes crucial, as it is in computer programming, to ensure universal properties.

Mutants In the study of genetic networks, biologists frequently suppress the expression of a gene or over-express it using genetic engineering. The resulting networks are labeled *mutant*, in contrast with the unaltered *wild* networks. In this section, the term “model” refers to either wild or mutant networks. It is important to define a mutant

⁴ The length of `p` is set to 5 because it is the maximal length of a non looping path for this example.

network from its comparable wild one and express different properties on each network while ensuring that they share the same kinetic parameters.

We proceed by introducing the predicates `model(M)` (M is a model), `mutant(M,N,V)` (in the model M , the gene N is mutated and its expression value is V) and `mutant(M,N)` (in the model M , the gene N is mutated). Moreover, we extend the predicate `path(P)` to `path(P,M)` (P is a path in the model M) in order to distinguish between paths in different models. Depending on the model, the value V of gene N in the first step of a path may be fixed or left ambiguous.

The non-monotonicity of ASP allows to use the same expression of rules defining the successors of a state (see Section 5.2) for either the wild model or mutant models. In order to obtain this expression, we add the literals `not mutant(N,M)`, `path(P,M)` to the body of the rule defining the value of N for a successor of state differing in N from this state. This prevents that this rule resets the focal value of a mutant gene. In case N is a mutant gene in the model M , the following rule makes constant the value of N for all state of any path P :

`species(N,V,I,P) :- mutant(M,N,V), path(P,M), step(I,P).`

where `step(I,P)` means that the path P has a I th step.

For the example in Figure 1, asking for the wild model with at least one steady state and for a mutant with $a = 0$ and a steady state in the mutant model such that $b = 0$, we get only the transition graphs G_1, G_2, G_4 and G_5 of Figure 2.

Inference of Properties From a coherent set of constraints, one is naturally lead to search for *predictions*. Within our approach, predictions are seen as properties that hold in all networks consistent with the biological constraints. To be worthwhile, this kind of automatic learning should be *non supervised*, i.e. not predicting a priori fixed properties but, instead, non fixed properties but nonetheless belonging to a language. The design of languages expressing properties is a problem per se which has to be discussed with biologists. Ideally, for predictions to be a guide for further experiments, biologically relevant properties should be experimentally verifiable. So we do not address this issue and will just give insight into this question.

Such a language could be composed with logical clauses [164], i.e. union of literals. They are particularly interesting because general logical formulas can be translated in this canonical form and because they are easily computationally represented. Inferring properties consequently consists to infer clauses. In this view, the only task to pay attention to is the choice of the set of atomic formulas. Corblin *et al.* [12] exhibit two languages where these atomic formulas are inequalities between kinetic parameters belonging to the same focal equation. For example (Fig. 2) the formula $(K_b < K_b^a) \vee \neg(K_b < K_b^{ab})$ belongs to such a language and is actually true in all models.

Automatic deduction of common characteristics across all models is simple with the `--cautious` option of the solver [162]. This option provides all the atomic formulas which are true across all models. By specifying the property languages using appropriate ASP predicates, one can easily obtain all the properties consistent across all models. Of course, practical limitations may be necessary, for example on the size of the clauses. Note that only three literals are needed to get properties which are implication

of an atom with the conjunction of two literals as premise. For example, $a_1 \Leftarrow a_2 \vee a_3$ may be deduced from $a_1 \wedge \neg a_2 \wedge \neg a_3$.

ASP is specifically interesting when used for the inference of properties: as models are minimal, the number of deduced properties is at least as high as the number of properties deduced using classical logic. For example, the program

```
a :- not b.  
b :- not a.
```

has only two minimal models: $\{a\}$ and $\{b\}$, while the model $\{a, b\}$ is not minimal. We can thus infer the exclusion of $\{a \wedge b\}$, which would not be possible using classical logic.

Minimization It can happen that the set of constraints accepts a very large number of consistent models, typically if the knowledge about the biological network is not very large. In these cases a frequent request concerns the existence of specific models. For defining such a specificity, optimization criteria are commonly proposed. For example, it may be interesting from a biological point of view to focus on models for which the total number of thresholds is minimal. It may even be possible that a Boolean model (with only one threshold for each gene) could be consistent. Of course, nothing prevent a priori that a gene has as many different thresholds as thoutput degree in the interaction graph.

A search for the minimum number of required thresholds in the network is possible using the `#minimize` operator, dual of `#maximize`. By defining the predicate `threshold_max(N, T)` (T is the number of thresholds of N), we ask for this minimum with the rule:

```
#minimize [threshold_max(N, T) : val(N, T) : node(N)=T].
```

To every atom `threshold_max(N, T)` that is true is associated a weight T. This rule automatically minimizes the sum of the weights.

4 Applications

The three applications that are presented below illustrate the following advantages of the declarative approach: inconsistency checking and repairing, minimization of interaction and threshold numbers, and temporal series modeling. We have recently applied this approach to three classical networks for comparison purposes (see Section 7.2).

Carbon Starvation Response in *Escherichia coli* The declarative approach has been applied to the re-examination of the regulation network of the carbon starvation in *E. coli* presented by Ropers *et al.* [18]. As long as environmental conditions are favorable, a population of *E. coli* bacteria grows quickly. The bacteria are in a state called exponential phase. Upon a nutritional stress due to carbon starvation, the bacteria are no longer able to maintain a fast growth rate. They enter in a stationary phase. Their response can be reversed as soon as the environmental conditions become favorable again. Modeling with the generate-and-test approach classically used for constructing GRN models led to a unique, instantiated and inconsistent model.

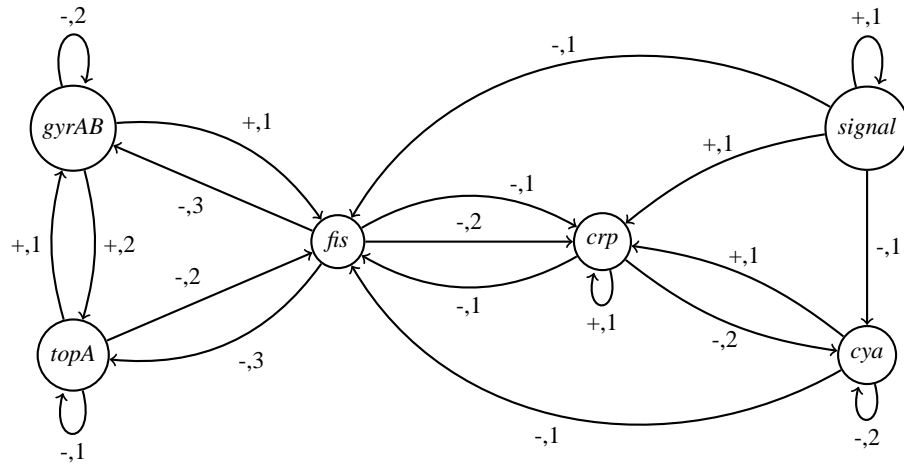


Fig. 3. Interaction graph of the regulation of the carbon starvation response in *Escherichia coli* [12].

A declarative analysis of this network, using CLP and SAT solvers, has been presented by Corblin *et al.* [12]. The network (Fig. 3) and biological observations on interactions, paths (stationary states and paths leading from the exponential phase to the stationary phase and vice-versa) and even characteristics of the shape of the DNA (*supercoiling*) were described using constraints. This analysis was resumed with an ASP implementation. We illustrate here the repairing of inconsistency.

Logical inconsistency was established. This showed rigorously the non-existence of alternative models, i.e. with a reasoning not based on the inconsistency of only one particular instantiated model. Repairing inconsistency was related to additivity constraints to the extent that they were not supported experimentally. Then repairing process proposed two solutions, that is to remove one constraint among $K_{gyrAB}^{fis} \leq K_{gyrAB}$ and $K_{topA} \leq K_{topA}^{fis}$. After biological investigations, it appeared that the first one should not be removed, but that the second could be, as it can be considered as not biologically plausible.

Computer performances stay very acceptable for solving such requests which require numerous recombination computations. For example, it is for determining the removable constraints that Corblin *et al.* [12] reports the highest computer time (around 25 minutes), with Prolog and a SAT solver loosely cooperating. This result was understandable because of the size of the solution space in this case. The same issue took four seconds when solved by our ASP implementation (with a Core 2 Duo 3GHz, 4GB RAM).

***Drosophila* Embryo Gap Genes Network** This ASP approach has been applied by Corblin *et al.* [165] to the regulatory network controlling the earliest steps of *Drosophila* embryo segmentation, i.e. the gap genes and their cross-regulations, under the additional control of maternal gene products [166,167,168]. Three kinds of data were con-

sidered: 1) published molecular genetic studies enabling the identification of the main actors (seven genes), as well as the establishment or the potentiality of cross-regulatory interactions, 2) qualitative information on the spatio-temporal expression profiles of the main genes involved in the process, giving seven regions with different stable states, 3) data available on the gap gene expression profiles for seven loss-of-function mutations, affecting maternal or gap genes. On the basis of this combination of interaction and gene expression constraints, the challenge was to identify the model(s) involving all established regulatory edges, along with a minimal set of potential ones, while minimizing the number of distinct thresholds.

In a first step, the consistency of the data was proven in 3338 seconds, using a Linux PC with an Intel Core2Duo processor at 2.4GHz and 2.9GB of memory. Then, a unique regulatory network structure was obtained in 1016 seconds which included only two potential interactions (on 11). Surprisingly, from this network, there was a unique instantiation of the thresholds minimizing the number of threshold values per component (obtained in 368 seconds). Finally, some properties concerning the kinetic parameters were deduced: 52 parameters fixed (over 72), 12 inequalities connecting a threshold and a parameter, and 36 inequalities connecting two parameters.

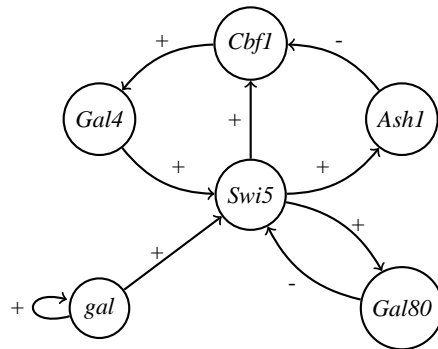


Fig. 4. Interaction graph of the IRMA network. The node *gal* is an input of the network. The values of the thresholds are not known.

IRMA Interaction Network The IRMA (standing for “In vivo benchmarking of Reverse-engineering and Modeling Approaches”) network [169] comprises five genes: *Swi5*, *Ash1*, *Cbf1*, *Gal4* and *Gal80*, as well as one input (*gal*) and eight interactions (Fig. 4). These genes were chosen in the synthesis of the network so that different types of interactions were considered, including transcription regulation and protein-protein interaction, thereby capturing the behavior of eukaryotic GRNs. Cantone *et al.* [169] explored the dynamics of the IRMA network by measuring each gene’s expression level in response to two different perturbations using qRT-PCR. In the first set of experiments, they shifted yeast cells from a glucose to galactose medium (“switch-on” experiments) and in the second set of experiments they shifted the cells from a galactose to

glucose medium (“switch-off” experiments). The presence of galactose allows for increased transcription of Swi5 and is thus “switch-on”, while the opposite is true for the “switch-off” experiments. From these data, two temporal series, composed of averaged gene expressions over five “switch-on” and four “switch-off” independent experiments, have been extracted.

Finding possible models of the IRMA network respecting these time series is a challenge proposed by Batt *et al.* [170]. The network is given in such terms that the order between the kinetic parameters is known. So the issue is to find a consistent order between thresholds and these parameters and between the thresholds themselves. Time series are formalized by CTL formulas of the form: $EF(prop_1 \wedge EF(\dots EF(prop_n)\dots))$ where $n = 12$ for the switch-off experiments and $n = 10$ for the switch-on experiments. A property $prop_i$ relates to the values of the components of a state and also to the derivative signs of these components. Batt *et al.* [170] propose a modeling taking into account *singular states* (states admitting for a component a threshold value) leading to more states, together with the use of the model checking tool NuSMV. They claim, reviewing their work, that they provide more precise results and efficient coding.

When applying an ASP declarative approach to this problem (not yet published), we designed the appropriate constraints for expressing that a path satisfies a time series, while remaining in the Thomas framework, i.e. without singular states. The same number of parametrizations (64) were exhibited as in the paper of Batt *et al.* [170]⁵, in 139 seconds (compared to 885 seconds, on a similar computer), showing that, at least for this analysis, taking account of singular states was not necessary.

5 Optimized Modeling of Thomas’ GRN and Additivity Constraints

Focal equations, successor relationship, observability and additivity constraints have been described in Section 3. In this section, we present and discuss their ASP modeling.

Two issues arise for modeling additivity constraints: 1) escaping from a possible inconsistency that would result if these constraints would be imposed, 2) getting only the “most general” networks that is, intuitively, those which accept as many as possible additivity constraints compatible with the biological data. Recall that such a minimization of the number of resulting models leads to increasing the number of properties which can be deduced.

We first present the main lines of a naive modeling which enumerates all models and uses the para-logical maximization operator of the gringo language for obtaining only answer sets with a maximal number of satisfied additivity constraints. However, both enumerating too many literals and using para-logic operators are costly.

The refined modeling which follows attempts to reduce as far as possible these costs by taking advantage of the non monotonicity of ASP. Default rules [163] will be applied for lowering the enumeration of the models. Also, an appropriate conjunction of defaults will be introduced for minimizing the number of resulting models. Note

⁵ Parameterizations were found identical, except for two of them that were erroneous in the paper of Batt *et al.* [170].

that this modeling allows to associate additivity constraints even to edges that would not be labeled by any sign in the interaction graph but that would support nonetheless observability constraints as a result of the given behaviors.

5.1 A Naive Modeling

In this modeling, we introduce the following predicates to specify the focal value (defined by the logical kinetic parameter) of a gene N in a state (defined as being the I th step in a path P , see Section 3.2):

- $\text{param}(Ik)$, where $Ik = k(N, CC)$ represents the logical kinetic parameter $K_N^{l(\sigma)}$ and $CC = cc(N1, cc(N2, \dots, cc(Np, nil)))$ represents $l(\sigma)$. CC characterizes the cellular context where the concentration of all N_i , $i \in [1, p]$, are respectively above the threshold of the edges $N_i \rightarrow N$ and where the concentrations of the source genes of all other edges for which N is a target are under their edge's threshold. The definition of this predicate, which we do not detail here, is directly built from the specification of the interaction graph.
- $\text{kparam}(K, Ik)$ expresses that K is the focal value of the parameter Ik . The naive following definition ensures that parameters values are unique and in the right range:

$$1\{\text{kparam}(K, k(N, CC)) : \text{val}(N, K)\}1 \text{ :- param}(k(N, CC)) .$$
Unfortunately, this definition leads to an exhaustive enumeration of all networks. This will be revised in the next section.
- $\text{cell_context}(N, CC, I, P)$ indicates that the state at step I of the path P belongs to the cellular context CC of N . We do not give here the recursive definition of this predicate.
- $\text{focal}(N, K, I, P)$ represents the focal equation of N , i.e. K is the focal value of N of the state at the step I of P , according the following rule:

$$\text{focal}(N, K, I, P) \text{ :- kparam}(K, k(N, CC)) , \\ \text{cell_context}(N, CC, I, P) .$$

With these predicates, transitions between states of a path can be easily specified [171].

To solve the above mentioned issues about additivity constraints, an immediate naive way is to define first the predicates⁶:

$$\begin{aligned} \text{addit}(S, N1, N) & \text{ :- obs}(S, N1, N) , \\ & \text{opposite_sign}(S, Sp) , \text{not obs}(Sp, N1, N) . \\ \text{obs}(S, N1, N) & \text{ :- obs_cond}(S, N1, N, K, K_r, Ik, Ik_r) , \\ & \text{kparam}(K, Ik) , \text{kparam}(K_r, Ik_r) . \end{aligned}$$

where $\text{obs_cond}(S, N1, N, K, K_r, Ik, Ik_r)$ means that, considering two neighbor cellular contexts identified by the parameters Ik and Ik_r , such that in the first (resp. second) cellular context, the value of N is lower (resp. higher) than the threshold associated to the edge $N1 \rightarrow N$, their possible focal values K and K_r stay in the right order according to the sign S . S is p (resp. m) if the edge is positive (resp. negative) (see Section 5.2 for a proper definition). The predicate $\text{opposite_sign}(S, Sp)$ means

⁶ Note that this modeling accepts edges with observability constraints not necessarily coming from the interaction graph.

that S and S_p are opposite signs. Then the integrity constraint:

```
:- addit(S,N1,N), not -obs(Sp,N1,N), opposite_sign(S,Sp).
```

where $-obs(S, N1, N)$ is the (usual) negation of $obs(S, N1, N)$, will allow retaining only answer sets with $addit(S, N1, N)$ implying $-obs(S_p, N1, N)$. Furthermore, one could apply a rather radical criterion (discussed in Section 5.2) for reducing the number of models by maximizing, with the para-logical operator `#maximize`, the number of `addit` atoms.

We should notice the default character of the rule defining `addit`. In the general case where $obs(S_p, N1, N)$ is not established, this rule is applied, with possible consequences. But if, due to the addition of new data, $obs(S_p, N1, N)$ is established, there will be no inconsistency because of these consequences.

5.2 A Refined Modeling

Lowering the Enumeration of `kparam` Atoms As said above, this enumeration may lead to enumerate too many answer sets. Consequently, the `kparam` atoms must be carefully produced. Their production could come from three origins: observability constraints due to the interaction graph, additivity constraints and biological behaviors. We propose the following specific rules related to each origin.

Observability constraints coming from the interaction graph are modeled by producing `kparam` atoms with the rules:

```
2{kparam(K, Ik), kparam(K_r, Ik_r)}2
:- couple_kpr(K, Ik, K_r, Ik_r).
```

and

```
couple_kpr(K, Ik, K_r, Ik_r) :
  obs_cond(S, N1, N, K, K_r, Ik, Ik_r)
  :- sign(S, N1, N).
```

where the predicate $sign(S, N1, N)$ means that S is the sign of the edge $N1 \rightarrow N$. This rule reflects the fact that only one atom $couple_kpr(K, Ik, K_r, Ik_r)$ is necessary for ensuring the observability constraints, through the atoms $kparam(K, Ik)$ and $kparam(K_r, Ik_r)$ that it implies⁷. The proper definition of obs_cond is:

```
obs_cond(S, N1, N, K, K_r, Ik, Ik_r)
:- neighboring_cell_cont(N1, N, Ik, Ik_r),
   param_obs(S, N1, N, K, K_r).
```

where $neighboring_cell_cont(N1, N, Ik, Ik_r)$ ensures the existence of the two cellular contexts separated by the threshold of the edge $N1 \rightarrow N$ and $param_obs(S, N1, N, K, K_r)$, the right order of the possible focal values K and K_r regarding the sign S of the edge $N1 \rightarrow N$.

⁷ Unfortunately the language `gringo` does not allow syntactically to replace $couple_kpr(K, Ik, K_r, Ik_r)$ by $2\{kparam(K, Ik), kparam(K_r, Ik_r)\}2$ in the above rule. So we have to introduce a rule ensuring that the first formula is implied by the second. This is necessary in order to prevent unduly answer sets with more than necessary inequations on parameters representing observability constraints, i.e. one inequation coming from $couple_kpr(K, Ik, K_r, Ik_r)$ and some others from another source like behaviors while the last ones only are sufficient.

The `couple_kpr` atoms due to additivity constraints are introduced by the default rule:

```
couple_kpr(K, Ik, K_r, Ik_r) : -param_obs(Sp, N1, N, K, K_r)
                             :- neighboring_cell_cont(N1, N, Ik, Ik_r),
                                addit(S, N1, N),
                                opposite_sign(S, Sp).
```

where `-param_obs(Sp, N1, N, K, K_r)` is the usual negation of `param_obs(Sp, N1, N, K, K_r)`. Note that this time, expressing the logical conjunction representing an additivity constraint requires a `couple_kpr` atom for every couple of cellular contexts separated by the threshold of the edge $N1 \rightarrow N$.

Finally, what remains is to model the rules defining the successor of a state. For the case where the species N changes its value at step $I+1$, we introduce the following rules to set the focal value and the species value:

```
2{species(N, V, I, P), kparam(K, k(N, CC))}
  :- couple_ks(N, V, I, P, K, CC).
```

and⁸:

```
couple_ks(N, V_s, I+1, P, K, CC) :
  ineq_K(S, N, K, V) : incl(S, N, V, V_s)
                       :- diff(N, I, P), cell_context(N, CC, I, P),
                          species(N, V, I, P), not mutant(N, P).
```

where `ineq_K(S, N, K, V)` ensures that K and V are values of N ordered according to S , `incl(S, N, V, V_s)` that V_s is $V+1$ (resp. $V-1$) if $S = p$ (resp. m) and `diff(N, I, P)` that N is the only species (if any) allowed to change at step I . The value K of the logical kinetic parameter of the cellular context CC of N is fixed according to the transition between the two states. For the case where the species N does not change its value at step $I+1$, only a `species` atom has to be produced⁹:

```
species(N, V, I+1, P) :- not diff(N, I, P),
                          species(N, V, I, P).
```

and for the case where no species at all change, only a `kparam` atom has to be produced:

```
kparam(V, k(N, CC)) :- not_any_diff(I, P), species(N, V, I, P),
                       cell_context(N, CC, I, P).
```

where `not_any_diff(I, P)` means that no species (among the non-mutant ones) change between the steps I and $I+1$.

Conjunction of Defaults and Appropriate Use of the Para-logical Maximization Operator

It is necessary to precisely explain what is intended by the “most general” networks regarding additivity constraints. For that, we may raise two different questions. The first is: “for a set of instantiated parameter sets satisfying observability constraints and behaviors (e.g. paths), what is the logical definition of the only answer sets

⁸ Due to suitable integrity constraints, the atoms `kparam(K, I_k)` are unique for a given I_k and the atoms `species(N, V, I, P)` for given N, I, P . Consequently, atoms `couple_ks(N, V, I, P, K, CC)` are unique for given N, I, P, CC .

⁹ Note the default character of this rule always applicable except when `not diff(N, I, P)` holds.

to be retained regarding additivity constraints?”, the second: “among the answer sets that result from the first question, are there still some to be eliminated?”. For minimizing the number of resulting models, there are two means: a logical one, coming from the minimality of the stable models and adapted to the first question, and a para-logical one via the maximization operator possibly useful for the second question.

For a set of instantiated parameter sets satisfying observability constraints and behaviors, it appears natural to ask for only answer sets having additivity constraints for all edges of all species, if such an answer set exists. If not, one would like to keep only the answer sets verifying additivity constraints for all edges of the species for which it is possible. For example, for the network (Fig. 1) with a behavior implying only $K_b^{ab} = 2$, there are eight answer sets (actually represented by the graphs G_1, \dots, G_6 of Fig. 2) with additivity constraints for all edges of all species. But there are also other possible networks, for example, with one edge of b with no additivity constraints. Unfortunately, the above refined modeling provides such undesirable networks, due to possible additivity constraints for one edge that imply the non additivity for some other edges. This is the case when having additivity constraints for the edge $a \rightarrow b$ is ensured by the additional focal values $K_b = 1$, $K_b^a = 1$ and $K_b^b = 0$ (thus $(K_b \leq K_b^a) \wedge (K_b^b \leq K_b^{ab})$ holds). These parameter values forbid additivity constraints for the edge $b \rightarrow b$ since $K_b \leq K_b^b$ does not hold.

The logical way to obtain only models with a maximal number of additive edges requires representing each additivity constraints by default rules and searching for parameter values such that the application of one default rule does not prevent the application of another default rule. As far as we know, there is not yet an established methodology for expressing in ASP such *conjunction* defaults, i.e. forming a set of defaults whose members cannot block the application of any other default of the set. So we develop in the next section such a methodology and apply it to define precisely the notion of most general models.

6 A Methodology for Constructing a Conjunction of Defaults with Application to Additivity Constraints

In this section, we describe the implementation of such conjunctions of defaults and then use it for our ASP modeling of Thomas’ networks.

6.1 Methodology

A simple program, described in a first part, with only two default rules will help to illustrates the issue and to exhibit a methodology to solve it, especially since only conjunctions implying two defaults are necessary for our purpose as it is shown in a second part.

Expressing the Conjunction of Two Defaults Let us consider the two following default rules with cross influences between them. The defaults are modeled by the rules¹⁰:

¹⁰ The predicates prefixed by *adi* represent parameter values verifying additivity constraints on edge i , with *ad1* and *ad1_2* representing two different sets of parameter values verifying

ad1_2 | ad1 :- not op_ad1.

ad2_1 | ad2 :- not op_ad2.

and the influences between defaults by the rules:

op_ad2 :- ad1_2.

op_ad1 :- ad2_1.

These four rules have three answer sets: $\{ad1, ad2\}$, $\{ad1_2, op_ad2\}$ and $\{ad2_1, op_ad1\}$. The last two answer sets are due to the blocking of one default rule by the other.

The first issue is to express the conjunction of these two defaults, i.e. to transform these rules so that we only get the answer set $\{ad1, ad2\}$. The methodology consists firstly in introducing the literal c so that $not\ c$ represents the case where both op_ad1 and op_ad2 do not hold:

c :- op_ad1 .

c :- op_ad2 .

Secondly, each default rule needs to be doubled, with the body of one of the new rules completed by c and the body of the other one by $not\ c$. For example, the two new rules derived from the first default are:

ad1_2 | ad1 :- not op_ad1, c .

ad1_2 | ad1 :- not op_ad1, not c .

In this way, when $not\ c$ is true then $ad1_2$ (and $ad2_1$) cannot be deduced and consequently $\{ad1, ad2\}$ is the only answer set obtained. If the fact op_ad1 is added then the two answer sets obtained are: $\{op_ad1, ad2_1\}$ and $\{op_ad1, ad2\}$.

A shorter expressing way consists to complete the body of each of the original rules with a kind of *guard*, namely the formula $1\{c, not\ c\}1$, which is a tautology but provided with a default impact power:

ad1_2 | ad1 :- not op_ad1, $1\{c, not\ c\}1$.

ad2_1 | ad2 :- not op_ad2, $1\{c, not\ c\}1$.

From a theoretical point of view, it should be noted that this kind of guard represents a logical curiosity due to the logic underlying ASP. Usually, tautologies could be suppressed inside rules. It is not the case here as shown above: guarded and non-guarded rules of the example do not provide the same set of answer sets.

Consistency Issue Extending the methodology to cope with constructing more than two defaults is straightforward. But it makes us to remind a second issue which is very present in default logic studies [163], namely how to face inconsistency when applying default. This issue concerns simple defaults but the consequences are amplified in the case where a conjunction of defaults is considered. If applying a conjunction results in inconsistency, answer sets which could be obtained by applying only part of the default rules are prejudicially discarded. In the above example, this can occur in case where the literals $ad1$ and $ad2$ cannot hold altogether. In such cases, applying the conjunction of defaults results in inconsistency, while there are answer sets obtained by applying only one of the default rules. Clearly, consistency proof have to be done for any default and a fortiori for a conjunction of defaults.

these additivity constraints. Those prefixed by op_adi represent the (rarely proven) conditions preventing additivity constraints on edge i .

Then, it can be mandatory to search for a supplementary property to get a necessary and sufficient condition for attaining consistency. In our case, we will search for a property `cond_inconst` which implies that `ad1` and `ad2` cannot be both true, but also such that its negation, together with those of `op_ad1` and `op_ad2`, allows `ad1` and `ad2` being both true. Practically, extending the rule defining `c` to:

```
c :- 1{op_ad1, op_ad2, cond_inconst}.
```

will ensure that no answer sets will be discarded by applying the conjunction of the defaults.

Discovering such a condition for a conjunction of several defaults could be a difficult task. Fortunately, limiting the requirement to the conjunction of two defaults is both possible and advantageous. Indeed, the guard of a conjunction of a family of defaults is trivially expressed as the logical conjunction of the guards of each conjunction of two defaults of this family. Extending the above example to three defaults and transforming it illustrates this point. The initial program:

```
c :- 1{op_ad1, op_ad2, op_ad3}.
```

```
ad1_2 | ad1_3 | ad1 :- not op_ad1, 1{c, not c}1.
```

```
ad2_1 | ad2_3 | ad2 :- not op_ad2, 1{c, not c}1.
```

```
ad3_1 | ad3_2 | ad3 :- not op_ad3, 1{c, not c}1.
```

can be transformed into:

```
c12 :- 1{op_ad1, op_ad2}.
```

```
c13 :- 1{op_ad1, op_ad3}.
```

```
c23 :- 1{op_ad2, op_ad3}.
```

```
guards2 :- 1{c12, not c12}1, 1{c13, not c13}1,
```

```
1{c23, not c23}1.
```

```
ad1_2 | ad1_3 | ad1 :- not op_ad1, guards2.
```

```
ad2_1 | ad2_3 | ad2 :- not op_ad2, guards2.
```

```
ad3_1 | ad3_2 | ad3 :- not op_ad3, guards2.
```

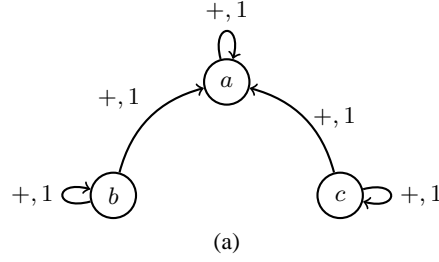
The advantages¹¹ are twofold. On one hand, the last formulation allows to tune for the best the resulting answer sets while the initial one does not provide any intermediate stage between applying the three defaults altogether or only one. On the other hand, if a supplementary condition for attaining inconsistency is needed, it is usually easier to discover it for only two defaults.

6.2 Application to Additivity Constraint Specification

Following the above methodology we have to check first that the default rule which produces additivity constraints (and which is conditioned by `addit(N1, N)`, see Section 5.2) does not result in inconsistency. This means that there should be, whatever are the observability constraints and the behaviors, an answer set respecting additivity constraints for an interaction $N1 \rightarrow N$ of sign S , provided that this interaction is additive, i.e. that `addit(S, N1, N)` holds (or, equivalently, that `obs(Sp, N1, N)` does not hold while `obs(S, N1, N)` does).

¹¹ As it could be seen, a drawback lies in the number of guards, quadratic in regards to the number of edges.

Proof. Let us consider the couples of parameters $K_N^{N1,\rho}$ and K_N^ρ for each cellular context ρ of N . If the parameters of such a couple are instantiated (i.e. known), the required additivity constraints are not violated since $\text{addit}(S, N1, N)$ holds. If not, there are appropriate values to give to these parameters: if only one parameter is instantiated, one can give for example an equal value to the other, if both parameters are unknown, the choice is left open for respecting the additivity constraints.



$$\begin{aligned}
 & (K_a < K_a^a \wedge K_a^a \geq 1) \vee (K_a^b < K_a^{ab} \wedge K_a^{ab} \geq 1) \\
 & \vee (K_a^c < K_a^{ac} \wedge K_a^{ac} \geq 1) \vee (K_a^{bc} < K_a^{abc} \wedge K_a^{abc} \geq 1) \\
 & (K_a < K_a^b) \vee (K_a^a < K_a^{ab}) \vee (K_a^c < K_a^{bc}) \vee (K_a^{ac} < K_a^{abc}) \\
 & (K_a < K_a^c) \vee (K_a^a < K_a^{ac}) \vee (K_a^b < K_a^{bc}) \vee (K_a^{ab} < K_a^{abc})
 \end{aligned}$$

(b)

Fig. 5. (a) Interaction graph. Thresholds have value 1 and the parameters take their value in $[0, 1]$. (b) Observability constraints for the edges $a \rightarrow a$, $b \rightarrow a$ and $c \rightarrow a$. The constraint related to the edge $a \rightarrow a$ is slightly more complicated because it is an auto-interaction. The observability constraints related to the edges $b \rightarrow b$ and $c \rightarrow c$ enforce that the parameters related to the nodes b and c are defined as $K_b = 0$, $K_b^b = 1$, $K_c = 0$ and $K_c^c = 1$.

For the example of Fig. 5 with the instantiations $K_a = 0$, $K_a^a = 1$, $K_a^b = 1$, $K_a^c = 1$ which ensure the observability constraints, there are some answer sets respecting the additivity constraints associated to the additive edge $b \rightarrow c$: among them, the one with $K_a^{ab} = 1$, $K_a^{bc} = 1$, $K_a^{ac} = 1$ and $K_a^{abc} = 1$.
□

According to the methodology, the conjunction of the two defaults related to two edges $N1 \rightarrow N$ and $N2 \rightarrow N$ is required next to be constructed. This is done by adding the following guard to the rules producing the additivity constraints related to these edges:

$$\text{guard}(N, N1, N2) \text{ :- } \neg \{ \text{one_no_pair_addit}(N, N1, N2), \\
 \text{not one_no_pair_addit}(N, N1, N2) \}.$$

with $\text{one_no_pair_addit}(N, N1, N2)$, meaning that either the edge $N1 \rightarrow N$ or the edge $N2 \rightarrow N$ is not additive, defined by the rule:

$$\text{one_no_pair_addit}(N, N1, N2) \text{ :- } \neg \{ \text{one_no_addit}(N1, N), \\
 \text{one_no_addit}(N2, N) \}.$$

where $\text{one_no_addit}(N1, N)$ holds if both $\text{obs}(S, N1, N)$ and $\text{obs}(Sp, N1, N)$ hold.

Again, it is necessary to check that, whatever are the observability constraints and the behaviors, the conjunction of defaults as defined above does not lead to inconsistency. Unfortunately, this is not the case.

Proof. By exhibiting a counter-example. Let us consider the example of Fig. 5, with the instantiations $K_a = 1$, $K_a^b = 0$, $K_a^{ab} = 0$, $K_a^{ac} = 0$, $K_a^{bc} = 0$ and $K_a^{abc} = 1$ which ensure the observability constraints¹². Clearly, there is no answer set respecting the conjunction of the two defaults related to the two edges $a \rightarrow a$ and $c \rightarrow a$ since the unknown parameters K_a^a and K_a^c should be higher (resp. lower) than or equal to $K_a = 1$ (resp. $K_a^{ac} = 0$) (see the lattice represented in Fig. 6)¹³.
□

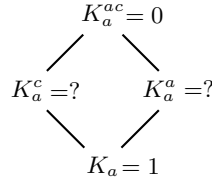


Fig. 6. K_a^a and K_a^c , not a priori known, cannot take values respecting both additivity constraints related to the edges $a \rightarrow a$ and $c \rightarrow a$.

Consequently, according to the methodology, the task to face is to search for a supplementary property to get a necessary and sufficient condition for attaining consistency and to take it into account to prevent a conjunction of two defaults related to two different edges $N1 \rightarrow N$ and $N2 \rightarrow N$ from leading to inconsistency. Let us call this property $\text{abnormal_lattice}(N, N1, N2)$ and take it into account by reformulating the definition of one_no_pair_addit :

$$\text{one_no_pair_addit}(N, N1, N2) \text{ :- } 1\{\text{one_no_addit}(N1, N), \\ \text{one_no_addit}(N2, N), \\ \text{abnormal_lattice}(N, N1, N2)\}.$$

Property. Let us suppose, in a first step, that the edges are both positive. Let us introduce lattices associated to these two edges, each made up with four parameters expressing their conjunct additivity constraints, namely $K_N^{N1.\rho} \leq K_N^{N1N2.\rho}$, $K_N^{N2.\rho} \leq K_N^{N1N2.\rho}$, $K_N^\rho \leq K_N^{N1.\rho}$ and $K_N^\rho \leq K_N^{N2.\rho}$, for each cellular context ρ containing neither $N1$, nor $N2$. Then $\text{abnormal_lattice}(N, N1, N2)$ is enunciated as: $\exists \rho | K_N^{N1N2.\rho} < K_N^\rho$. In the general case (whatever are the signs of the edges), the

¹² Note that $\text{one_no_addit}(b, a)$ holds, meaning that the edge $b \rightarrow a$ is not additive and then the same also for all conjunctions of defaults involving it.

¹³ Note that it can be proved that, in the absence of behaviors, the conjunction of defaults for all interactions does not lead to inconsistency.

lattices are constructed in the same way from the additivity constraints, with a maximum $K_N^{M,\rho}$ (like $K_N^{N1N2,\rho}$) and a minimum $K_N^{m,\rho}$ (like K_N^ρ), and the property¹⁴ is: $\exists \rho | K_N^{M,\rho} < K_N^{m,\rho}$.

Proof. Presented, without loss of generality, with both edges positive. In case where two adjacent nodes of a lattice are instantiated, the respect of the additivity constraints of the conjunction of defaults is detected via `one_no_addit(N1,N)` or `one_no_addit(N2,N)`. Consequently, `abnormal_lattice(N,N1,N2)` deals effectively only with instantiations of $K_N^{N1N2,\rho}$ and K_N^ρ . It is clearly sufficient as it infers that the possible values of $K_N^{N1,\rho}$ and $K_N^{N2,\rho}$ cannot respect the order induced by the considered lattice (see Fig. 6). It is necessary because if not present, it is always possible to give values to $K_N^{N1,\rho}$ and $K_N^{N2,\rho}$ which respect the order induced by the lattice, particularly in the case where $K_N^{N1N2,\rho}$ and K_N^ρ are instantiated to the same value.

□

Finally, the global guard, representing the composition of all double defaults related to a node N and which has to be added to the body part of the rule producing the additivity constraints is defined by the rule:

`guard(N) :- guard(N,N1,N2) :`
`N1!=N2: edge(N2,N) : edge(N1,N) .`

The following tests on the network presented in Fig. 5 illustrate the interest of the approach. Without any behaviors, we get 26 answer sets not completely instantiated: 14 without with all additivity constraints, six with additivity constraints on two edges and six with additivity constraints on one edge¹⁵. From these 26 answer sets, 101 completely instantiated are issued. After introducing the rule producing additivity, this number is reduced to 51.

To illustrate further the reduction of answer sets, let us consider a case where the observability constraints and the behaviors allow all additivity constraints, e.g. $K_a = 0$, $K_a^a = 0$, $K_a^b = 0$, $K_a^c = 1$, $K_a^{ab} = 1$. We get eight answer sets completely instantiated without considering additivity constraints, six answer sets by applying only unitary defaults, and one answer set with the global guard as defined above. Obtaining answer sets even in case of failure of a conjunction of defaults has also to be tested. For example, if we add to the preceding instantiations $K_a^{ac} = 0$, which forbids the conjunction of defaults on $a \rightarrow a$ and $c \rightarrow a$, we get anyway an answer set, with no additive constraints for $a \rightarrow a$. This illustrate well the non monotonic aspect of the programming technology (the addition of this fact does not lead to inconsistency).

Meanwhile, there remain cases that generate further questions. For example, for the network of Fig. 1 with at least two stationary states, this new modeling provides nonetheless three answer sets: one with the two edges targeting b being additive (graph G_6) and providing the stationary states $(0, 1)$ and $(0, 2)$, the two others respectively with one and not any of these edges being additive and providing the stationary states $(0, 1)$ and $(1, 0)$. The parameters values of these last answer sets come from the stationary

¹⁴ Defining `abnormal_lattice` requires literals whose number is exponential in regards to the number of edges minus 2. This is acceptable if the number of edges stay limited.

¹⁵ Unfortunately, we do not get only these fourteen answer sets. This is due to the production of observability constraints which do not respect some additivity constraints. Determining an acceptable guard for observability constraints stays an open problem.

states ($K_b = 1$ and $K_b^a = 0$) and the observability constraints ($K_b^{ab} = 2$, $K_b^b = 0$ or $K_b^b = 1$). They are acceptable models from the *logical* point of view developed above. Consequently, discriminating some answer sets among these three answer sets can require para-logic standards like the one used in Section 5.1, i.e. the winners are those having in the whole the greatest number of additive edges, which eliminates these two possibly undesirable models.

7 Conclusion

After a brief discussion, we give, as perspectives, a short introduction to two ongoing works.

7.1 Discussion

We gave the main lines of an ASP modeling of Thomas' GRNs and illustrated the declarative approach interest with biological applications which make use of inconsistency repairing, minimization of interactions and temporal series representation. To take into account properties only generally true, we presented adequate default rules and an optimized modeling both reducing the number of used atoms (an efficient way for improving performances in computational logics), and keeping only most generally accepted models in a logical way (as opposed to a para-logical one). For this purpose, we were led to express conjunction of defaults with surprising and powerful ASP logical expressions. A methodology is needed for constructing safe conjunctions of defaults since they could lead to undesirable inconsistency. So we developed such a methodology demanding possibly a supplementary condition but to be applied only on pairs of defaults. In the case of additivity constraints production, we exhibit such a condition acting on the lattices derived from the inequations due to these constraints on pairs of interactions.

Few other teams use a declarative approach for analyzing Thomas' networks. For this purpose, they use model checking tools and formalize paths with the temporal logic CTL, as in a seminal paper [22] on the subject. We showed advantages of our approach in Sections 3.2 and 4. It can be added that a constraint programming approach is well suited to avoid external processes to extract properties common to the set of consistent models, because such properties are themselves expressed as constraints. However, as mentioned in Section 3.2, imposing (opposed to checking), in a convenient way, CTL formulas of the form $AF\varphi$ with logic programming remains an issue. We discuss this point in Section 7.2.

Finally, it should be noted that ASP has also been applied successfully by other teams to the modeling of biological networks (see for example the work of Videla *et al.* [172]), but not specifically, at our knowledge, to the modeling of Thomas' GRNs. The apparently closest ASP based work is reported by Fayruzov *et al.* [173], but it deals only with simplistic instantiated deterministic Boolean networks, thus excluding issues coming from the multi-valued non deterministic Thomas' GRNs modeling in the declarative approach perspective, yet offered by the programming logic technology. Furthermore, it does not emphasize the non monotonicity offered by ASP.

7.2 Perspectives

We present two ongoing works initiated from two issues: constructing robust networks in a synthetic biology perspective (see Section 3.2) and expressing and learning general compositions of interactions (see Section 3.2).

Expressing general CTL formulas with ASP We just sketch here the main lines of an ASP implementation of general CTL formulas, focusing on EF et AF formulas. The following rules, very close to the logical definitions of these formulas, represent a good starting point:

```
eF(Prop,S) :- ap(Prop,S).
eF(Prop,S) :- not ap(Prop,S),
               transition(S,Sp), eF(Prop,Sp).
aF(Prop,S) :- ap(Prop,S).
aF(Prop,S) :- not ap(Prop,S),
               aF(Prop,Sp): transition(S,Sp).
```

where $ef(Prop,S)$ (resp. $af(Prop,S)$) stands for “when applied to the state S , the formula $EF(Prop)$ (resp. $AF(Prop)$) is true”, the predicate $ap(Prop,S)$ stands for “the state S verifies the property $Prop$ ” and $transition(S,Sp)$ for “the state Sp is a successor of the state S ”. The representation of properties $Prop$ is not detailed here.

Three difficulties appear:

- As seen previously in Sections 3.2 and 5.2, EF formulas are implemented in a simple way by defining constrained paths. These paths are limited in length and so the limited amount of constraints, linear according to the maximum length of a path, are tractable. Unfortunately, the number of states needed for an AF formula is exponential regarding this length. So the implementation previously presented is not appropriate. Another ASP modeling with an overall limitation on the number of states is necessary. It is currently undertaken.
- The second issue concerns how loops are tackled by the above rules which, remarkably, do not provide any test for checking circular paths. For example, one could find such a path generated by the following rules:

```
eF(Prop,s1) :- eF(Prop,s2), transition(s1,s2).
eF(Prop,s2) :- eF(Prop,s1), transition(s2,s1).
```

in case where $ap(Prop,s1)$ and $ap(Prop,s2)$ do not hold. Thus, it is of crucial importance to recall that the stable models, those only accepted by ASP, are minimal [174]. This means that if $transition(s1,s2)$ and $transition(s2,s1)$ hold, the only accepted answer sets are the one where $eF(Prop,s1)$ and $eF(Prop,s2)$ are false, whereas classical logic would consider also the model where they are true. This minimality is obviously very appropriate for our purpose.
- The third issue is related to the last rule defining the predicate aF . Its grounding supposes that all atoms with predicate $transition$ are known. That is not generally the case because transitions may depend on initially unknown parameters. A way to overcome this difficulty is to introduce in the definition of the predicate

```

aF, instead of transition(S, Sp), the literal hyp_AFtrans(Prop, S, Sp),
whose grounding stays finite. We get:
aF(Prop, S) :- ap(Prop, S).
aF(Prop, S) :- not ap(Prop, S),
                hyp_AFtrans(Prop, S, Sp): state(Sp).
hyp_AFtrans(Prop, S, Sp) :- transition(S, Sp), aF(Prop, Sp).
hyp_AFtrans(Prop, S, Sp) :- not transition(S, Sp).

```

Multiplexes For comparing her approach with ours, a Phd student from Ircyn lab. (Nantes) asked us to study paths in several networks. So we were brought to implement the ERBB receptor-regulated G1/S transition network presented by Sahin *et al.* [175] (18 species), the network describing the biological system triggering the tail resorption during the metamorphosis of tadpole presented by Khalis *et al.* [176] (8 species) and the T-cell Signaling network presented by Klamt *et al.* [177] (40 species)¹⁶. The last two ones and specially the one presented by Khalis *et al.* [176] considers sets of interactions which act conjointly, namely *multiplexes*. An immediate advantage of this notion is the reduction of the number of interactions. In our declarative view, another benefit is that it provides the definition of interesting properties to be learned from all consistent answer sets. Then an ongoing work aims at implementing properly in a declarative framework these multiplexes.

This implementation is based on a precise semantics of multiplexes and also general composition of interactions. We adopt for the composition of interactions a representation similar to the disjunctive normal form for logical formulas (already proposed par R. Thomas as SOP, Sum Of Products)¹⁷. This choice is particularly justified for an easy computational learning of multiplexes. The semantics of multiplexes is described in terms of observability constraints on the parameters. This can be sketched in the following way, without considering auto-interactions:

- Syntactically, the language of the interactions targeting a given species, say x , is defined by the following grammar of axiom I (I stands for interactions, Mul for multiplex and Iu for unitary interaction)¹⁸:

$$\begin{aligned}
I &::= Mul \mid Mul \textbf{ or } I \\
Mul &::= Iu \mid Iu \textbf{ and } Mul \\
Iu &::= Gene_id \mid Sig \ Gene_id \\
Sig &::= + \mid -
\end{aligned}$$

¹⁶ Performances of our implementation compare very favorably with other approaches mentioned by E. Ben Abdallah. She thinks that the limitation of the length of a path is a drawback of our proposition. Of course it is, if one just wants to exhibit a consistent answer set. But that is not our primary goal. An incremental version of our implementation would be more appropriate for this purpose.

¹⁷ Our definition differs from the definition given by Khalis *et al.* [176], but it has the same expression power.

¹⁸ This grammar ensures the priority of **and** on **or**. As we will see, formulas are associative and commutative regarding the associated observability constraints.

- The semantics of I is defined as the value of a predicate $oc(I)$ which is a logical function of the parameters.

Examples:

$$oc(+a) = K_x^a > K_x.$$

$$oc(+a \textbf{ and } +b) = (K_x^{ab} > K_x) \wedge (K_x^a = K_x^b = K_x)$$

$$oc(+a \textbf{ or } +b) = (K_x^{ab} > K_x \vee K_x^a > K_x) \wedge (K_x^{ab} > K_x^a \vee K_x^b > K_x)$$

This predicate is constructed following the well-known *composition principle*¹⁹. Its definition is presented in the Annex 8, together with the extended one of additivity constraints.

Acknowledgments This work was supported by Microsoft Research through its PhD Scholarship Programme. We acknowledge funding by the Agence Nationale de la Recherche through the CADMIDIA project.

8 Annex. Semantics of the composition of interactions

We give first the definition of $oc(I)$, where I conforms to the grammar presented in Section 7.2 (interactions targeting the same species x) and then some examples of additivity constraints.

To each formula Mul and Iu is associated a couple of cellular contexts $[M, m]$ of x , where M (resp. m) is the one of the formula indexing the maximal (resp. minimal) kinetic parameter, e.g. to the formula $+a$, indicating a positive interaction $a \rightarrow x$, is associated the couple $[a, \phi]$ and to $-b$ is associated $[\phi, b]$. To each formula I is associated a list of such couples.

- The value of $oc(Iu)$ is the observability constraint associated to Iu , which is the inequality derived from its associated couple $[M, m]$, that is $K_x^M > K_x^m$. For example:

$$oc(+a) = K_x^a > K_x.$$

- The value of $oc(Iu \textbf{ and } Mul)$, determined from the associated couples $[M1, m1]$ (resp. $[M2, m2]$) of Iu (resp. Mul), is the conjunction of 1) the inequality $K_x^{M1.M2} > K_x^{m1.m2}$ and 2) the equalities of $K_x^{m1.m2}$ with all other parameters involved by Iu and Mul ²⁰. For example:

$$oc(+a \textbf{ and } +b) = (K_x^{ab} > K_x) \wedge (K_x^a = K_x^b = K_x)$$

This expresses that both species a and b must be above their threshold to trigger an activation of x .

The couple associated to the formula $Iu \textbf{ and } Mu$ is $[M1.M2, m1.m2]$, e.g. $[ab, \phi]$ is associated to $+a \textbf{ and } +b$. For example:

$$oc(-c \textbf{ and } +a \textbf{ and } +b) = (K_x^{ab} > K_x^c)$$

$$\wedge (K_x^{abc} = K_x^{ac} = K_x^{bc} = K_x^a = K_x^b = K_x = K_x^c)$$

since the couple associated to $-c$ is $[\phi, c]$ and the one associated to $+a \textbf{ and } +b$ is $[ab, \phi]$.

¹⁹ The value of a formula is a composition of the values of its components.

²⁰ The operator “.” denotes concatenation.

- The value of $oc(Mul \text{ or } I)$, determined from the associated couple $[M, m]$ associated to Mul and the list of couples $L_CC_I = [[M_1, m_1], \dots, [M_n, m_n]]$ associated to I , is the conjunction of the union of formulas deduced from $oc(Mul)$ where all cellular contexts are “augmented” with certain cellular contexts coming from L_CC_I and the union of two formulas deduced from $oc(Mul)$ by replacing each context cellular ρ it contains by $M.\rho$ or $m.\rho$. More precisely, let be S_CC_I the set of cellular contexts $p_1 \dots p_i \dots p_n$ where $p_i = M_i$ or $p_i = m_i$ and $F_{[\rho \rightarrow cc.\rho]}$ representing the formula F where all cellular contexts ρ are substituted by $cc.\rho$. Then²¹:
 $oc(Mul \text{ or } I) = \bigvee_{\omega \in S_CC_I} oc(Mul)_{[\rho \rightarrow \omega.\rho]} \wedge (oc(I)_{[\rho \rightarrow M.\rho]} \vee oc(I)_{[\rho \rightarrow m.\rho]})$
The list of couples associated to $Mul \text{ or } I$ is $[M, m].L_CC_I$.

Examples:

$$oc(+a \text{ or } +b) = (K_x^{ab} > K_x^b \vee K_x^a > K_x) \wedge (K_x^{ab} > K_x^a \vee K_x^b > K_x)$$

since $[a, \phi]$ is associated to $+a$ and $[[b, \phi]]$ to $+b$.

$$oc(+a \text{ or } -b) = (K_x^{ab} > K_x^b \vee K_x^a > K_x) \wedge (K_x^a > K_x^{ab} \vee K_x > K_x^b)$$

since $[a, \phi]$ is associated to $+a$ and $[[\phi, b]]$ to $-b$.

$$oc(+a \text{ and } +b \text{ or } -c) = ((K_x^{ab} > K_x) \wedge (K_x^a = K_x^b = K_x))$$

$$\vee ((K_x^{abc} > K_x^c) \wedge (K_x^{ac} = K_x^{bc} = K_x^c))$$

$$\wedge ((K_x^{abc} < K_x^{ab}) \vee (K_x^c < K_x))$$

since $[ab, \phi]$ is associated to $+a \text{ and } +b$ and $[[\phi, c]]$ to $-c$.

The definition of additivity is an extension of the one given in Section 3.2 for a simple interaction. This time, it is related to a multiplex: it is additive if the observability constraints of this multiplex holds with a sign but does not with the opposite sign. Then, two notions must be clarified: observability constraints associated to a multiplex and sign of a multiplex.

Recall the observability constraints of the edge $b \rightarrow a$ in the example of Fig. 5:

$$(K_a < K_a^b) \vee (K_a^a < K_a^{ab}) \vee (K_a^c < K_a^{bc}) \vee (K_a^{ac} < K_a^{abc})$$

They express that there exists at least a case, among those provided by the cellular contexts arising from all edges targeting a , where an activation of a due to this edge has been observed. Extending the notion of observability constraints to multiplexes needs to define the predicate $oc_cont(Mult, I)$ giving the observability constraints related to $Mult$ if all interactions targeting the same node x are represented by I . The following definition is easily deduced from the one of $oc(Mult \text{ and } I)$:

$$oc_cont(Mul, Mul \text{ or } I) = \bigvee_{\omega \in S_CC_I} oc(Mul)_{[\rho \rightarrow \omega.\rho]}$$

Examples:

$$oc_cont(+a, +a \text{ or } +b) = K_x^{ab} > K_x^b \vee K_x^a > K_x$$

NB. Note that, due to commutativity of formulas I regarding observability constraints, we have $oc_cont(+a, +a \text{ or } +b) = oc_cont(+a, +b \text{ or } +a)$.

$$oc_cont(+a \text{ and } +b, +a \text{ and } +b \text{ or } -c)$$

$$= ((K_x^{ab} > K_x) \wedge (K_x^a = K_x^b = K_x))$$

$$\vee ((K_x^{abc} > K_x^c) \wedge (K_x^{ac} = K_x^{bc} = K_x^c)).$$

In the presented framework, a multiplex is considered to be an activation. Consequently, to express an inhibition it is necessary to deal with the sign of the unitary interactions belonging to a multiplex. For example, $+a$ is a multiplex representing an

²¹ The value of $oc(Mul)$ is deduced from $oc(Mul \text{ or } I)$ by taking $L_CC_I = []$.

interaction and $-a$ an inhibition. The same for $-a$ **and** $-b$ which represents an inhibition as compared with $+a$ **and** $+b$, i.e. an activation when both a and ab are under their thresholds. Then, the multiplex expressing the inhibition corresponding to a multiplex Mul is obtained by inverting the sign of every unitary interactions it contains.

Now, we can easily define the additivity constraints of a multiplex according to Section 3.2. Let us consider the predicate $add_cont(Mul, I)$ giving the additivity constraints of Mul , with I representing all interactions targeting the same node x . We have: $add_cont(Mul, Mul \text{ or } I) = inv_mul(Mul, Mulp) \wedge \neg oc_cont(Mulp, Mulp \text{ or } I)$

Examples:

$$\begin{aligned}
add_cont(+a, +a \text{ or } +b) &= \neg oc_cont(-a, -a \text{ or } +b) \\
&= \neg(K_x^{ab} < K_x^b \vee K_x^a < K_x) = K_x^{ab} \geq K_x^b \wedge K_x^a \geq K_x \\
add_cont(+a \text{ and } +b, +a \text{ and } +b \text{ or } -c) \\
&= \neg oc_cont(-a \text{ and } -b, -a \text{ and } -b \text{ or } -c) \\
&= \neg [((K_x^{ab} < K_x) \wedge (K_x^{ab} = K_x^b = K_x^a)) \vee ((K_x^{abc} < K_x^c) \wedge (K_x^{abc} = K_x^{bc} = K_x^{ac}))] \\
&= ((K_x^{ab} \geq K_x) \vee atm1D(K_x^{ab}, K_x^b, K_x^a)) \wedge ((K_x^{abc} \geq K_x^c) \vee atm1D(K_x^{abc}, K_x^{bc}, K_x^{ac}))
\end{aligned}$$

where $atm1D(v_1, \dots, v_i, \dots)$ means that at least for one couple (i, j) , v_i differs from v_j .

Bibliographie

- [1] T. W. Fawcett and A. D. Higginson. Heavy use of equations impedes communication among biologists. *Proc. Natl. Acad. Sci. U.S.A.*, 109(29) :11735–11739, July 2012.
- [2] A. R. Carvunis, E. Gomez, N. Thierry-Mieg, L. Trilling, and M. Vidal. Systems biology : from yesterday’s concepts to tomorrow’s discoveries. *Med Sci (Paris)*, 25(6-7) :578–584, 2009.
- [3] A. D. Sheftel, D. R. Richardson, J. Prchal, and P. Ponka. Mitochondrial iron metabolism and sideroblastic anemia. *Acta Haematol.*, 122(2-3) :120–133, 2009.
- [4] M. W. Hentze, M. U. Muckenthaler, and N. C. Andrews. Balancing acts : molecular control of mammalian iron metabolism. *Cell*, 117(3) :285–297, April 2004.
- [5] M. A. Schaub, T. A. Henzinger, and J. Fisher. Qualitative networks : a symbolic approach to analyze biological signaling networks. *BMC Syst Biol*, 1 :4, 2007.
- [6] Z. P. Gerdtzen. Modeling metabolic networks for mammalian cell systems : General considerations, modeling strategies, and available tools. In *Genomics and Systems Biology of Mammalian Cell Culture*, volume 127 of *Advances in Biochemical Engineering Biotechnology*, pages 71–108. Springer Berlin Heidelberg, 2012.
- [7] R. Thomas. Logical description, analysis, and synthesis of biological and other networks comprising feedback loops. *Adv. Chem. Phys.*, 55 :247–282, 1983.
- [8] R. Thomas and R. d’Ari. *Biological Feedback*. CRC Press, 1990.
- [9] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. II. logical analysis of regulatory networks in terms of feedback circuits. *CHAOS*, 11(1) :180–195, 2001.
- [10] J. Demongeot, R. Thomas, and M. Thellier. A mathematical model for storage and recall functions in plants. *C. R. Acad. Sci. III, Sci. Vie*, 323(1) :93–97, January 2000.
- [11] M. Kaufman, F. Andris, and O. Leo. A logical analysis of T cell activation and anergy. *Proc. Natl. Acad. Sci. U.S.A.*, 96(7) :3894–3899, March 1999.
- [12] F. Corblin, S. Tripodi, É. Fanchon, D. Ropers, and L. Trilling. A declarative constraint-based method for analyzing discrete genetic regulatory networks. *Biosystems*, 98 :91–104, 2009.
- [13] F. Corblin, É. Fanchon, and L. Trilling. Applications of a formal approach to decipher discrete genetic networks. *BMC Bioinformatics*, 11(1), 2010.
- [14] F. Achcar, J.-M. Camadro, and D. Mestivier. A Boolean probabilistic model of metabolic adaptation to oxygen in relation to iron homeostasis and oxidative stress. *BMC Systems Biology*, 5(1), 2011.

- [15] A. Sackmann, D. Formanowicz, P. Formanowicz, and J. Blazewicz. New insights into the human body iron metabolism analyzed by a Petri net based approach. *Biosystems*, 96(1) :104–113, 2009.
- [16] E. H. Snoussi. Qualitative dynamics of piecewise-linear differential equations : a discrete mapping approach. *Dynamics and Stability of Systems*, 4(3) :565–583, 1989.
- [17] H. De Jong, J. L. Gouze, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull. Math. Biol.*, 66(2) :301–340, March 2004.
- [18] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems*, 84(2) :124–152, 2006.
- [19] D. Ropers, V. Baldazzi, and H. de Jong. Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli*. *IEEE/ACM Trans Comput Biol Bioinform*, 8(1) :166–181, 2011.
- [20] J. Ahmad, G. Bernot, J.-P. Comet, D. Lime, and O. Roux. Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs*, 3 :231–251, 2006.
- [21] J.-P. Comet, J. Fromentin, G. Bernot, and O. Roux. A formal model for gene regulatory networks with time delays. In J.H. Chan, Ong Y.-S., and Cho S.-B., editors, *1st International Conference on Computational Systems-Biology and Bioinformatics (CSBio'2010)*, volume 115 of *Communications in Computer and Information Science (CCIS)*, pages 1–13, Bangkok (Thailand), November 2010. Springer.
- [22] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks : extending Thomas' asynchronous logical with temporal logic. *Journal of Theoretical Biology*, 229(3) :339–347, 2004.
- [23] M. Davidich and S. Bornholdt. The transition from differential equations to Boolean networks : a case study in simplifying a regulatory network model. *J. Theor. Biol.*, 255(3) :269–277, December 2008.
- [24] G. Kervizic and L. Corcos. Dynamical modeling of the cholesterol regulatory pathway with Boolean networks. *BMC Syst Biol*, 2 :99, 2008.
- [25] J. Aracena, É. Fanchon, M. Montalva, and M. Noual. Combinatorics on update digraphs in Boolean networks. *Discrete Applied Mathematics*, 159(6) :401–409, 2011.
- [26] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, 79(8) :2554–2558, April 1982.
- [27] T. Hervé, J. M. Dolmazon, and J. Demongeot. Random field and neural information. *Proceedings of the National Academy of Sciences*, 87(2) :806–810, 1990.
- [28] J. Aracena, S.B. Lamine, M.A. Mermet, O. Cohen, and J. Demongeot. Mathematical modeling in genetic networks : relationships between the genetic expression and both chromosomal breakage and positive circuits. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 33(5) :825–834, October 2003.
- [29] J. Aracena, J. Demongeot, and E. Goles. Positive and negative circuits in discrete neural networks. *Neural Networks, IEEE Transactions on*, 15(1) :77–83, January 2004.
- [30] H. Ben Amor, F. Corblin, É. Fanchon, A. Elena, L. Trilling, J. Demongeot, and N. Glade. Formal methods for Hopfield-like networks. *Acta Biotheor.*, 61(1) :21–39, March 2013.
- [31] J. Demongeot. Random automata networks. In F. Fogelman-Soulie, Y. Robert, and M. Tchuente, editors, *Automata networks in computer science : theory and applications*, pages 47–57. 1987.

- [32] R. David and H. Alla. *Discrete, Continuous, and Hybrid Petri Nets*. Springer, 2005.
- [33] P. J. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. U.S.A.*, 95(12) :6750–6755, June 1998.
- [34] Á. M. Halász, H. J. Lai, M. McCabe Pryor, K. Radhakrishnan, and J. S. Edwards. Analytical solution of steady-state equations for chemical reaction networks with bilinear rate laws. *IEEE/ACM Trans Comput Biol Bioinform*, 10(4) :957–969, 2013.
- [35] E. Asarin, T. Dang, and A. Girard. Hybridization methods for the analysis of nonlinear systems. *Acta Informatica*, 43(7) :451–476, 2007.
- [36] M. B. Sass, A. N. Lorenz, R. L. Green, and R. A. Coleman. A pragmatic approach to biochemical systems theory applied to an alpha-synuclein-based model of Parkinson’s disease. *J. Neurosci. Methods*, 178(2) :366–377, April 2009.
- [37] M. W. Covert, C. H. Schilling, and B. Palsson. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 213(1) :73–88, November 2001.
- [38] O. Maler and D. Nickovic. Monitoring temporal properties of continuous signals. In Y. Lakhnech and S. Yovine, editors, *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*, volume 3253 of *Lecture Notes in Computer Science*, pages 152–166. Springer Berlin Heidelberg, 2004.
- [39] A. Donzé, É. Fanchon, L. M. Gattepaille, O. Maler, and P. Tracqui. Robustness analysis and behavior discrimination in enzymatic reaction networks. *PLoS ONE*, 6(9) :e24246, September 2011.
- [40] S. Stoma, A. Donzé, F. Bertaux, O. Maler, and G. Batt. STL-based analysis of TRAIL-induced apoptosis challenges the notion of type I/type II cell line classification. *PLoS Comput. Biol.*, 9(5) :e1003056, 2013.
- [41] J. J. Tyson, K. C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers : dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.*, 15(2) :221–231, April 2003.
- [42] H. Kitano. Biological robustness. *Nat. Rev. Genet.*, 5(11) :826–837, November 2004.
- [43] K. Csillery, M. G. Blum, O. E. Gaggiotti, and O. Francois. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol. (Amst.)*, 25(7) :410–418, July 2010.
- [44] A. Georgoulas, A. Clark, A. Ocone, S. Gilmore, and G. Sanguinetti. A subsystems approach for parameter estimation of ODE models of hybrid systems. *Electronic Proceedings in Theoretical Computer Science*, 92 :30–41, 2012.
- [45] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31) :187–202, 2009.
- [46] T. J. Hickey and D. K. Wittenberg. Rigorous modeling of hybrid systems using interval arithmetic constraints. In R. Alur and G. J. Pappas, editors, *Hybrid Systems : Computation and Control*, volume 2993 of *Lecture Notes in Computer Science*, pages 402–416. Springer Berlin Heidelberg, 2004.
- [47] A. Donzé. Breach, a toolbox for verification and parameter synthesis of hybrid systems. In *CAV*, pages 167–170, 2010.
- [48] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. SUNDIALS : Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.*, 31(3) :363–396, September 2005.

- [49] J. Wang and K. Pantopoulos. Regulation of cellular iron metabolism. *Biochem. J.*, 434(3) :365–381, March 2011.
- [50] Y. Yu, Z. Kovacevic, and D. R. Richardson. Tuning cell cycle regulation with an iron key. *Cell Cycle*, 6(16) :1982–1994, August 2007.
- [51] G. Cairo and S. Recalcati. Iron-regulatory proteins : molecular biology and pathophysiological implications. *Expert reviews in molecular medicine*, 9(33) :1–13, 2007.
- [52] M. P. Yeager and R. A. Coleman. In silico evidence for glutathione- and iron-related pathogenesis in Parkinson’s disease. *J. Neurosci. Methods*, 188(1) :151–164, April 2010.
- [53] V. Hower, P. Mendes, F. M. Torti, R. Laubenbacher, S. Akman, V. Shulaev, and S. V. Torti. A general map of iron metabolism and tissue-specific subnetworks. *Molecular bioSystems*, 5 :422–443, May 2009.
- [54] K. Pantopoulos, S. K. Porwal, A. Tartakoff, and L. Devireddy. Mechanisms of mammalian iron homeostasis. *Biochemistry*, 51(29) :5705–5724, July 2012.
- [55] A. Funahashi, M. Morohashi, H. Kitano, and N. Tanimura. Celldesigner : a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*, 1(5) :159–162, November 2003.
- [56] F. Samaniego, J. Chin, K. Iwai, T. A. Rouault, and R. D. Klausner. Molecular characterization of a second iron-responsive element binding protein, iron regulatory protein 2. Structure, function, and post-translational regulation. *J. Biol. Chem.*, 269(49) :30904–30910, December 1994.
- [57] B. R. Henderson and L. C. Kuhn. Differential modulation of the RNA-binding proteins IRP-1 and IRP-2 in response to iron. *J. Biol. Chem.*, 270(35) :20509–20515, September 1995.
- [58] N. K. Gray and M. W. Hentze. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J.*, 13(16) :3882–3891, August 1994.
- [59] A. A. Vashisht, K. B. Zumbrennen, X. Huang, D. N. Powers, A. Durazo, D. Sun, N. Bhaskaran, A. Persson, M. Uhlen, O. Sangfelt, C. Spruck, E. A. Leibold, and J. A. Wohlschlegel. Control of iron homeostasis by an iron-regulated ubiquitin ligase. *Science*, 326(5953) :718–721, October 2009.
- [60] B. J. Iacopetta and E. H. Morgan. The kinetics of transferrin endocytosis and iron uptake from transferrin in rabbit reticulocytes. *J. Biol. Chem.*, 258(15) :9108–9115, August 1983.
- [61] A. D. Sheftel, A. S. Zhang, C. Brown, O. S. Shirihai, and P. Ponka. Direct interorganellar transfer of iron from endosome to mitochondrion. *Blood*, 110(1) :125–132, July 2007.
- [62] C. Seiser, M. Posch, N. Thompson, and L. C. Kuhn. Effect of transcription inhibitors on the iron-dependent degradation of transferrin receptor mRNA. *J. Biol. Chem.*, 270(49) :29400–29406, December 1995.
- [63] R. Erlitzki, J. C. Long, and E. C. Theil. Multiple, conserved iron-responsive elements in the 3’-untranslated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2. *J. Biol. Chem.*, 277(45) :42579–42587, November 2002.
- [64] N. Hubert and M. W. Hentze. Previously uncharacterized isoforms of divalent metal transporter (DMT)-1 : implications for regulation and cellular function. *Proc. Natl. Acad. Sci. U.S.A.*, 99 :12345–12350, September 2002.

- [65] A. P. West, M. J. Bennett, V. M. Sellers, N. C. Andrews, C. A. Enns, and P. J. Bjorkman. Comparison of the interactions of transferrin receptor and transferrin receptor 2 with transferrin and the hereditary hemochromatosis protein HFE. *J. Biol. Chem.*, 275(49) :38135–38138, December 2000.
- [66] B. J. Iacopetta, E. H. Morgan, and G. C. Yeoh. Transferrin receptors and iron uptake during erythroid cell development. *Biochim. Biophys. Acta*, 687(2) :204–210, May 1982.
- [67] I. De Domenico, D. M. Ward, G. Musci, and J. Kaplan. Evidence for the multimeric structure of ferroportin. *Blood*, 109(5) :2205–2209, March 2007.
- [68] D. L. Zhang, R. M. Hughes, H. Ollivierre-Wilson, M. C. Ghosh, and T. A. Rouault. A ferroportin transcript that lacks an iron-responsive element enables duodenal and erythroid precursor cells to evade translational repression. *Cell Metab.*, 9(5) :461–473, May 2009.
- [69] D. M. Ward and J. Kaplan. Ferroportin-mediated iron transport : expression and regulation. *Biochim. Biophys. Acta*, 1823(9) :1426–1433, September 2012.
- [70] L. Cianetti, P. Segnalini, A. Calzolari, O. Morsilli, F. Felicetti, C. Ramoni, M. Gabbianelli, U. Testa, and N. M. Sposi. Expression of alternative transcripts of ferroportin-1 during human erythroid differentiation. *Haematologica*, 90(12) :1595–1606, December 2005.
- [71] L. Cianetti, M. Gabbianelli, and N. M. Sposi. Ferroportin and erythroid cells : an update. *Adv. Hematol.*, 2010, 2010.
- [72] P. M. Harrison and P. Arosio. The ferritins : molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta*, 1275(3) :161–203, July 1996.
- [73] N. D. Chasteen and P. M. Harrison. Mineralization in ferritin : an efficient means of iron storage. *J. Struct. Biol.*, 126(3) :182–194, June 1999.
- [74] D. M. Lawson, A. Treffry, P. J. Artymiuk, P. M. Harrison, S. J. Yewdall, A. Luzzago, G. Cesareni, S. Levi, and P. Arosio. Identification of the ferroxidase centre in ferritin. *FEBS Lett.*, 254(1-2) :207–210, August 1989.
- [75] H. Shi, K. Z. Bencze, T. L. Stemmler, and C. C. Philpott. A cytosolic iron chaperone that delivers iron to ferritin. *Science*, 320(5880) :1207–1210, May 2008.
- [76] I. De Domenico, M. B. Vaughn, L. Li, D. Bagley, G. Musci, D. M. Ward, and J. Kaplan. Ferroportin-mediated mobilization of ferritin iron precedes ferritin degradation by the proteasome. *EMBO J.*, 25(22) :5396–5404, November 2006.
- [77] M. W. Hentze, T. A. Rouault, S. W. Caughman, A. Dancis, J. B. Harford, and R. D. Klausner. 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(19) :6730–6734, October 1987.
- [78] N. Aziz and H. N. Munro. Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region. *Proc. Natl. Acad. Sci. U.S.A.*, 84(23) :8478–8482, December 1987.
- [79] E. Pourcelot, M. Lénon, N. Mobilia, J. Y. Cahn, J. Arnaud, É. Fanchon, J. M. Moulis, and P. Mossuz. Iron for proliferation of cell lines and hematopoietic progenitors : Nailing down the intracellular functional iron concentration. *Biochim. Biophys. Acta*, 1853(7) :1596–1605, July 2015.
- [80] M. Shvartsman, E. Fibach, and Z. I. Cabantchik. Transferrin-iron routing to the cytosol and mitochondria as studied by live and real-time fluorescence. *Biochem. J.*, 429(1) :185–193, July 2010.

- [81] D. R. Richardson, D. J. Lane, E. M. Becker, M. L. Huang, M. Whitnall, Y. Suryo Rahmanto, A. D. Sheftel, and P. Ponka. Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. *Proc. Natl. Acad. Sci. U.S.A.*, 107(24) :10775–10782, June 2010.
- [82] G. C. Ferreira and J. Gong. 5-Aminolevulinate synthase and the first step of heme biosynthesis. *J. Bioenerg. Biomembr.*, 27(2) :151–159, April 1995.
- [83] D. B. Mirel, K. Marder, J. Graziano, G. Freyer, Q. Zhao, R. Mayeux, and K. C. Wilhelmsen. Characterization of the human mitochondrial aconitase gene (ACO2). *Gene*, 213(1-2) :205–218, June 1998.
- [84] T. C. Cox, M. J. Bawden, A. Martin, and B. K. May. Human erythroid 5-aminolevulinate synthase : promoter analysis and identification of an iron-responsive element in the mRNA. *EMBO J.*, 10(7) :1891–1902, July 1991.
- [85] C. K. Wu, H. A. Dailey, J. P. Rose, A. Burden, V. M. Sellers, and B. C. Wang. The 2.0 Å structure of human ferrochelatase, the terminal enzyme of heme biosynthesis. *Nat. Struct. Biol.*, 8(2) :156–160, February 2001.
- [86] D. R. Crooks, M. C. Ghosh, R. G. Haller, W. H. Tong, and T. A. Rouault. Posttranslational stability of the heme biosynthetic enzyme ferrochelatase is dependent on iron availability and intact iron-sulfur cluster assembly machinery. *Blood*, 115 :860–869, January 2010.
- [87] D. Krell, M. Assoku, M. Galloway, P. Mulholland, I. Tomlinson, and C. Bardella. Screen for IDH1, IDH2, IDH3, D2HGDH and L2HGDH mutations in glioblastoma. *PLoS ONE*, 6 :e19868, May 2011.
- [88] L. Oliveira and J. C. Drapier. Down-regulation of iron regulatory protein 1 gene expression by nitric oxide. *Proc. Natl. Acad. Sci. U.S.A.*, 97(12) :6550–6555, June 2000.
- [89] F. Zhang, W. Wang, Y. Tsuji, S. V. Torti, and F. M. Torti. Post-transcriptional modulation of iron homeostasis during p53-dependent growth arrest. *J. Biol. Chem.*, 283(49) :33911–33918, December 2008.
- [90] M. W. Hentze and L. C. Kuhn. Molecular control of vertebrate iron metabolism : mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl. Acad. Sci. U.S.A.*, 93(16) :8175–8182, August 1996.
- [91] M. Castoldi and M. U. Muckenthaler. Regulation of iron homeostasis by microRNAs. *Cell. Mol. Life Sci.*, 69(23) :3945–3952, June 2012.
- [92] C. Sangokoya, J. F. Doss, and J. T. Chi. Iron-responsive miR-485-3p regulates cellular iron homeostasis by targeting ferroportin. *PLoS Genet.*, 9(4) :e1003408, April 2013.
- [93] E. Nemeth and T. Ganz. Regulation of iron metabolism by hepcidin. *Annu. Rev. Nutr.*, 26 :323–342, 2006.
- [94] E. Nemeth, M. S. Tuttle, J. Powelson, M. B. Vaughn, A. Donovan, D. M. Ward, T. Ganz, and J. Kaplan. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science*, 306(5704) :2090–2093, December 2004.
- [95] J. A. M. Borghans, R. J. de Boer, and L. A. Segel. Extending the quasi-steady state approximation by changing variables. *Bulletin of Mathematical Biology*, 58(1) :43–63, 1996.
- [96] S. W. Omholt, X. Kefang, Ø. Andersen, and E. Plahte. Description and analysis of switchlike regulatory networks exemplified by a model of cellular iron homeostasis. *Journal of Theoretical Biology*, 195(3) :339–350, 1998.

- [97] E. H. Flach and S. Schnell. Use and abuse of the quasi-steady-state approximation. *Syst Biol (Stevenage)*, 153(4) :187–191, July 2006.
- [98] C. J. Doherty and S. A. Kay. Circadian control of global gene expression patterns. *Annu. Rev. Genet.*, 44 :419–444, 2010.
- [99] E. Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev.*, 4 :345–376, October 1914.
- [100] W. G. Xu and Q. S. Li. The dimensionlessness of chemical kinetic equation. *Journal of Mathematical Chemistry*, 31(3) :237–250, April 2002.
- [101] S. H. Lam and D. A. Goussis. The CSP method for simplifying kinetics. *International Journal of Chemical Kinetics*, 26(4) :461–486, 1994.
- [102] W. A. Stein et al. *Sage Mathematics Software (Version 6.7)*. The Sage Development Team, 2014. <http://www.sagemath.org>.
- [103] J. Chifman, A. Kniss, P. Neupane, I. Williams, B. Leung, Z. Deng, P. Mendes, V. Hower, F. M. Torti, S. A. Akman, S. V. Torti, and R. Laubenbacher. The core control system of intracellular iron homeostasis : a mathematical model. *J. Theor. Biol.*, 300 :91–99, May 2012.
- [104] S. Mitchell and P. Mendes. A computational model of liver iron metabolism. *PLoS Comput. Biol.*, 9(11) :e1003299, November 2013.
- [105] T. Dang and T. Dreossi. Falsifying oscillation properties of parametric biological models. In *Proceedings Second International Workshop on Hybrid Systems and Biology, HSB 2013*, volume 125 of *EPTCS*, pages 53–67, September 2013.
- [106] L. Granvilliers and F. Benhamou. Algorithm 852 : RealPaver : an interval solver using constraint satisfaction techniques. *ACM Transactions on Mathematical Software*, 32(1) :138–156, March 2006.
- [107] N. C. Andrews. Forging a field : the golden age of iron biology. *Blood*, 112(2) :219–230, July 2008.
- [108] D. Galaris and K. Pantopoulos. Oxidative stress and iron homeostasis : mechanistic and health aspects. *Crit Rev Clin Lab Sci*, 45(1) :1–23, 2008.
- [109] I. De Domenico, D. McVey Ward, and J. Kaplan. Regulation of iron acquisition and storage : consequences for iron-linked disorders. *Nat. Rev. Mol. Cell Biol.*, 9(1) :72–81, January 2008.
- [110] R. C. Hider and X. Kong. Iron speciation in the cytosol : an overview. *Dalton Trans*, 42(9) :3220–3229, February 2013.
- [111] P. Arosio and S. Levi. Cytosolic and mitochondrial ferritins in the regulation of cellular iron homeostasis and oxidative damage. *Biochim. Biophys. Acta*, 1800(8) :783–792, August 2010.
- [112] J. C. Salgado, A. Olivera-Nappa, Z. P. Gerdtzen, V. Tapia, E. C. Theil, C. Conca, and M. T. Nunez. Mathematical modeling of the dynamic storage of iron in ferritin. *BMC Syst. Biol.*, 4 :147, 2010.
- [113] P. Aisen. Transferrin receptor 1. *The International Journal of Biochemistry & Cell Biology*, 36(11) :2137–2143, 2004.
- [114] R. Mayr, A. R. Janecke, M. Schranz, W. J.H. Griffiths, W. Vogel, A. Pietrangelo, and H. Zoller. Ferroportin disease : A systematic meta-analysis of clinical and molecular findings. *Journal of Hepatology*, 53(5) :941–949, November 2010.

- [115] J.-L. Gouzé. Positive and negative circuits in dynamical systems. *J. Biol. Syst.*, 6(1) :11–15, March 1998.
- [116] A. Donzé and O. Maler. Robust satisfaction of temporal logic over real-valued signals. In *FORMATS*, pages 92–106, 2010.
- [117] W. A. Stein et al. *Sage Mathematics Software (Version 6.2)*. The Sage Development Team, 2014. <http://www.sagemath.org>.
- [118] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4) :308–313, 1965.
- [119] E. Zamora-Sillero, M. Hafner, A. Ibig, J. Stelling, and A. Wagner. Efficient characterization of high-dimensional parameter spaces for systems biology. *BMC Syst. Biol.*, 5 :142, 2011.
- [120] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2) :81–93, June 1938.
- [121] The MathWorks, Inc. MATLAB and Statistics Toolbox Release 2012a.
- [122] R.R. Sokal and F.J. Rohlf. *Biometry*. W. H. Freeman, 1995.
- [123] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer Publishing Company, Incorporated, 4 edition, 2002.
- [124] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2011.
- [125] A. C. Cameron and Trivedi P. K. *Microeconometrics : methods and applications*. Cambridge University Press, May 2005.
- [126] H. Bengtsson. *R.matlab : Read and write of MAT files together with R-to-MATLAB connectivity*, 2014. R package version 3.0.0.
- [127] E. Pourcelot, N. Mobilia, A. Donzé, F. Louis, O. Maler, P. Mossuz, É. Fanchon, and J. M. Moulis. Cellular iron regulation in animals : need and use of suitable models. In *Nutzen-Risiko-Bewertung von Mineralstoffen und Spurenelemente : Biochemische, physiologische und toxikologische Aspekte*, pages 73–89. KIT Verlag, 2013.
- [128] M. D. Knutson, M. Oukka, L. M. Koss, F. Aydemir, and M. Wessling-Resnick. Iron release from macrophages after erythrophagocytosis is up-regulated by ferroportin 1 overexpression and down-regulated by hepcidin. *Proc. Natl. Acad. Sci. U.S.A.*, 102(5) :1324–1328, February 2005.
- [129] C. Dycke, P. Charbonnier, K. Pantopoulos, and J. M. Moulis. A role for lysosomes in the turnover of human iron regulatory protein 2. *Int. J. Biochem. Cell Biol.*, 40(12) :2826–2832, 2008.
- [130] A. M. Weissman, R. D. Klausner, K. Rao, and J. B. Harford. Exposure of K562 cells to anti-receptor monoclonal antibody OKT9 results in rapid redistribution and enhanced degradation of the transferrin receptor. *J. Cell Biol.*, 102(3) :951–958, March 1986.
- [131] R. Moirand, A. M. Mortaji, O. Loréal, F. Paillard, P. Brissot, and Y. Deugnier. A new syndrome of liver iron overload with normal transferrin saturation. *The Lancet*, 349(9045) :95–97, January 1997.
- [132] S. Epsztejn, O. Kakhlon, H. Glickstein, W. Breuer, and I. Cabantchik. Fluorescence analysis of the labile iron pool of mammalian cells. *Anal. Biochem.*, 248(1) :31–40, May 1997.

- [133] O. S. Chen, K. P. Blemings, K. L. Schalinske, and R. S. Eisenstein. Dietary iron intake rapidly influences iron regulatory proteins, ferritin subunits and mitochondrial aconitase in rat liver. *J. Nutr.*, 128(3) :525–535, March 1998.
- [134] R. C. Hunt and L. Marshall-Carlson. Internalization and recycling of transferrin and its receptor. Effect of trifluoperazine on recycling in human erythroleukemic cells. *J. Biol. Chem.*, 261(8) :3681–3686, March 1986.
- [135] A. A. Kurzhanskiy P. Gagarinov. 2014.
- [136] J. Chifman, R. Laubenbacher, and S. V. Torti. A systems biology approach to iron metabolism. *Adv. Exp. Med. Biol.*, 844 :201–225, 2014.
- [137] T. J. S. Lopes, T. Luganskaja, M. Vujić Spasić, M. W. Hentze, M. U. Muckenthaler, K. Schümann, and J. G. Reich. Systems analysis of iron metabolism : the network of iron pools and fluxes. *BMC Systems Biology*, 4(1), 2010.
- [138] F. Achcar. *Modélisations Booléennes probabilistes de l’homéostasie du fer pour l’exploration des liens entre stress oxydant et perturbations métaboliques : de la cellule à l’organisme*. PhD thesis, Université Pierre et Marie Curie, 2010.
- [139] D. Formanowicz, A. Sackmann, P. Formanowicz, and J. Błażewicz. Petri net based model of the body iron homeostasis. *J Biomed Inform*, 40(5) :476–485, October 2007.
- [140] J. Błażewicz, D. Formanowicz, P. Formanowicz, A. Sackmann, and M. Sajkowski. Modeling the process of human body iron homeostasis using a variant of timed Petri nets. *Discrete Applied Mathematics*, 157(10) :2221–2231, 2009.
- [141] V. Becker, M. Schilling, J. Bachmann, U. Baumann, A. Raue, T. Maiwald, J. Timmer, and U. Klingmüller. Covering a broad dynamic range : information processing at the erythropoietin receptor. *Science*, 328(5984) :1404–1408, June 2010.
- [142] W. Wang, X. Di, R. B. D’Agostino, S. V. Torti, and F. M. Torti. Excess capacity of the iron regulatory protein system. *J. Biol. Chem.*, 282(34) :24650–24659, August 2007.
- [143] J. Ma, S. Haldar, M. A. Khan, S. D. Sharma, W. C. Merrick, E. C. Theil, and D. J. Goss. Fe²⁺ binds iron responsive element-RNA, selectively changing protein-binding affinities and regulating mRNA repression and activation. *Proc. Natl. Acad. Sci. U.S.A.*, 109(22) :8417–8422, May 2012.
- [144] J. B. Goforth, S. A. Anderson, C. P. Nizzi, and R. S. Eisenstein. Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. *RNA*, 16(1) :154–169, January 2010.
- [145] R. M. Coulson and D. W. Cleveland. Ferritin synthesis is controlled by iron-dependent translational derepression and by changes in synthesis/transport of nuclear ferritin RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, 90(16) :7613–7617, August 1993.
- [146] L. V. Sharova, A. A. Sharov, T. Nedorezov, Y. Piao, N. Shaik, and M. S. Ko. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.*, 16(1) :45–58, February 2009.
- [147] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppim, and T. Tuller. Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput. Biol.*, 7(9) :e1002127, September 2011.
- [148] R. Y. Y. Chan, C. Seiser, H. M. Schulman, L. C. Kühn, and P. Ponka. Regulation of transferrin receptor mRNA expression. *Eur. J. Biochem.*, 220(3) :683–692, March 1994.

- [149] E. Mattia, K. Rao, D. S. Shapiro, H. H. Sussman, and R. D. Klausner. Biosynthetic regulation of the human transferrin receptor by desferrioxamine in K562 cells. *J. Biol. Chem.*, 259(5) :2689–2692, March 1984.
- [150] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape : a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11) :2498–2504, November 2003.
- [151] 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 M. A. Aon, V. Saks, and U. Schlattner, editors, *Systems Biology of Metabolic and Signaling Networks*, volume 16 of *Springer Series in Biophysics*, pages 63–114. Springer Berlin Heidelberg, 2014.
- [152] T. Z. Kidane, E. Sauble, and M. C. Linder. Release of iron from ferritin requires lysosomal activity. *Am. J. Physiol., Cell Physiol.*, 291(3) :C445–455, September 2006.
- [153] K. Pantopoulos, N. K. Gray, and M. W. Hentze. Differential regulation of two related RNA-binding proteins, iron regulatory protein (IRP) and IRPB. *RNA*, 1(2) :155–163, April 1995.
- [154] E. Pourcelot. *Homéostasie cellulaire du fer dans les cellules leucémiques myéloïdes*. PhD thesis, Université Grenoble Alpes, 2015.
- [155] T. Sun, W. Yang, J. Liu, and P. Shen. Modeling the basal dynamics of p53 system. *PLoS ONE*, 6(11) :e27882, November 2011.
- [156] X. Brazzolotto, M. Andriollo, P. Guiraud, A. Favier, and J. M. Moulis. Interactions between doxorubicin and the human iron regulatory system. *Biochim. Biophys. Acta*, 1593(2–3) :209–218, February 2003.
- [157] E. P. Testa, U. Testa, P. Samoggia, G. Salvo, A. Camagna, and C. Peschle. Expression of transferrin receptors in human erythroleukemic lines : regulation in the plateau and exponential phase of growth. *Cancer Res.*, 46(10) :5330–5334, October 1986.
- [158] L. L. Chu and R. A. Fineberg. On the mechanism of iron-induced synthesis of apoferritin in HeLa cells. *J. Biol. Chem.*, 244(14) :3847–3854, July 1969.
- [159] Y. Deville, M. Janssen, and P. Van Hentenryck. Consistency techniques in ordinary differential equations. In M. Maher and J.-F. Puget, editors, *Principles and Practice of Constraint Programming — CP98*, volume 1520 of *Lecture Notes in Computer Science*, pages 162–176. Springer Berlin Heidelberg, 1998.
- [160] N. Mobilia, A. Rocca, S. Chorlton, É Fanchon, and L. Trilling. Logical modeling and analysis of regulatory genetic networks in a non monotonic framework. In *Bioinformatics and Biomedical Engineering*, volume 9043, pages 599–612, 2015.
- [161] C. Baral. *Knowledge Representation, Reasoning, and Declarative Problem Solving*. Cambridge University Press, 2003.
- [162] M. Gebser, R. Kaminski, B. Kaufmann, M. Ostrowski, T. Schaub, and S. Thiele. *A user's guide to gringo, clasp, clingo, and iclingo (version 3.x)*, October 2010.
- [163] P. Besnard. *An Introduction to Default Logic*. Springer, 1989.
- [164] G. Bossu and P. Siegel. Saturation, nonmonotonic reasoning and the closed-world assumption. *Artif. Intell.*, 25 :13–63, January 1985.
- [165] F. Corblin, É Fanchon, L. Trilling, C. Chaouiya, and D. Thieffry. Automatic inference of regulatory and dynamical properties from incomplete gene interaction and expression data. In *Information Processing in Cells and Tissues*, pages 25–30, 2012.

- [166] L. Sánchez and D. Thieffry. A logical analysis of the *Drosophila* Gap-gene system. *J. Theor. Biol.*, 211(2) :115–141, July 2001.
- [167] J. Jaeger, M. Blagov, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, S. Surkova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, 167 :1721–1737, 2004.
- [168] F. Alves and R. Dilão. Modeling segmental patterning in *Drosophila* : Maternal and gap genes. *J. Theor. Biol.*, 241(2) :342–359, July 2006.
- [169] I. Cantone, L. Marucci, F. Iorio, M. A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, and M. P. Cosma. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, 137(1) :172–181, April 2009.
- [170] G. Batt, M. Page, I. Cantone, G. Goessler, P. Monteiro, and H. de Jong. Efficient parameter search for qualitative models of regulatory networks using symbolic model checking. *Bioinformatics*, 26(18) :i603–610, September 2010.
- [171] A. Rocca, N. Mobilia, É. Fanchon, T. Ribeiro, L. Trilling, and K. Inoue. *ASP for Construction and Validation of Regulatory Biological Networks*, pages 167–206. John Wiley & Sons, Inc., August 2014.
- [172] S. Videla, C. Guziolowski, F. Eduati, S. Thiele, M. Gebser, J. Nicolas, J. Saez-Rodriguez, T. Schaub, and A. Siegel. Learning boolean logic models of signaling networks with ASP. *Theoretical Computer Science*, 2014.
- [173] T. Fayruzov, J. Janssen, D. Vermeir, C. Cornelis, and M. De Cock. Modelling gene and protein regulatory networks with Answer Set Programming. *Int J Data Min Bioinform*, 5(2) :209–229, March 2011.
- [174] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. *Logic Programming*, 2 :1070–1080, 1988.
- [175] O. Sahin, H. Fröhlich, C. Löbke, U. Korf, S. Burmester, M. Majety, J. Mattern, I. Schupp, C. Chaouiya, D. Thieffry, A. Poustka, S. Wiemann, T. Beissbarth, and D. Arlt. Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst Biol*, 3(1), 2009.
- [176] Z. Khalis, J.-P. Comet, A. Richard, and G. Bernot. The SMBioNet method for discovering models of gene regulatory networks. *Genes, Genomes and Genomics*, 3(1) :15–22, 2009.
- [177] S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, and E. D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7(1) :56, 2006.