



HAL
open science

Modélisation en langage Kappa de la dynamique des cellules étoilées hépatiques durant le développement et la réversion de la fibrose

Matthieu Bouguéon

► **To cite this version:**

Matthieu Bouguéon. Modélisation en langage Kappa de la dynamique des cellules étoilées hépatiques durant le développement et la réversion de la fibrose. Médecine humaine et pathologie. Université de Rennes, 2023. Français. NNT : 2023URENB073 . tel-04388949v2

HAL Id: tel-04388949

<https://inria.hal.science/tel-04388949v2>

Submitted on 26 Apr 2024

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.

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 637
Sciences de la Vie et de la Santé
Spécialité : *Génétique, Génomique, Bioinformatique*

Par

Matthieu BOUGUÉON

Modélisation en langage Kappa de la dynamique des cellules étoilées hépatiques durant le développement et la réversion de la fibrose

Thèse présentée et soutenue à Rennes, le 21 Decembre 2023
Unité de recherche : IRSET - Équipe DyMec, Irista - Équipe Dyliss

Rapporteurs avant soutenance :

Anna NIARAKIS Professeur des universités - Université de Toulouse
Cédric LHOUSSAINE Professeur des universités - Université de Lille
Fabien CRAUSTE Directeur de recherches CNRS - UMR 8541 CNRS, Univ Paris Cité

Composition du Jury :

Président :	Cédric LHOUSSAINE	Professeur des universités - Université de Lille
Rapporteurs :	Anna NIARAKIS	Professeur des universités - Université de Toulouse
	Cédric LHOUSSAINE	Professeur des universités - Université de Lille
	Fabien CRAUSTE	Directeur de recherches CNRS - UMR 8541 CNRS, Univ Paris Cité
Examineurs :	Sophie LOTERSZTAJN	Directrice de recherches Inserm - UMR 1149 Inserm, Univ Paris Cité
Dir. de thèse :	Nathalie THÉRET	Directrice de recherches Inserm - UMR1085 Inserm, EHESP, Univ Rennes
Co-dir. de thèse :	Anne SIEGEL	Directrice de recherches CNRS - UMR 6074 CNRS, Inria, Univ Rennes
Co-encadrant :	Jérôme FERET	Chargé de recherches Inria - ENS, Paris

REMERCIEMENTS

Tout d'abord, je tiens à remercier Anna Niarakis, Cédric Lhoussaine et Fabien Crauste pour leur temps passé à relire et commenter mes travaux de thèse. Je souhaite aussi remercier Sophie Lotersztajn d'avoir acceptée d'être membre de mon jury. Merci à tous pour vos remarques et l'intérêt porté à mes travaux.

Il est maintenant temps de remercier mes encadrants (officiels et officieux). Tout d'abord Anne Siegel, je te remercie pour ton écoute et ton empathie qui m'auront permis de venir à bout de cette thèse dans de meilleures conditions. Merci aussi pour ta capacité à me faire comprendre la qualité et la quantité de travail que j'ai pu réaliser durant cette thèse, même si mon syndrome de l'imposteur n'est toujours pas loin. Je tiens aussi à remercier Jérôme Féret pour les nombreuses opportunités que tu m'as apportées. Merci aussi pour toutes ces bonnes adresses sur Paris, j'espère avoir l'occasion d'en tester de nouvelles prochainement. Merci aussi de m'avoir réconcilié avec les mathématiques ainsi que pour le soutien et les conseils que tu m'as apportés durant ces trois années. Merci à Vincent Legagneux, tu as tellement apporté durant cette thèse que tu fais pour moi partie de mes encadrants. Merci pour toutes ces réunions (qui avaient toujours tendances à durer trop longtemps), merci pour tous tes conseils et le temps passé à relire ce manuscrit dans lequel beaucoup trop de phrases ne voulaient rien dire en première écriture. Enfin, merci à Nathalie Théret pour m'avoir donnée l'opportunité de travailler sur ce projet, merci pour ton temps, ta patience et ta capacité à me pousser à donner le meilleur de moi-même. J'espère que les prochains mois à travailler ensemble permettront d'améliorer ce modèle qui nous aura tant résisté. Finalement, je vous remercie tous pour votre humanité qui m'aura été grandement nécessaire durant ces trois années, merci encore.

Je remercie tous les collègues de Symbiose, (même si mon cœur est à Dyliss) merci de m'avoir montré qu'il était encore possible de travailler en recherche dans de bonnes conditions. Merci pour les soirées Harry Potter, Seigneur des Anneaux et toutes les bonnes bouffes (même si c'est plutôt à vous de me remercier d'avoir cuisiné;)) Merci à Victor E et Garance pour leur folie (surtout toi Victor) et leur écoute. Merci à Nicolas pour m'avoir permis d'améliorer mon chrono à la course même si un peu intense comme approche :) Merci à Kerian pour les parties de JDR, les parties de Magic et surtout pour Thanatos!

Merci à Cham et Florian qui, malgré la distance, avez su rester toujours présents. Je vous remercie mon Pas Capitaine, on en aura bavé en Master et surtout en stage mais c'est fini maintenant on est libre, il suffit de faire le grand saut ! Merci à JCVD sans qui mes présentations auraient été moins fun. Merci à Thomas et Morgane pour toutes ces années, je sais ce n'est pas les potagers connectés que j'avais promis. Merci pour les hébergements... promis cette fois c'était la dernière... Merci à l'équipe de Koh Lanta, j'espère survivre à l'épreuve de poteaux mais bon après le ski c'est bien parti.

Je tiens aussi à remercier ma famille, pour leur soutien et leur patience durant toutes ces années. Bien que cela n'a pas toujours été simple, c'est en aussi grâce à vous si j'en suis la aujourd'hui.

Enfin, je souhaite te remercier mon amour, merci d'avoir été là pour me soutenir et m'encourager. Je t'aime.

TABLE DES MATIÈRES

I	Introduction	13
1	La fibrose hépatique	14
1.1	Les fonctions du foie	14
1.2	Les cellules du foie	16
1.3	Les maladies chroniques du foie	18
1.4	Les cellules étoilées du foie : moteur de la fibrose	22
1.5	Les modèles d'études expérimentales de la fibrose	26
2	Modéliser les systèmes biologiques	28
2.1	Catégories de modèles	28
2.1.1	Les modèles statiques : identifier les relations	30
2.1.2	Les modèles dynamiques : identifier les comportements	31
2.1.2.1	Les modèles discrets.	31
2.1.2.2	Les modèles continus.	32
2.1.2.3	Les modèles stochastiques : comprendre l'hétérogénéité du vivant.	32
2.1.2.4	Les modèles spatiaux : comprendre la dynamique spatio-temporelle	34
2.2	Exécution d'un modèle	34
2.2.1	Les approches exactes	34
2.2.2	Les approches approchées	35
2.2.3	Les approches hybrides : le rapport efficacité/précision	36
2.3	Conception d'un modèle	37
2.3.1	Choisir l'abstraction	37
2.3.2	Valider le modèle	38
2.3.3	Analyser la sensibilité et la robustesse du modèle	39
2.3.4	Choisir le formalisme en fonction des caractéristiques du sujet d'études	39
2.4	Le langage Kappa	41
2.4.1	Les modèles basés sur la réécriture de graphes à sites	41

2.4.2	La syntaxe de Kappa	43
2.4.3	La plateforme de modélisation Kappa	51
2.4.3.1	Simulation : KaSim.	51
2.4.3.2	Analyse statique : KaSa.	52
2.4.3.3	Analyse de Causalité : KaStor.	53
2.4.3.4	Équations différentielles ordinaires et réduction de mo- dèles : KaDe.	53
2.4.4	Les exemples d'utilisation du formalisme Kappa dans la littérature .	54
2.4.5	Les limites et extensions de Kappa	56
2.4.5.1	Les sites d'interaction.	56
2.4.5.2	Les notions d'espace en Kappa.	56
2.4.5.3	Les manques de l'interface utilisateur.	57
3	Apport de la biologie computationnelle dans l'étude de la fibrose et son diagnostic	58
3.1	Modéliser la fibrose hépatique	58
3.2	Modéliser la fibrose dans d'autres organes	63
3.3	Modéliser le signal TGFB1, moteur de la fibrose	67
4	Hypothèses de travail	73
II	Résultats	77
5	Le langage Kappa pour la modélisation multi-échelle, application à la dynamique des cellules étoilées	78
5.1	Model of activation of stellate cells	79
5.1.1	Overview of model	79
5.1.2	Some elements of biochemistry	80
5.1.2.1	Reaction half-time	80
5.1.2.2	Conversion	80
5.1.2.3	Production equilibrium	81
5.1.2.4	Erlang distributions	82
5.1.3	Interaction rules	82
5.1.3.1	The behavior of TGFB1 proteins	82
5.1.3.2	Renewal of quiescent HSCs	84

5.1.3.3	Activation and differentiation	85
5.1.3.4	Prolifération of activated hepatic stellate cells	87
5.1.3.5	Prolifération of myofibroblasts	88
5.1.3.6	Apoptosis and senescence of myofibroblasts	88
5.1.3.7	Inactivation of myofibroblasts	89
5.1.3.8	Behavior of inactivated hepatic stellate cells	90
5.1.3.9	Prolifération of reactivated cells	93
5.1.3.10	Degradation of reactivated <i>MFB</i>	94
5.1.3.11	Behavior of receptors	94
5.1.4	Parameters	96
5.2	Results	97
5.2.1	Static analysis	97
5.2.2	Underlying reaction network	99
5.2.3	Simulations	99
5.3	Discussion	101
5.4	Conclusion	105
6	Modélisation de la dynamique des cellules étoilées hépatiques, introduction des <i>tokens</i> et des <i>counters</i>	107
6.1	Trois modèles basés sur des règles pour la dynamique des cellules étoilées du foie	108
6.2	Syntaxe de modélisation	112
6.3	Calibration et estimation des paramètres	118
6.4	L'inactivation des myofibroblastes réactivés est essentielle pour maintenir l'accumulation de collagène 1 au cours d'une agression chronique du foie .	124
6.5	L'accumulation de collagène I est sensible à la dynamique de la stimulation par le TGFB1	128
7	Validation et prédictions biologiques du modèle :	131
7.1	Validation du modèle par de nouvelles données expérimentales chez la souris	131
7.2	Association entre l'accumulation des cellules étoilées inactivées et le score fibrose	133

III Discussion et Perspectives	141
A Annexes	153
A.1 Différents formalismes en biologie des systèmes	153
A.2 Communications et Formations	159
Bibliography	161

TABLE DES FIGURES

1	Organisation du foie à différentes échelles	15
2	Représentation schématique à l'échelle tissulaire et cellulaire du développement de la fibrose.	19
3	Représentation schématique des composants de la Matrice Extracellulaire .	20
4	Imagerie SHG dans la progression de la fibrose hépatique	21
5	Illustration de la dynamique des CEF	23
6	Cycle de recherche en biologie des systèmes	29
7	Comparaison des différentes approches de modélisation en fonction de leur degré de précision par rapport à leur coût de calcul	35
8	Comparaison des différentes approches de modélisation en fonction de l'échelle en distinguant les approches discrètes et continues.	40
9	Schéma de l'application d'une règle locale	43
10	Graphe de site	44
11	Modèle jeu en langage Kappa	46
12	Modèle de fibrose hépatique par Dutta-Moschetto <i>et al</i>	60
13	Modèle de fibrose hépatique par Wang <i>et al</i>	61
14	Modèle de fibrose hépatique par Yoshizawa <i>et al</i>	62
15	Modèle de fibrose hépatique par Friedman <i>et al</i>	63
16	Modèle de fibrose cardiaque par Rikard <i>et al</i>	66
17	Représentation schématique de l'activation du TGFB1	69
18	Carte de contact du modèle Kappa d'activation du TGFB1	72
19	Dynamique des cellules étoilées hépatique dans le premier modèle	79
20	Le comportement de la protéine <i>TGFB1</i>	83
21	Le renouvellement des cellules étoilées hépatiques quiescentes	84
22	Activation des cellules étoilées hépatiques et différenciation en myofibroblastes	86
23	Prolifération des cellules étoilées hépatiques activées	87

TABLE DES FIGURES

24	Prolifération des myofibroblastes	88
25	Voie d'apoptose/sénescence	89
26	Voie d'inactivation	90
27	Dégradation et réversion des cellules étoilées hépatiques inactivées	91
28	Redifférenciation des cellules étoilées hépatiques inactivées	92
29	Prolifération des cellules réactivées	93
30	Apoptose/sénescence des myofibroblastes réactivés	94
31	Cycle des récepteurs	94
32	Dynamique des récepteurs	95
33	Résultats de l'analyse statique	98
34	Simulations du premier modèle	100
35	Trois modèles pour décrire la dynamique des cellules étoilées (HSC)	110
36	Règle de renouvellement des HSC quiescentes	112
37	Règle de fixation du TGFB1 aux qHSC	113
38	Règles de dégradation et recyclage du récepteur au TGFB1	114
39	Règle du processus d'activation des HSC	115
40	Règle de prolifération des aHSC	116
41	Règles impliquées dans la dynamique du Collagène 1.	117
42	Règles d'élimination des MFB par : apoptose/senescence et inactivation	117
43	Analyses temporelles des cellules et de COL1 dans les modèles sans ou avec inactivation des react_MFB	125
44	Analyses temporelles des cellules et de COL1 dans les modèles sans ou avec réversion des iHSC vers un état quiescent	127
45	Analyse de l'évolution temporelle de COL1 en fonction de la variation de la concentration, du nombre et de la périodicité des stimulations par le deTGFB1	130
46	Comparaison des prévisions du modèle et des nouvelles données expérimentales acquises dans un modèle de fibrose induite par le CCl4	134
47	Comparaison des prévisions du modèle et des données expérimentales acquises dans des modèles de fibrose induite par la diméthyl_nitrosamine, le thioacétamide et un régime riche en graisse	135
48	Accumulation des iHSC au cours des répétitions du cycle agression/réversion	137

49 L'accumulation des iHSC est associée à la fibrose hépatique chez les patients atteints de NASH 140

LISTE DES TABLEAUX

1	Comparaison des différents formalismes	41
2	Modèles de signalisation du TGFB1	70
3	Tableau des paramètres du premier modèle	96
4	Tableau des paramètres du modèle final	122
5	Liste des gènes utilisés pour l'analyse GSEA	138
6	Tableau des formations réalisées	160

PREMIÈRE PARTIE

Introduction

LA FIBROSE HÉPATIQUE

1.1 Les fonctions du foie

Le foie possède plusieurs fonctions, parmi lesquelles : la fabrication de la bile et de la plupart des protéines sanguines, la transformation et le stockage des nutriments absorbés par le tube digestif, ainsi que la détoxification du sang. Le foie contrôle des centaines de voies métaboliques dont les produits sont relargués dans le sang et redistribués à tout l'organisme. On trouve notamment parmi ces produits le glycogène, l'urée et le cholestérol, qui proviennent respectivement du métabolisme des glucides, des protéines et des lipides. Une fonction du foie est de stocker certaines molécules, par exemple la vitamine A ou le glycogène, et de les relarguer dans l'organisme en fonction des besoins. Enfin, le foie possède une fonction de détoxification des substances endogènes comme l'urée, ou exogènes comme l'alcool ou les médicaments. Les fonctions hépatiques impliquent des échanges importants entre les hépatocytes qui assurent ces fonctions (voir paragraphe 1.2) et la circulation sanguine. Ces échanges sont rendus possibles par une vascularisation importante du foie et par son organisation spécifique. Le foie est composé de quatre lobes divisés en huit segments organisés en plusieurs millions d'unités fonctionnelles appelées lobules hépatiques (voir en Figure 1 Organisation vasculaire du foie), dans lesquels les hépatocytes entourent des micro-vaisseaux sanguins appelés sinusoides. Les sinusoides sont alimentées par l'artère hépatique qui fournit un sang riche en oxygène et par la veine porte qui fournit les nutriments en provenance du système digestif. Avec le canal biliaire, l'artère hépatique et la veine porte forment un ensemble structurel appelé espace porte (voir Figure 1 Organisation d'un lobule hépatique). À l'autre extrémité de la sinusoides, le sang est évacué du lobule par la veine centrale qui rejoint la veine cave inférieure par l'intermédiaire des veines sus-hépatiques. Le sang circule entre les travées d'hépatocytes dans les sinusoides hépatiques, ce qui permet l'échange de substance avec les cellules. Les sinusoides sont délimitées par des cellules endothéliales spécialisées (voir 1.2) qui favorisent les échanges entre le sang et les hépatocytes. Cependant, les hépatocytes ne

sont pas directement au contact de ces cellules endothéliales dont ils sont séparés par un espace interstitiel péri-sinusoïdal appelé espace de Disse (Voir Figure 1 Organisation du sinusoiïde).

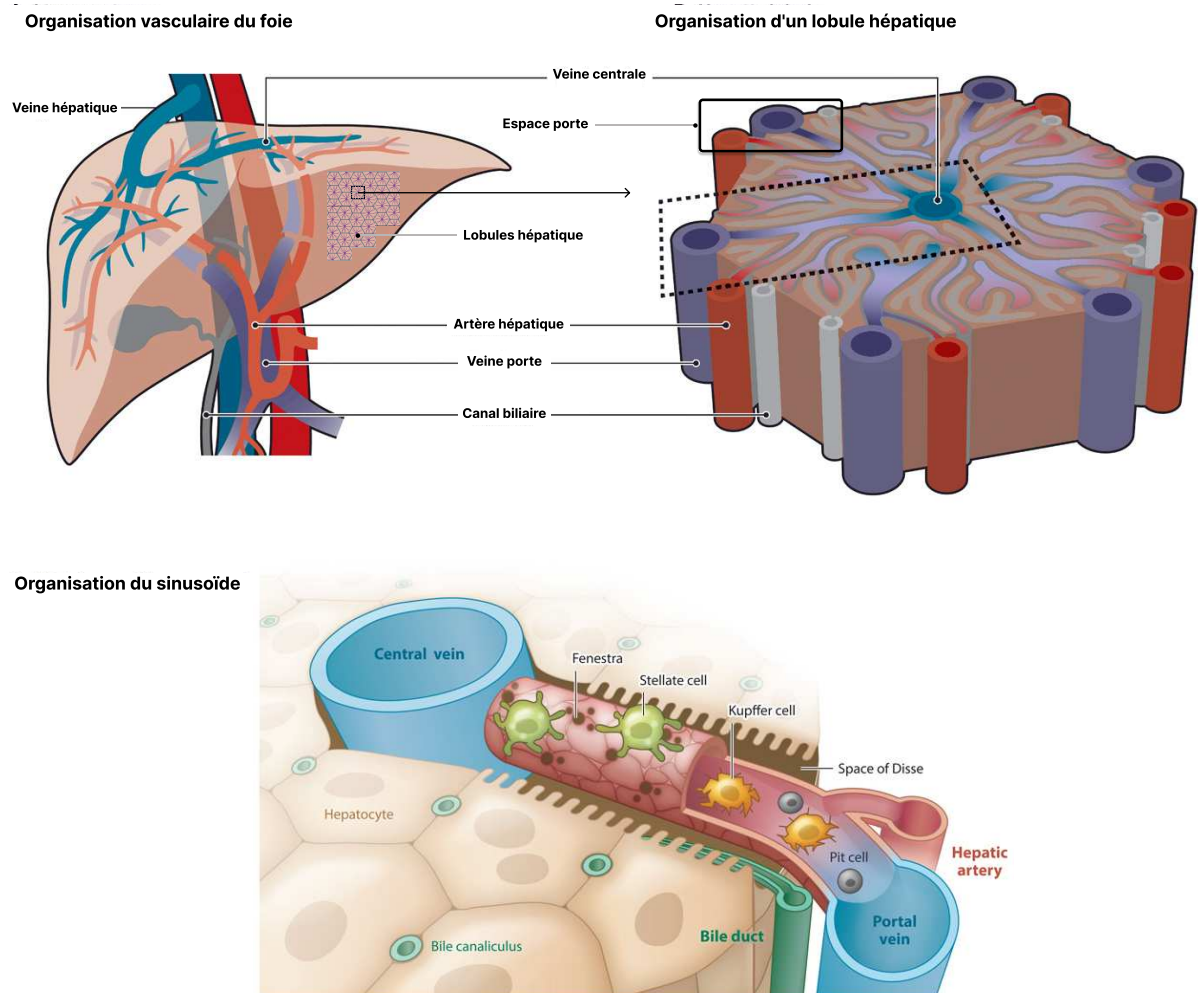


FIGURE 1 – Organisation vasculaire du foie, Organisation d’un lobule hépatique (adapté de [93]), Organisation du sinusoiïde [79]

Le foie est doublement vascularisé par la veine porte et l’artère hépatique qui apportent respectivement le sang riche en nutriment et oxygène, qui ressortira par la veine hépatique. Sur la partie extérieure de chaque lobule est présent un espace porte composé : du canal biliaire, de l’artère hépatique et de la veine porte par lequel le sang arrive. Ce sang passe ensuite dans le sinusoiïde pour ressortir par la veine centrale. Le sinusoiïde est un capillaire fenêtré, composé de cellules endothéliales sinusoiïdales (rouge), de Cellules étoilées hépatiques qui enveloppent le capillaire et de cellules de Kupffer (jaune). De part et d’autre du sinusoiïde se trouve des hépatocytes (beige) qui sont séparés du sinusoiïde par l’espace de Disse.

1.2 Les cellules du foie

Le foie est constitué de différents types cellulaires que l'on retrouve dans tous les lobules hépatiques et parmi lesquelles on peut distinguer deux catégories : les cellules parenchymateuses et non-parenchymateuses. Les cellules parenchymateuses sont des cellules épithéliales (hépatocytes et cholangiocytes) qui assurent les principales fonctions hépatiques. Les cellules non parenchymateuses correspondent à toutes les autres cellules présentes dans les lobules, parmi lesquelles on peut citer : les cellules endothéliales sinusoidales du foie (CESF), les cellules de Kupffer et les cellules étoilées du foie (CEF). En plus de ces deux catégories, le foie possède des cellules souches intra-hépatiques appelées cellules progénitrices du foie (CPF) qui peuvent se différencier en hépatocytes ou en cholangiocytes.

Les hépatocytes

Les hépatocytes sont les cellules les plus nombreuses du foie, et assurent la majorité des fonctions métaboliques du foie, ainsi que la sécrétion de bile. Ces cellules sont organisées en travées autour de l'espace de Disse. Elles présentent une polarité particulière, avec un domaine apical délimitant les canalicules biliaires et un domaine basal au contact de l'espace de Disse. Les microvillosités du domaine basal permettent d'augmenter leur surface d'échange avec les sinusoides hépatiques. L'activité métabolique des hépatocytes dépend de leur localisation le long du sinusoides. Les hépatocytes proches de l'artère hépatique, riche en oxygène, participent aux métabolismes oxydatifs (p.e, gluconéogénèse, β -Oxydation) alors que les hépatocytes proches de la veine porte, riche en nutriments et déchets, sont impliqués principalement dans les processus de dégradation (p.e, glycolyse) et de détoxification.

Les cholangiocytes

Les cholangiocytes sont, comme les hépatocytes, des cellules épithéliales. La majorité de ces cellules bordent les canaux biliaires situés dans les espaces portes des lobules hépatiques. Le rôle de ces cellules est de réguler la composition de la bile en sécrétant ou en absorbant l'eau ou les solutés (p.e, électrolytes, bicarbonates) présents dans la bile. Ces cellules ont aussi un rôle de maintien de l'intégrité de la barrière épithéliale des canaux biliaires.

Les cellules endothéliales sinusoidales du foie (CESF)

Les cellules endothéliales sinusoidales du foie sont les cellules constituant la paroi du sinusoides. Ces cellules sont spécialisées grâce à la présence de pores, appelés fenestrae, qui traversent le cytoplasme. Cette particularité structurelle des CESFs permet de faciliter les échanges entre le sang circulant et les hépatocytes et est donc indispensable à la fonction hépatique.

Les cellules de Kupffer

Les cellules de Kupffer sont les macrophages résidents du foie, elles se trouvent dans la lumière des sinusoides. En tant que cellules immunitaires, les cellules de Kupffer participent à la réponse immédiate en cas d'agression par des agents pathogènes ou des toxines endogènes. Ces cellules jouent donc un rôle crucial dans le maintien de l'homéostasie hépatique.

Les cellules progénitrices du foie (CPF)

Les cellules progénitrices du foie (CPF) se trouvent dans le canal de Hering situé dans l'espace porte et qui fait la jonction entre les canalicules et les canaux biliaires. En cas de lésion hépatique, elles participent à la régénération des hépatocytes et des cholangiocytes et expriment des marqueurs fœtaux spécifiques de ces deux lignées cellulaires.

Les cellules étoilées du foie

Les cellules étoilées du foie (CEF) sont des cellules mésenchymateuses, situées dans l'espace de Disse, entre les hépatocytes et le sinusoides. Dans des conditions physiologiques, les CEF se trouvent à l'état quiescent et leur fonction principale est de stocker la vitamine A. Cependant, en réponse à une lésion hépatique, ces cellules sont activées par des cytokines inflammatoires, perdent leur capacité de stockage, changent de phénotype et prolifèrent. Les rôles et les fonctions de ces cellules seront vues plus en détails dans la section 1.4

1.3 Les maladies chroniques du foie

Le développement des maladies chroniques du foie est associé principalement aux infections virales (virus de l'Hépatite C et B), aux expositions à des substances toxiques (p.e. alcool, médicaments), aux maladies du foie gras non alcoolique (NAFLD) principalement associée à la résistance à l'insuline, au diabète de type 2, à l'obésité et aux troubles métaboliques, la plus commune étant la stéatose hépatique non-alcoolique (NASH). Les agressions hépatiques provoquent des lésions, elles entraînent un processus de réparation qui consiste en une réaction inflammatoire suivie d'une régénération cellulaire conduisant à la restauration du tissu. Dans le cas d'agressions chroniques, l'inflammation persiste et conduit au développement d'une fibrose hépatique. La cirrhose est le stade final de la fibrose et se caractérise par une désorganisation de l'architecture du foie et constitue le risque majeur de survenue du carcinome hépatocellulaire (CHC). La fibrose se caractérise par une accumulation excessive de matrice extracellulaire (MEC) dans l'espace périsinusoïdal induisant des changements dans le phénotype des cellules environnantes, notamment une réduction des microvillosités des hépatocytes et une perte des fenestrations des cellules endothéliales qui signent la "capillarisation" du sinusoiide hépatique. L'accumulation des composants de la MEC réduit les échanges de substances solubles entre le compartiment sanguin et les cellules parenchymateuses, conduisant à un dysfonctionnement du foie (Figure 2).

La Matrice Extracellulaire

La MEC est une structure non-cellulaire composée de glycoprotéines, de protéoglycanes, de collagènes, de fibres élastiques et de glycosaminoglycanes. C'est un réseau moléculaire complexe et dynamique qui joue un rôle majeur dans le support physique des cellules et la communication cellulaire (Figure 3). Elle module la biodisponibilité et l'activité des cytokines et des facteurs de croissance et joue un rôle biomécanique (en exerçant des forces de tension sur les cellules).

Dans le foie normal, la MEC est principalement constituée de collagènes et notamment des collagènes I, III, IV et V, ainsi que de glycoprotéines (fibronectine, laminine) et de protéoglycanes (acide hyaluronique). Certaines de ces protéines sont des constituants de la membrane basale qui sépare les épithélia et les vaisseaux sanguins du tissu conjonctif. Dans le foie, l'espace de Disse représente cependant un cas particulier : il contient des constituants de la lame basale tels que le collagène IV et la laminine, mais pour former

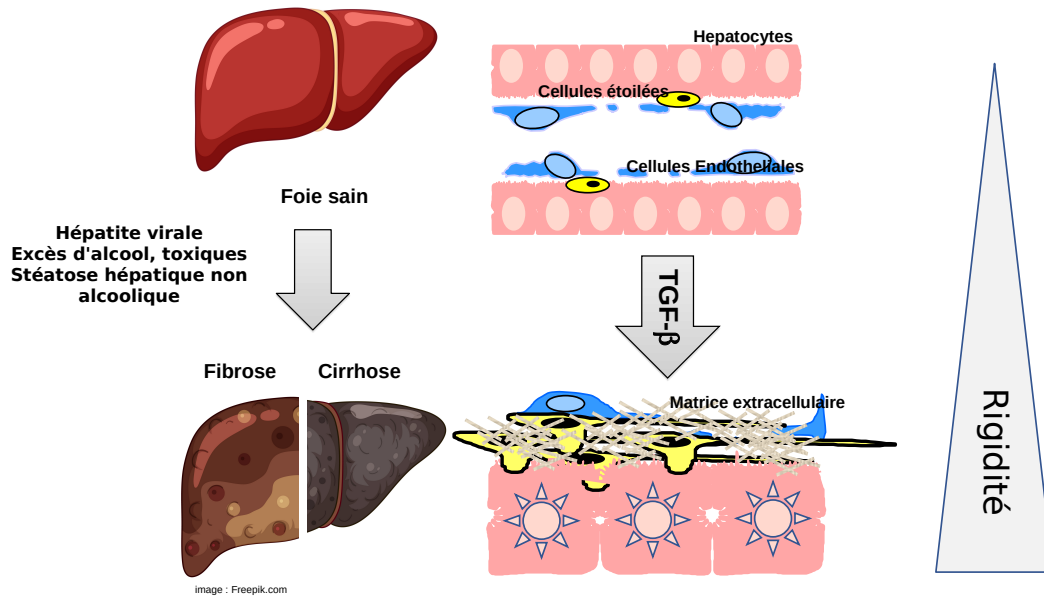


FIGURE 2 – Représentation schématique à l'échelle tissulaire et cellulaire du développement de la fibrose. Les agressions hépatiques provoquent le mort des hépatocytes, ce qui induit la mise en place d'un processus inflammatoire dont le TGFβ1 fait partie. Le dépôt de MEC par les CEF, sous induction du TGFβ1, induit une augmentation de la rigidité. En cas d'agression chronique, l'augmentation de la rigidité induira le développement d'une fibrose qui pourra, en cas de maintien des agressions, se développer en cirrhose.

une matrice bien plus lâche qu'une lame basale ordinaire. Cette caractéristique de l'espace de Disse permet la diffusion des molécules entre le sang sinusoidal et les hépatocytes, via les fenestreae des cellules endothéliales. Dans les maladies chroniques, non seulement la MEC s'accumule dans le sinusoïde, mais sa composition se modifie avec un enrichissement notamment en collagène de type 1. Le développement des approches protéomiques a contribué à mieux caractériser les altérations de la MEC au cours de la fibrose. Le matrisome, c'est-à-dire l'ensemble de composants de la MEC, est modifié très précocement en réponse à une atteinte hépatique [178] mais il présente une grande hétérogénéité dépendante des modèles de fibrose étudiés [10].

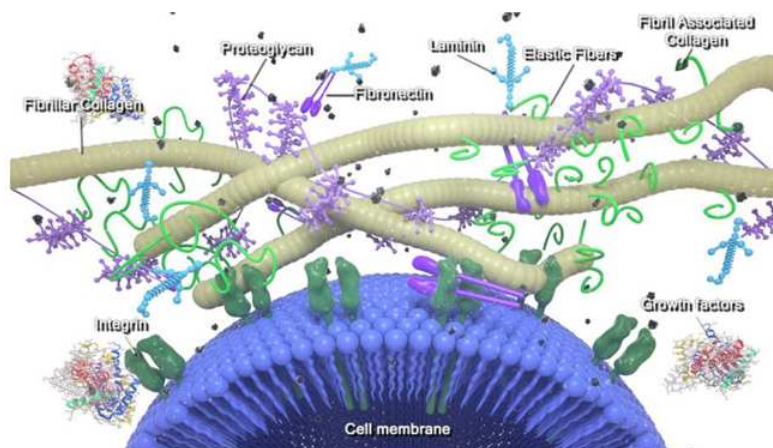


FIGURE 3 – Représentation schématique des composants de la Matrice Extracellulaire, détaillant leur organisation et interaction avec les cellules [142].

Évaluation clinique de la fibrose

La biopsie hépatique constitue l'examen de référence. Elle permet une appréciation anatomo-pathologique de la fibrose en quantifiant et localisant le dépôt de matrice sur les coupes histologiques après coloration au rouge Sirius ou au Trichrome de Masson qui révèle le collagène fibrillaire. Ces analyses histochimiques de tissu conduisent à l'établissement de scores de fibrose semi-quantitatifs. Le score METAVIR s'appuie sur l'évaluation de la fibrose et de l'activité nécrotico-inflammatoire des tissus. Définition du dictionnaire médicale : *La fibrose comprend cinq grades : F0 (absence), F1 (atteinte minimale), F2 (lésions caractérisées par des septas fibreux en pont reliant deux espaces portes ou un espace porte et une veine sus-hépatique) considérée comme « significative », grade à partir duquel un traitement est habituellement proposé. Le grade F3 est pré-cirrhogène avec une fibrose arciforme qui n'est pas encore annulaire. Le grade F4 correspond à une cirrhose histologique, c'est-à-dire une fibrose annulaire qui entoure des nodules de régénération. L'activité nécrotico-inflammatoire comprend quatre grades : A0 (sans activité), A1 (activité minimale), A2 (activité modérée), A3 (activité sévère).* Plus récemment, les techniques de microscopie de génération de seconde harmonique (SHG) se sont développées, elles permettent une visualisation des fibres de collagène sur les coupes sans marquage [97]. Cette méthode est plus sensible et plus quantitative que la coloration au rouge Sirius et permet également d'observer la texture du collagène, par exemple l'orientation des fibres (Figure 4). Bien que plus précise, cette technique requière une instrumentation qui n'est pas encore accessible en routine à l'hôpital.

Il existe des techniques non invasives qui reposent d'une part sur l'évaluation de marqueurs sériques et d'autre part sur des mesures de l'élasticité hépatique

La prise de sang permet la détection de marqueurs de fibrose comme des composants issus de la matrice extracellulaire (p.e. le propeptide N-terminal du procollagène de type III et l'inhibiteur des métalloprotéase TIMP1) ou des marqueurs hépatiques comme dans le Fibrotest® (haptoglobine, gGT, bilirubine, a2-macroglobuline, apolipoprotéine A1); le test Fibromètre® (a2-macroglobuline, acide hyaluronique, plaquettes, taux de prothrombine, ASAT, ALAT, urée) et l'Actitest (Alpha-2-macroglobuline, haptoglobine, apolipoprotéine-A1, bilirubine totale, g-glutamyl-transpeptidase). le score FIB-4 associe l'âge du patient, le taux de plaquettes et les transaminases (ASAT et ALAT), il est recommandé comme outil de dépistage avec une orientation vers des tests plus spécifiques en cas de positivité.

Les approches de mesure de l'élasticité hépatique utilisée en évaluation clinique selon un rapport de l'Association Française pour l'Etude du Foie (AFEF) sont : l'élastographie impulsionnelle (VICE) qui est la plus utilisée et communément appelée Fibroscan®; l'élastographie par ondes de cisaillement ponctuelles (SWE); la SWE avec ondes bidimensionnelles (2D-SWE) et l'élasto-IRM [13].

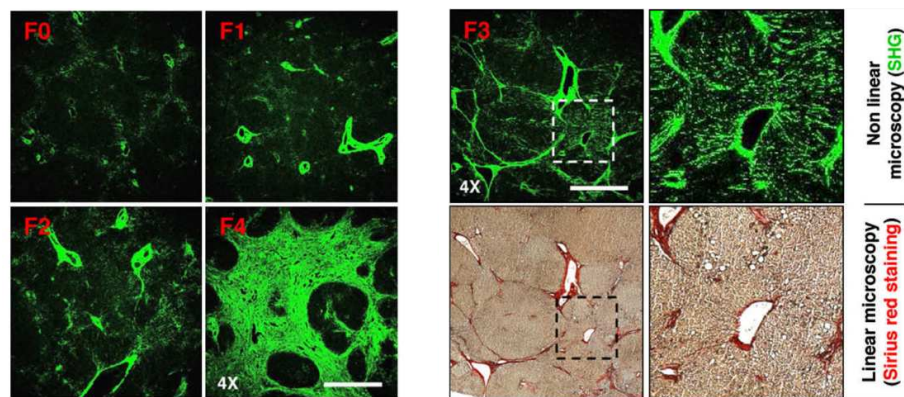


FIGURE 4 – Imagerie SHG dans la progression de la fibrose hépatique[97]

Panel de gauche : Images SHG représentant l'évolution caractéristique des dépôts de collagène fibrillaire d'un foie normal (F0-Metavir) à un foie cirrhotique (grade F4-Metavir). Les grades F-Metavir ont été déterminés par un pathologiste à l'aide du système de notation Metavir. Panel de droite : Comparaison entre l'imagerie SHG et la microscopie linéaire après coloration au rouge de Sirius (biopsie F3-Metavir). Les images de droite montrent l'imagerie à fort grossissement réalisée dans la zone délimitée par le carré en pointillés.

1.4 Les cellules étoilées du foie : moteur de la fibrose

Les cellules étoilées du foie ont été découvertes et décrites par le Dr Carl von Kupffer en 1876. Elles sont toutefois distinctes des cellules dites "de Kupffer" qui sont les macrophages résidents du foie localisés dans les sinusoides. Les cellules étoilées, détectées en utilisant une méthode histochimique au chlorure d'or, sont localisées dans l'espace de Disse, entre le tissu parenchymateux (hépatocytes) et le sinusoides. En 1952, le Pr Toshio Ito mit en lumière la capacité de stockage de gouttelettes lipidiques de ces cellules [203]. La standardisation du nom "cellules étoilées du foie" (en anglais : Hepatic Stellate Cells, HSCs) a pris place en 1996, après avoir été désignées sous les termes de cellules perisinusoidale, péricytes, lipocytes, cellules interstitielles, cellules stockant les lipides, cellules stockant la vitamine A, ou Ito cells (« Hepatic Stellate cell nomenclature Letter », 1996). Elles sont aujourd'hui communément appelées CEF (Cellules étoilées du foie).

Dynamique fonctionnelle des cellules étoilées

Les cellules étoilées ont un phénotype quiescent, non prolifératif dans le foie sain, où elles stockent 80 à 90 % du rétinol total du corps dans des gouttelettes lipidiques [227]. En réponse à une atteinte hépatique et à la réaction inflammatoire induite, les CEF sont activées et perdent leurs gouttelettes lipidiques pour se transdifférencier en myofibroblastes (MFB) (Figure 5) qui contribuent à la réparation du tissu. En absence de nouvelles atteintes hépatiques, les lésions sont réparées, les CEF activées et MFB sont éliminés, le produit cicatriciel est résorbé et l'architecture hépatique reconstituée. Si le signal inflammatoire persiste, l'accumulation de cellules produisant du collagène conduit à la formation d'une fibrose. Cependant, la réversibilité ou la régression de la fibrose hépatique a été observée dans des modèles expérimentaux et des études cliniques après l'élimination de l'agent inflammatoire [144, 39]. Dans ce paragraphe, nous détaillons les mécanismes d'activation et d'élimination des CEF.

Activation des CEF

Il est généralement admis que le processus d'activation se déploie en deux phases [257] : (1) l'initiation qui correspond aux changements précoces de l'expression des gènes et du phénotype en réponse aux stimuli issus des lésions (2) La perpétuation qui résulte ensuite des effets de ces stimuli sur le maintien du phénotype activé. Les lésions provoquant la mort cellulaire des hépatocytes (hépatolyse) entraînent le relargage de corps

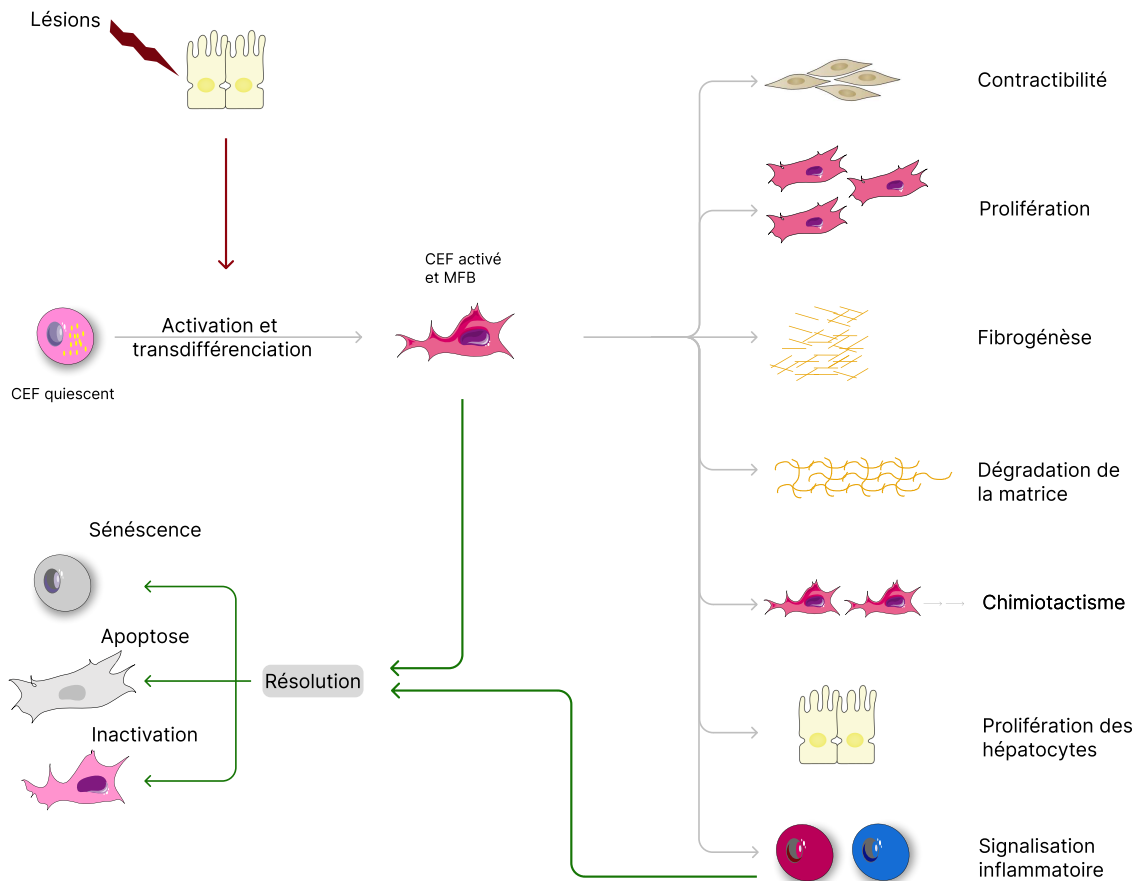


FIGURE 5 – Illustration de la dynamique des CEF suite à des agressions (inspiré de [257]). Les agressions hépatiques induisent l’activation des CEF quiescent qui vont ensuite se transdifférencier en MFB. Les CEF activés et MFB acquièrent de nouveaux phénotypes tels que : la contractibilité, la prolifération, la production et de la dégradation de MEC, le chimiotactisme, la production de molécules inflammatoires. En cas de réversion, les MFB sont éliminés par trois voies, l’apoptose, la sénescence et l’inactivation.

apoptotiques, d'espèces réactives de l'oxygène (ROS) et de DAMP (DAMP, Damage Associated Molecular Pattern) mitochondriaux qui tous contribuent à l'activation des CEF notamment en recrutant des cellules immunitaires, dont les cellules de Kupffer et les monocytes. En réponse à ces signaux, les cellules immunitaires orchestrent une réponse inflammatoire en produisant des cytokines et des facteurs de croissance comme les interleukines IL-1 et IL-6, et surtout le TGFB1 qui vont initier l'activation des CEF. Au cours de ce processus d'activation, les CEF acquièrent de nouvelles propriétés : prolifération, inflammatoire, chimiotactisme, sécrétion de composants matriciels, activation des voies de survie, contractilité avec notamment l'expression de la protéine Alpha smooth muscle actin (α -SMA). Les CEF contribuent à la perpétuation de leur phénotype myofibroblastique en produisant des cytokines fibrogéniques comme le TGFB1 et en rigidifiant son microenvironnement (sécrétion de protéines fibrillaires et d'inhibiteurs de collagénases). La rigidité de la matrice induit un signal mécanique nécessaire au maintien du phénotype myofibroblastique [197, 38, 157, 152].

Les CEF activées/MFB possèdent donc une grande plasticité phénotypique qui contribue à maintenir un état propice à leur perpétuation. La coexistence de ces multiples phénotypes a été particulièrement bien caractérisée grâce à l'apport des méthodes de Séquençage des ARNm sur cellules uniques [153, 288, 249, 214, 283]. Il est intéressant de noter que les programmes transcriptionnels qui conduisent à l'initiation des CEF sont toujours actifs chez des patients présentant une cirrhose [66].

Élimination des CEF

Différents mécanismes pouvant contribuer à l'élimination des CEF ont été identifiées : la mort programmée, la sénescence et l'inactivation.

- Les mécanismes de mort programmée se regroupent aujourd'hui sous différentes terminologies suivant les mécanismes/acteurs impliqués et si l'apoptose est le mécanisme le plus documenté pour l'élimination des CEF, d'autres mécanismes comme la necroptose, la pyroptose et la ferroptose ont été plus récemment identifiés [98]. L'analyse des mécanismes de l'apoptose des CEF en lien avec la réversibilité de la fibrose a fait l'objet de nombreux travaux [82]. Ces événements sont globalement liés à la production de facteurs pro-apoptotiques dans l'environnement extracellulaire par les cellules immunitaires. Par exemple, les cellules NK (natural killers) libèrent de l'IFN- γ qui induit l'apoptose des MFB. Dans une boucle de rétro-signalisation, l'IFN- γ stimule en retour les cellules NK pour induire d'autres voies de signalisation pro-apoptotiques telles que TRAIL, FASL ou

la production de granzyme [132, 133, 247]. Plus récemment, l'induction de l'apoptose des CEF par les lymphocytes T CD8+ mémoire a été démontré dans un modèle murin de NASH induit par un régime riche en graisses et en cholestérol [149], ceci en initiant, eux aussi, la voie FASL ou *via* l'utilisation de granzyme/perforine.

- La sénescence cellulaire est un processus de vieillissement qui se traduit par un arrêt de la division cellulaire. Comme l'apoptose, la sénescence des CEF a été impliquée dans le processus de réversion de la fibrose. Les CEF sénescents contribuent à limiter la progression de la fibrose [156, 223]. Elles ne prolifèrent plus et perdent leur phénotype myofibroblastique, avec une diminution de production des protéines fibrogéniques (Collagènes I et III, fibronectine) et une surexpression des métalloprotéases matricielles (MMPs). Par ailleurs, les cellules sénescents produisent et sécrètent des protéines spécifiques (cytokines inflammatoires, chimiokines, facteurs de croissance et facteurs de remodelage de la matrice) qui contribuent au remodelage du microenvironnement et à l'inflammation chronique [194].

- L'inactivation est un processus de réversion où les CEF activées/MFB retournent à un état proche de l'état des CEF quiescentes, mais avec un phénotype et un profil d'expression génique différent. Il y a très peu d'information concernant le devenir de ces CEF inactivées, seule leur capacité de réactivation plus rapide que les CEF quiescentes a été mise en lumière pour le moment [143, 256, 80].

L'apport des approches de séquençage ARN en cellules individuelles

Les approches de séquençage ARN en cellules individuelles (single cell RNA-seq ou scRNAseq) ont récemment aidé dans l'étude des dynamiques de ces différents phénotypes dans le contexte de la progression fibrotique. Dès 2019, Krenkel *et al.* [153] ont ainsi identifié dans un modèle murin de fibrose induite par le tétrachlorure de carbone (CCl₄), 4 clusters de MFB avec des phénotypes différents : i) sécrétion de composants matriciels, ii) inflammatoire, iii) prolifération et iv) phénotype mixte associant divers modulateurs de la MEC et des marqueurs, dont certains sont des marqueurs des fibroblastes portaux. Par la suite, plusieurs travaux [288, 249, 214, 283], ont non seulement confirmé l'hétérogénéité des phénotypes de MFB en identifiant en général 4 clusters, mais ils ont aussi apporté une vision dynamique entre ces phénotypes. En effet, ces nouvelles études ont utilisé différents algorithmes (Monocle[253] et Slingshot [241]) permettant de reconstruire

des trajectoires/chemins de la dynamique d'activation des CEF et de retracer un historique hypothétique entre les différents phénotypes identifiés par clustering. Le principe de ces algorithmes repose sur le re-ordonnement de chaque cellule (à partir de ses données de séquençage) selon leur position dans un chemin de transformation (activation, différenciation). Grâce à ces approches de scRNAseq :

- les résultats de Zhang *et al.* [288] obtenus sur un modèle murin de fibrose induite par le CCl4 mettent en évidence, par approche pseudo-time", trois voies/trajectoires majeures d'activation des CEF i) deux conduisant à deux clusters intermédiaires de MFB distincts, l'un à phénotype plus fibrosant, l'autre plus inflammatoire, ii) une voie/trajectoire unique conduisant à des MFB associés à la réparation. À côté de ces trois clusters, les quatre autres clusters présentent des profils transcriptomiques mixtes.
- Les résultats de Terkelsen *et al.* [249] mettent en évidence une dynamique transcriptomique des CEF au cours de leur activation qui se caractérise d'abord par un profil d'expression de gènes associés au "soutien à la morphogenèse des vaisseaux sanguins", à la prolifération et à la différenciation cellulaire vers un profil lié à l'adhésion cellulaire, l'organisation de la MEC, au développement du tissu musculaire lisse et à la traduction des protéines.
- Les résultats de Yang *et al.* [283] montrent que les trajectoires d'activation des CEF est différentes selon le modèle murin de fibrose utilisé, qu'il soit induit par le CCl4 ou par ligation du canal cholédoque.
- Les résultats de Rosenthal *et al.* [214] obtenu sur un modèle murin de fibrose induite par un régime type occidental (NASH) et sa réversion mettent en évidence la trajectoire d'apparition des CEF inactivées au cours de la régression. C'est ce jeu de données qui sera utilisé pour valider les prédictions de notre modèle (voir Partie Résultats 7).

1.5 Les modèles d'études expérimentales de la fibrose

L'étude de la fibrose hépatique repose sur l'utilisation de matériel humain (biopsies), de modèles animaux (essentiellement Rat et Souris) et de modèles in vitro (culture et coculture de cellules hépatiques primaires ou en lignées). Chez l'homme, les biopsies hépatiques sont utilisées pour l'analyse histochimique des tissus lésés (localisation des marqueurs protéiques, métaboliques ou ARN), associée à la caractérisation clinique des patients (score

de fibrose, profil inflammatoire, etc.). Pour les modèles animaux [284, 281], les atteintes hépatiques peuvent être induites par des méthodes chirurgicales, diététiques ou chimiques. La méthode chirurgicale la plus répandue consiste à mimer une cholestase obstructive par ligation du canal cholédoque. Des régimes alimentaires riches en graisses sont à la base des modèles de stéatose hépatique non alcoolique (NASH), une pathologie dont l'incidence mondiale a augmenté, due à la généralisation de certaines habitudes alimentaires occidentales. Parmi les méthodes chimiques, la plus répandue consiste à traiter les souris par ingestion de tétrachlorure de carbone (CCl₄) [225]. Cette approche permet de développer des modèles d'atteintes aiguës (une seule injection de CCl₄) ou chroniques (plusieurs injections par semaine pendant plusieurs semaines). Les modèles d'atteinte aiguë permettent de suivre la réaction inflammatoire et la réparation tissulaire (cicatrisation) et, à l'échelle cellulaire, l'activation des cellules étoilées et leur élimination. Les modèles chroniques permettent d'étudier les mécanismes cellulaires impliqués dans le développement et la progression de la fibrose. C'est ce modèle qui a été utilisé pour étudier la plasticité phénotypique et la dynamique des populations de CEF pendant la progression, la réversion et la récurrence de la fibrose hépatique [143]. Le suivi dynamique des populations de CEF est rendu possible grâce à des stratégies de marquage génétique : expression d'une protéine fluorescente sous le contrôle d'un promoteur spécifique des cellules étoilées. L'utilisation des modèles cellulaires présente l'avantage de réduire l'utilisation des animaux de laboratoire, mais les observations faites dans ces modèles nécessitent toutefois d'être validées chez l'animal. Ces modèles consistent à cultiver ou co-cultiver différents types de cellules hépatiques : Hépatocytes, CEF, cellules immunitaires, endothéliales, etc. Ces modèles utilisent des cellules primaires purifiées à partir de pièces opératoires, mais cette source est très difficile à obtenir en routine. Une alternative consiste à utiliser des lignées cellulaires permanentes dérivées de ces différents types cellulaires. En fonction des conditions de culture, le phénotype de ces lignées peut toutefois s'éloigner de celui des cellules d'origine. La revue de Yanguas *et al.* [284] décrit en détail les différentes lignées cellulaires pouvant être utilisées *in vitro* ainsi que les avantages et inconvénients de chacune d'entre elles.

MODÉLISER LES SYSTÈMES BIOLOGIQUES

La modélisation des systèmes biologiques est une branche de la biologie computationnelle qui vise à étudier le comportement des organismes à différentes échelles : organes, tissus, cellules, molécules en décrivant les liens et les interactions entre les éléments composant le système et en appliquant des approches mathématiques et divers concepts théoriques. La diversité des approches et des méthodes utilisées en biologie des systèmes permet autant d'étudier les voies d'expression de gènes que la régénération d'un organe complet.

Nous proposons la définition suivante : Un modèle computationnel est une abstraction numérique de la réalité biologique, dans lequel l'évolution des molécules est décrite par des lois dynamiques dans un formalisme mathématique adapté au niveau d'abstraction. La simulation *in silico* des lois du système dynamique permet de prédire les évolutions dans le temps de l'état des molécules. L'objectif d'un modèle est de répondre à des questions biologiques sous forme de prédictions qui vont aiguiller de nouvelles études expérimentales qui pourront en retour nourrir le modèle (Figure 6). Le processus de modélisation permet aussi de mettre en évidence des manques dans les connaissances ou leurs contradictions. Les modèles permettent de tester des hypothèses qui nécessiteraient la réalisation d'expérimentations complexes, coûteuses et difficiles à mettre en place par le biologiste. Un autre avantage des modèles est la possibilité de les réutiliser en les adaptant à de nouvelles questions biologiques.

2.1 Catégories de modèles

La description d'un modèle, quel que soit son type, reposera la plupart du temps sur un ou plusieurs graphes dont les nœuds représentent des espèces (p.e. protéines, cellules) et les arêtes représentent des relations entre ces espèces. Il existe différents types de

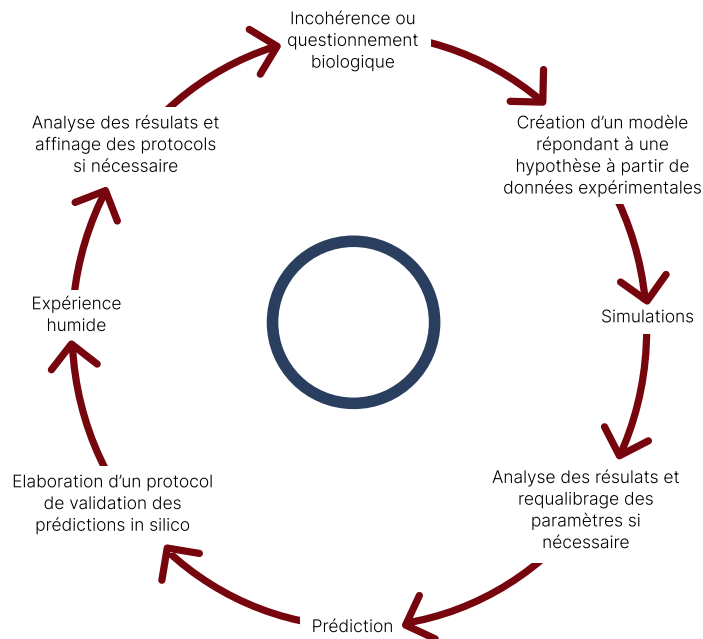


FIGURE 6 – Cycle de recherche de biologie des systèmes. Ce cercle montre les différentes étapes nécessaires à l’élaboration d’une étude des systèmes biologiques. L’interdépendance entre les expériences *humides* et *sèches* est indispensable pour ces approches. Les expérimentations *humides* ont besoin des prédictions des modèles pour orienter et réduire le champ des possibilités. Les expériences *sèches* ont, en retour, besoin des données expérimentales pour définir l’hypothèse de travail et être implémentés à partir des données obtenues (inspiré de [147]).

graphes. Les graphes *non dirigés* dans lesquels aucune direction n’est associée aux arêtes reliant les nœuds. Les graphes *dirigés*, dans lesquels une direction est associée aux arêtes reliant les nœuds (représentés sous forme de flèches). Il est possible d’ajouter un poids aux arêtes, celui-ci peut être de valeur variable, positive ou négative, on parle alors de graphes *pondérés*. Le but de ces poids est de représenter la direction et le poids des influences entre les nœuds du modèle (inhibition, production) [276]. Dans certains cas, le graphe peut posséder deux sous-ensembles de nœuds différents tels que chaque arête relie un nœud d’un des sous-ensembles à un nœud de l’autre sous-ensemble. On parle alors de graphes *bipartis*.

Il existe deux grandes catégories de modèles : les modèles statiques basés sur l’analyse des relations entre les espèces et modèles numériques dynamiques, généralement des sys-

tèmes dynamiques discrets ou continus qui suivent le comportement d'espèces au cours du temps.

2.1.1 Les modèles statiques : identifier les relations

Les modèles statiques reposent principalement sur la structure de graphes et la recherche de liens de causalité entre les éléments qui les composent. Il existe diverses approches d'analyse topologique permettant d'extraire des informations de ces graphes [200]. Nous en citons quelques-unes ci-dessous.

- La recherche de motifs consiste à rechercher des motifs d'interconnexions qui sont plus fréquents que dans un graphe aléatoire. Un motif répété indique que ce motif peut jouer un rôle important dans le modèle et qu'il peut s'agir d'un élément de régulation.
- La recherche de *clusters* permet d'identifier des nœuds d'un graphe qui partagent statistiquement plus de liens que dans un graphe généré de manière aléatoire. Ceci permet d'identifier des sous-groupes de réactions au sein du graphe et de mettre en lumière certaines parties ou groupes de réactions fortement connectées.
- La recherche de centralité permet de définir l'importance de certains nœuds dans le graphe en identifiant, par une mesure locale, par exemple le nombre de nœuds auxquels ils sont reliés. Une des approches de centralité la plus communément utilisée en analyse de graphes est la centralité d'intermédiaires, elle repose sur l'identification du ou des nœuds les plus représentés sur les chemins les plus courts entre deux autres nœuds du graphe. Plus un nœud est présent sur les chemins les plus courts, plus celui-ci aura un fort impact sur le reste du graphe.

L'utilisation et l'analyse topologique des graphes permet donc de mettre à jour des éléments clefs dans des processus biologiques complexes et d'identifier des sous-groupes de réactions dans les modèles. Il est aussi possible de visualiser l'impact de l'ajout ou du retrait d'éléments sur le reste du graphe, permettant de tester des hypothèses. De nombreux logiciels existent pour la visualisation et l'analyse de la topologie de ces graphes. Cytoscape [230] est un des outils les plus utilisés et les plus complets avec plusieurs centaines de *plugins* permettant non seulement d'analyser des propriétés mathématiques comme l'analyse de centralité intermédiaire ou la recherche de cluster, mais aussi d'intégrer un très grand nombre d'annotations biologiques. Une liste de ces outils est disponible

ici¹. L'identification des relations entre les éléments d'un système est importante, elle permet de mettre en lumière des éléments clefs théoriques. Cependant, il manque une chose importante à ces modèles statiques qui caractérise le vivant : l'évolution au cours du temps.

2.1.2 Les modèles dynamiques : identifier les comportements

Un modèle dynamique consiste à étudier le comportement d'un système biologique en fonction du temps. Ces modèles peuvent se concentrer sur l'étude des évènements seulement et seront alors *qualitatifs* ou alors, ils peuvent aussi prendre en compte le nombre d'éléments de chaque espèce et sont alors *quantitatifs*. Le suivi de la quantité des espèces et la représentation du temps d'un modèle peut être faite de deux manières : *discrète* ou *continue*. Les modèles dynamiques peuvent être divisés en deux catégories : les modèles *déterministes* et *non-déterministes*. Il existe différents types de modèles non-déterministe, par exemple des modèles *stochastiques* ou bien, certains réseaux booléens non-déterministe.

Par définition, les modèles *déterministes* suivent une loi mathématique définie comme étant le comportement observé, autrement dit si on sait ce que l'on a en entrée de la réaction, on connaît la valeur exacte de ce que l'on aura en sortie [277]. Pour les modèles *stochastiques*, une notion de probabilité est introduite à cette loi, ce qui fait que la sortie obtenue appartient à une distribution de possibilités pour chaque entrée. Dans les modèles dynamiques, nous évoquerons aussi les modèles hybrides qui, comme leur nom l'indique, mixent les concepts, et les approches spatiales qui permettent de décrire les déplacements des éléments du système biologique.

2.1.2.1 Les modèles discrets.

Au sein de ces modèles, les échelles sont définies par des valeurs discrètes, par exemple, chaque espèce du système est quantifiée par son nombre d'individus, c'est le cas des modèles basés sur des agents (ABM). Il existe aussi des modèles dans lesquels les éléments du modèle sont définis par des valeurs booléennes 0 et 1 ou vrai et faux. Le temps, lui aussi, peut être discrétisé à une échelle propice à l'étude (p.e. minute, heure). Avec la complexification des modèles, il devient donc de plus en plus intéressant de travailler avec des modèles discrets qui permettent de réduire le coût de calcul en ne suivant qu'un

1. <https://bioinformaticsonline.com/pages/view/35386/list-of-visualization-tools-for-network-biology>

échantillonnage sur un temps lui-même échantillonné. Échantillonner le temps signifie définir un pas τ . Le système ne mettra à jour les quantités des éléments qu'après chaque τ et où la quantité d'un élément est fonction de sa quantité à $\tau - 1$.

La principale difficulté des approches discrètes repose sur la sélection de ce τ , si celui-ci est trop long, des informations seront perdues. C'est pour cela que certains modèles dynamiques discrétisent le temps en ne symbolisant que le temps où un événement se produit : systèmes à événements discrets.

2.1.2.2 Les modèles continus.

Au sein de ces modèles, le temps se déroule comme dans la réalité, sans discontinuité, et les espèces sont représentées le plus souvent par leur concentration. Les modèles quantitatifs, qu'ils soient discrets ou continus, sont fortement dépendants de la disponibilité des données quantitatives nécessaires au calibrage, et qui sont parfois difficiles à obtenir.

Le formalisme le plus utilisé pour les approches déterministes est celui des *équations différentielles ordinaires* (ODE) [129, 158]. Le suivi de la concentration de chaque élément du modèle se fait par une ODE qui lui est propre et qui est potentiellement influencée par d'autres éléments au cours du temps.

$$\frac{dx}{dt} = f(x(t)), \quad x(0) = x_0 \in D,$$

où f est la fonction appliquée à x pour modifier sa valeur et où f est défini par des constantes qui lui sont propres ou des variables liées aux autres éléments du modèle. Une variante des ODEs sont les PDEs (*équations aux dérivées partielles*) qui introduisent une notion spatiale en plus du suivi dans le temps.

2.1.2.3 Les modèles stochastiques : comprendre l'hétérogénéité du vivant.

Les modèles stochastiques sont des modèles dynamiques qui introduisent une composante aléatoire dans la résolution des événements. Ils peuvent être utilisés pour décrire l'hétérogénéité du vivant, par exemple les variations de comportements de cellules. Cette hétérogénéité peut être expliquée par trois facteurs : le facteur génétique, le facteur environnemental et le facteur chance [125]. Le premier correspond au génotype de la cellule, le second correspond à l'impact de phénomènes extérieurs (p.e. stress oxydatif, UV) et la dernière est liée à des mécanismes intrinsèques aléatoires de la cellule (p.e. mutations, probabilité de collision des particules). Les modèles déterministes sont capables de définir les

deux premières causes d'hétérogénéité, mais pas la dernière. Cependant, ces trois facteurs sont interdépendants, chacun pouvant influencer l'autre et définir des comportements qui ne seront observables que si ces trois facteurs sont définis.

Il n'est pas toujours pertinent d'utiliser des approches stochastiques pour modéliser le vivant. En effet, l'étude de très grandes populations (cellules, molécules) est très coûteuse d'un point de vue computationnel (temps de calcul, mémoire) et diminue l'intérêt de la stochasticité en réduisant la variabilité. La variabilité (ou "*bruit*") est définie par $\frac{1}{\sqrt{N}}$ où N est le nombre d'entités dans le système [246].

Un processus stochastique se définit par la capacité d'un système à se déplacer de manière aléatoire entre différents états x dans un espace donné M au cours du temps t . [122, 254]. Ce processus stochastique caractérisé par un système de transition peut être représenté sous forme d'un graphe de transition. Ce graphe $M = (V, E)$ est un graphe dirigé tel que V est l'ensemble des états possibles x et les arcs $(u, v) \in E$ modèlent les transitions possibles entre des états u et $v \in x$. Les états x et le temps t qui décrivent le système peuvent être discrets, continus ou les deux. Dans un système stochastique (discret), le prochain événement et le temps entre deux événements successifs sont tous deux définis par des distributions de probabilité. Sous hypothèse de la loi d'action de masse, chaque événement a une propensité et la probabilité que le prochain événement soit celui-ci est proportionnelle à sa propensité. Cette propensité se définit comme étant la probabilité qu'un événement s'applique sur un temps infinitésimal. Le temps entre deux événements est tiré selon une loi exponentielle dont le paramètre est égal à la somme des propensités de tous les événements potentiels dans l'état courant du système. Ainsi, plus il y a d'événements possibles ou plus leurs propensités sont élevées, plus le temps entre deux événements est court (en probabilité). Un processus X , i.e un ensemble d'événements $\{x(t)\}$, est définie par la totalité des suites d'événements qui peuvent se réaliser dans le système. À partir des définitions précédentes, le processus X peut être défini comme un processus de Markov. Ainsi, si le temps est discrétisé, le processus X peut être décrit comme une chaîne de Markov dont les transitions au cours du temps suivent une loi exponentielle [104]. Cette définition ne s'appliquait initialement que pour un gaz, puis il a été démontré qu'elle pouvait aussi s'appliquer aux liquides, ce qui important étant donné que la majeure partie des réactions biologiques ont lieu dans un milieu aqueux [100]. Les travaux de Gillespie fournissent une démonstration étapes par étapes de l'algorithme SSA [104].

2.1.2.4 Les modèles spatiaux : comprendre la dynamique spatio-temporelle

Les principales approches spatiales sont les modèles d'équation aux dérivées partielles (PDE), les modèles basés sur des agents (ABM) et les automates cellulaires. Les PDE sont une version spatiale des ODEs appliquées dans un environnement en deux dimensions, proposé par Euler [84]. Les approches ABM et d'automates cellulaires consistent à générer une grille en 2D ou 3D qui permet de décrire le déplacement et les interactions des espèces avec des coordonnées x, y . Pour les ABM, les agents représentant les espèces se déplacent sur cette grille, mais pour les automates cellulaires, les cases de la grille (appelées *cellules*) sont les espèces du modèle, possédant leurs propres caractéristiques. La valeur de ces *cellules* est impactée par leur entourage.

2.2 Exécution d'un modèle

Il existe différentes approches, définies par la représentation de l'échelle de temps entre discret et continu ou par le degré d'exactitude et d'abstraction mathématique (Figure 7). De la même manière que lorsqu'on fait un choix entre déterminisme et stochastique, il est important de bien choisir l'approche en fonction de la complexité des données et de choisir celle possédant le meilleur rapport précision/performance. On distingue deux types d'approches : les approches dites "exactes" et les approches dites "approchées" dont voici quelques exemples concernant les modèles stochastiques.

2.2.1 Les approches exactes

- Master équation : il s'agit de l'approche la plus précise et la plus complexe, détaillant l'état du système à chaque instant t . À chaque instant t , la probabilité de tous les états possibles du système, que le système change d'état ou reste dans le même état à $t + 1$, est défini sous forme d'un ensemble d'ODE correspondant à la Master équation [102, 171]. Cette équation permet d'avoir un résultat exact sur le comportement du système qui suit un processus stochastique, elle est cependant tellement coûteuse qu'elle n'est utilisée que sur des petits modèles principalement en chimie.
- SSA Gillespie (stochastic simulation algorithm) : il s'agit d'une approche basée sur les trajectoires, ce qui permet de ne pas calculer tous les états possibles, autrement dit à chaque *tour* est choisi un événement potentiel, expliquant pourquoi la sortie

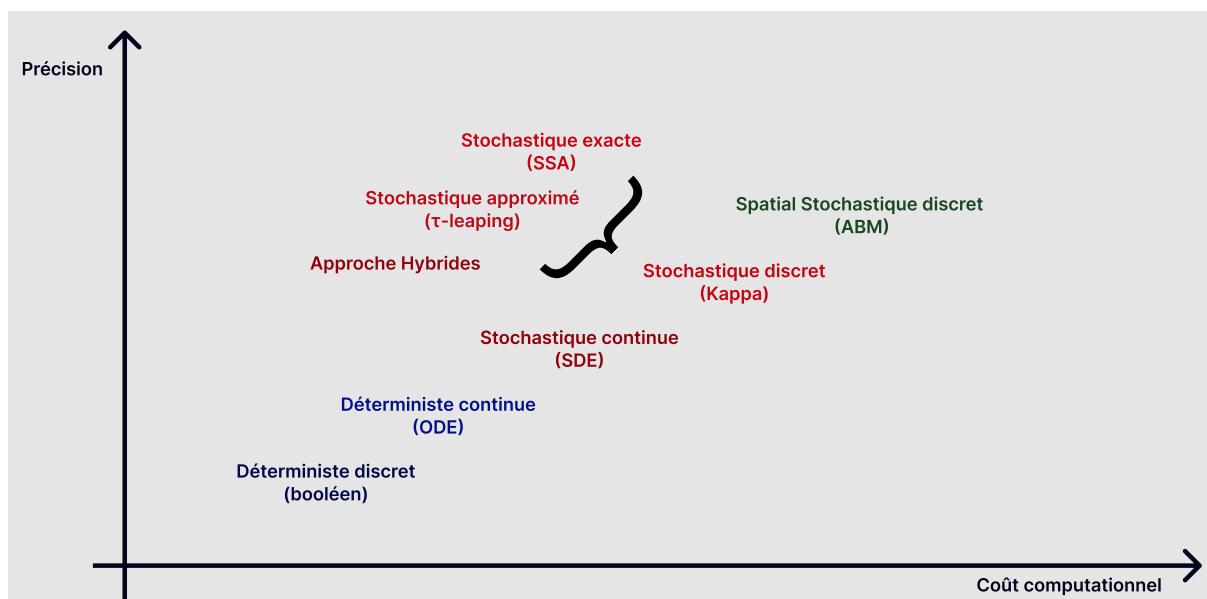


FIGURE 7 – Comparaison des différentes approches de modélisation en fonction de leur degré de précision par rapport à leur coût de calcul. Les approches déterministes sont moins précises mais aussi moins coûteuses en coût de calcul que les approches stochastiques plus précises. Inspiré de [246].

de chaque simulation est légèrement différente [104]. La réalisation de plusieurs simulations permet de générer plusieurs trajectoires correspondant à la distribution décrite par la Master équation. C'est donc une approche très précise et l'une des plus utilisées dans les modèles stochastiques. Elle a cependant un certain coût (quoique plus faible que celui de la Master équation), car elle doit simuler toutes les réactions, ce qui augmente la précision, mais représente un coût important en calcul et limite donc son utilisation aux modèles de taille modérée [246].

2.2.2 Les approches approchées

- τ leaping : il s'agit d'une variante de SSA où un saut est effectué (plus important que la discrétisation du temps) ce qui permet de calculer moins d'événements et donc de réduire le coup de calcul au détriment de la précision. La définition de τ est donc extrêmement importante, car il limite l'exactitude en lien avec une perte d'information durant le saut, notamment sur : le nombre de réactions, le timing des réactions et l'impact des réactions manquantes sur le reste du système pouvant changer les trajectoires [103].

- Langevin : elle s’applique dans le cas où le processus de Markov serait continu. La continuité du temps permet de poser l’équation de Langevin où la variabilité est définie par une variable Gaussienne qui est une variable aléatoire définie par sa moyenne et sa variance. L’équation est ensuite résolue par le mouvement Brownien (aussi appelé processus de Wiener) qui permet d’intégrer cette variable Gaussienne dans le temps et donc d’obtenir une distribution Gaussienne. Cette approche correspond à celle utilisée par les *équations différentielles stochastiques* SDE. Cependant, le passage à un temps continu implique une abstraction de la dynamique stochastique discrète, ce qui permet de faciliter le calcul au détriment de la précision [235].

N’ont été cités ici que les principaux algorithmes stochastiques utilisés, mais il en existe d’autres comme RSSA (regularized smoothing stochastic approximation) qui est une variante de l’algorithme SSA [250], les fragments stochastiques (Stochastic fragments) qui une méthode de réduction exacte de la sémantique stochastique en RBM [91] ou encore FRM (First and next reaction methods) [101] ; la revue de Simoni *et al.* propose une description plus précise des différents algorithmes stochastiques utilisés dans l’étude des systèmes biologiques [235].

Il est important de noter que ces approches de réduction exactes et approchées sont aussi appliquées aux modèles déterministes, comme les méthodes d’analyses tropicales [208] ou de réduction formelle des ODEs [89] pour en citer quelques-unes. Cette thèse ne portant pas sur ce sujet, nous n’entrerons pas dans le détail de ces approches, des revues spécialisées proposent des états de l’art exhaustif de ces méthodes de réduction pour les modèles déterministe [209, 166].

2.2.3 Les approches hybrides : le rapport efficacité/précision

Les modèles hybrides peuvent servir à modéliser des systèmes multi-échelles, un des enjeux est de trouver des méthodes de simulations efficaces. Cela peut passer par une résolution indépendante et successive des différentes échelles de temps (de la plus rapide à la plus lente). Une autre solution est de diviser le modèle en fonction des différentes échelles de temps, le but étant d’appliquer par exemple une approche stochastique exacte à l’échelle la plus lente et une approche stochastique approximée ou déterministe à l’échelle la plus rapide. L’échelle la plus rapide est le plus souvent la plus représentée avec un nombre important d’éléments et inversement pour l’échelle la plus lente [191, 116, 3]. La revue de Joseph Walpole *et al.* [269] détaille les différentes méthodes pour l’application de

modèles multi-échelles. Un algorithme hybride, HRSSA est proposé par Marchetti *et al.* en 2016, où les réactions rapides sont simulées en suivant des ODEs ce qui permet d'obtenir des résultats plus rapidement tout en conservant une certaine précision comparée aux approches complètement stochastiques ou déterministes [173].

2.3 Conception d'un modèle

Chaque modèle a pour objectif de répondre à une (ou plusieurs) questions, autrement dit le *quoi* biologique. Ce *quoi*, qui sera modélisé, permet de commencer à chercher les données et modèles déjà existant sur le sujet. Cette première étape permet d'étudier et de récolter les données disponibles. Celles constituent le *Qui* du modèle, elles sont à la fois la base et les limites de celui-ci. À partir des données, il est déjà possible définir si le modèle sera : qualitatif ou quantitatif et dans le dernier cas s'il sera discret ou continu. L'identification de ce *Qui* permet de définir le périmètre du modèle en termes de temporalité et d'espace, c'est-à-dire le *Quand et le Où*. Parfois, la définition du périmètre du modèle impliquera plusieurs échelles (temporelles et spatiales) et conduira au développement d'un modèle multi-échelles. Dans ce cas, le modèle pourra être implémenté de deux façons différentes : *top_down* ou *bottom_up*. L'approche *top_down* consiste à partir d'une description des éléments de l'échelle la plus grande (p.e. organe) et d'y ajouter des comportements des éléments des échelles plus faibles (p.e. cellules et molécules) qui pourrait expliquer son comportement. À l'inverse, l'approche *bottom_up* consiste à partir d'une description des éléments de l'échelle la plus faible et de les ajouter pour essayer de décrire un élément de plus grande échelle [180]. L'ensemble de ces informations permet de savoir si le modèle peut être déterministe ou stochastique et d'identifier de potentielles abstractions à faire durant l'implémentation du modèle.

2.3.1 Choisir l'abstraction

Un modèle est une représentation abstraite de la réalité biologique, tous les éléments impliqués dans le système biologique ne sont pas présents et certaines parties d'un modèle peuvent être réfutables (hypothèse) ou erronées (erreurs d'implémentation). Une abstraction peut être simplement liée à l'absence de connaissance ou bien liée à une différence de dynamique entre un ensemble de réactions par rapport au reste du modèle .

Il existe plusieurs méthodes d'abstraction, utilisant des algorithmes dédiés [135], des

simplifications mathématiques ou de modélisations [245] dont voici une liste non exhaustive [195] :

- enlever un ensemble de réactions ;
- prédéfinir des éléments (concentration, flux, réaction) ;
- abstraction mathématique (p.e, booléenne) ;
- former des groupes d'espèces (p.e, métabolites).

Pour citer George Box : "Tous les modèles sont faux, mais certains sont utiles" [35] cela signifie que l'abstraction n'est pas problématique en soi, pourvu qu'elle soit justifiée (basée sur la connaissance) et permette de décrire une dynamique.

2.3.2 Valider le modèle

Définir si un modèle est valide dépend avant tout de la question de départ, de l'approche utilisée et de l'objectif du modèle. Dans leur livre, Edda Klipp *et al.*[148] proposent quatre objectifs possibles de modélisation :

- description "précise" de chaque donnée par une approche mathématique ;
- description "prédictive" d'un processus biologique (généralisable à d'autres données) ;
- description "exacte" d'un processus biologique (abstraction du processus étudié) ;
- description "simple" des principes clefs (préséance du qualitatif sur le quantitatif).

En fonction des objectifs du modèle, l'exigence sur la validation ne sera pas la même, par exemple un modèle qui tend à décrire précisément un processus biologique doit refléter les données expérimentales, c'est-à-dire présenter une faible différence entre les résultats de simulations et les observations expérimentales. Un modèle s'intéressant aux principes clefs d'un processus, pourra être validé avec une différence plus importante relativement aux observations, pourvu que la tendance générale soit respectée.

Les méthodes de validation restent semblables quel que soit l'objectif de modélisation fixé. Seule varie l'exigence de conformité aux observations. Un moyen simple de valider un modèle est la validation externe [64]. Elle permet d'évaluer dans quelle mesure le modèle est généralisable en le testant avec des conditions autres que celles utilisées pour la calibration. Cela permet par ailleurs de mettre en lumière le risque d'*overfitting* (sur-adaptation du modèle aux données, le rendant non généralisable).

Il est donc important de définir les objectifs du modèle en prenant en compte la question biologique posée, les données disponibles et l'approche de modélisation. La sélection du modèle le plus adéquat se fait en comparant les prédictions obtenues avec les données

expérimentales. Si dans certains cas, il est simple d'observer un comportement aberrant et donc d'éliminer rapidement un modèle, il est parfois nécessaire de faire appel à des approches statistiques pour comparer les modèles avec les données expérimentales afin de sélectionner celui qui est le plus conforme [137].

2.3.3 Analyser la sensibilité et la robustesse du modèle

La sensibilité et la robustesse d'un modèle sont les deux faces d'une même pièce qui décrivent la capacité du modèle à être impacté par la variation des paramètres externes et internes. La sensibilité correspond à la capacité d'un modèle à répondre de manière adéquate à la modification d'un stimulus. À l'inverse, la robustesse est la capacité du modèle à rester stable en présence de petites variations. Il faut trouver un compromis entre ces deux aspects : un modèle trop sensible est instable, alors qu'un modèle trop robuste ne répondra pas de manière adéquate aux changements extérieurs.

L'analyse de sensibilité et de robustesse d'un modèle se fait généralement en modifiant la valeur des paramètres (internes ou externes) et en comparant les résultats obtenus. Cela permet d'identifier les paramètres auxquels le modèle est sensible et ceux qui ont peu d'impact sur le comportement de ce dernier.

Il existe plusieurs approches pour calculer un score de sensibilité/robustesse [234]. Ces approches peuvent être divisées en deux catégories selon qu'elles soient locales ou globales [289]. Les approches locales consistent à définir l'impact sur une partie du modèle d'un jeu de paramètres qui est modifié. Les approches globales consistent à quantifier l'effet général de la variation de paramètres d'entrée sur le modèle.

L'article de Marino *et al.* propose une méthodologie complète pour effectuer des analyses de sensibilité globales, en utilisant des outils analytiques existants, à la fois dans des contextes déterministes et stochastiques [175].

2.3.4 Choisir le formalisme en fonction des caractéristiques du sujet d'études

Une fois que l'on a répondu aux *Quoi*, *Qui*, *Quand*, *Où* qui ont permis d'identifier les contraintes du modèle, il faut définir le *Comment* qui correspond à la sélection du formalisme qui décrira la modèle.

Il existe divers formalismes adaptés pour chaque type de modèles. Comme décrit dans l'annexe A.1, on peut trouver, parmi eux, les modèles Booléen, les réseaux de Petri, les

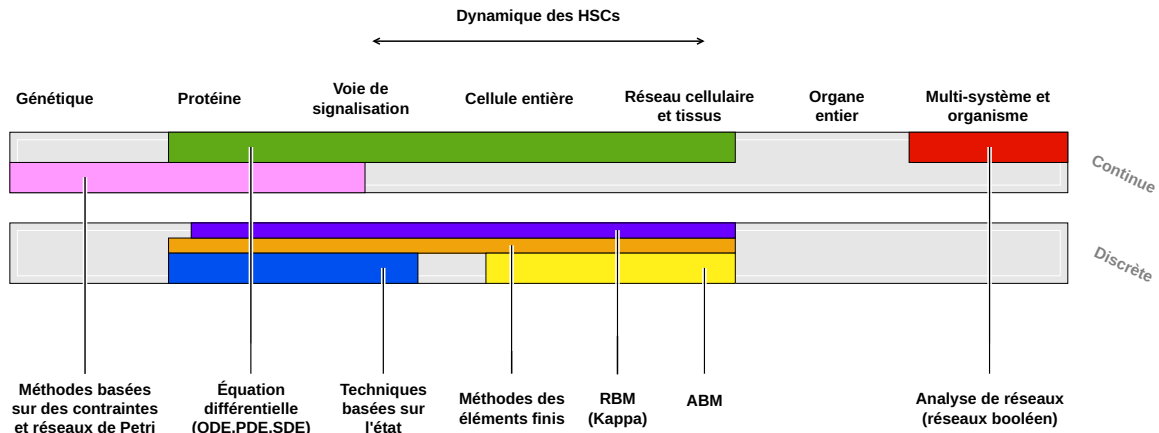


FIGURE 8 – Comparaison des différentes approches de modélisation en fonction de l'échelle en distinguant les approches discrètes et continues. Inspiré de [269].

équations différentielles ordinaires (ODE) et leurs variations (partielle [PDE] et stochastique [SDE]), les modèles basés sur les agents (ABM), les automates cellulaires et les langages basés sur des règles (RBM) ; pour ne citer qu'eux.

Chaque formalisme possède des spécificités, qui lui permette d'être plus ou moins adapté à certaines échelles spatiales et temporelles (voir Figure 8). L'implémentation de notre modèle numérique de la dynamique des CEF durant le développement et la réversion de la fibrose impose certaines contraintes de modélisation qui ont orienté notre choix de formalisme. Le Tableau 1 compare différents formalismes en fonction de ces contraintes. Le choix d'un formalisme se fait en sélectionnant celui répondant au mieux aux besoins et contraintes imposés par la question biologique posée.

Dans le cadre de nos travaux de thèse, nous ne décrivons que le langage Kappa issu du formalisme RBM et qui est l'approche correspondant au mieux à nos besoins qui seront décrits Section 4. Une description générale et non exhaustive pour chacune des autres approches fait l'objet de l'Annexe A.1 (adaptée de la revue de Machado *et al.* [170]).

Formalisme	Type Données	Nature des variables	Échelles	Nombre d'états	Déterminisme
Réseaux Booléen	Qual		Mono		D / S
Réseaux de Petri	Quan	Dis	Multi	1	D / S
ODEs	Quan	Cont	Multi	1	D
PDEs	Quan	Cont	Multi	1	D
SDEs	Quan	Cont	Multi	1	S
Automates cellulaires	Quan	Dis	Multi	Multi	D / S
ABM	Quan	Dis	Multi	Multi	D / S
RBM	Quan	Dis	Multi	Multi (locale)	D / S

TABLEAU 1 – Comparaison des différents formalismes en fonction des différentes caractéristiques.

Les caractéristiques sélectionnées sont celles que nous avons utilisées pour définir le formalisme le plus adapté pour répondre à notre problématique biologique. Le *Types de données* définit si le formalisme est qualitatif ou quantitatif; la *Nature des variables* définit si ces variables sont continues ou discrètes; l'*échelle* définit si le formalisme supporte les approches multi-échelles; le *Nombre d'états* définit si le formalisme permet de décrire plusieurs états pour une même espèce; le *Déterminisme* définit si le formalisme est déterministe et/ou stochastique. Qual : Qualitatif; Quan : Quantitatif; Dis : Discret; Cont : Continu; D : Déterministe; S : Stochastique.

2.4 Le langage Kappa

Le formalisme Kappa est un langage basé sur des règles de réécriture de graphes à sites dont nous rappelons le principe dans cette section. Nous détaillerons ensuite la syntaxe Kappa puis nous décrivons l'environnement Kappa avec la plateforme de simulation Kasim et les outils qui ont été développés. Nous illustrerons cette section par des exemples d'usage du langage Kappa dans des modèles issus de la littérature et nous terminerons en discutant des limites du langage Kappa.

2.4.1 Les modèles basés sur la réécriture de graphes à sites

Les modèles basés sur la réécriture de graphes à sites, ou basés sur des règles (RBM) sont centrés sur le comportement d'agents comme les modèles ABM et ont une syntaxe inspirée de la chimie. Les règles décrivent les processus biologiques en spécifiant les contraintes nécessaires à leurs applications. Les éléments à gauche des règles définissent les conditions d'application et les éléments à droite des règles sont la résultante de leurs applications. Il y a cependant une différence majeure avec les modèles ABM, il s'agit de descriptions locales. Pour les approches basées sur des agents, tous les comportements

possibles des agents doivent être décrits, qu'il s'agisse d'un objet [69] ou d'un processus [50]. L'énorme avantage des approches RBM est qu'elles ne nécessitent pas d'information sur l'ensemble des configurations des agents, seules les parties nécessaires pour rendre possible les interactions doivent être spécifiées comme le montre l'exemple Figure 9. Cela permet de faciliter grandement l'écriture du modèle en diminuant la quantité de règles nécessaires à sa mise en place et donc de gagner en temps et en lisibilité. Les modèles RBM sont donc plus appropriés pour la modélisation d'espèces complexe possédant de nombreux états et sites (p.e. cellules, protéines). Un état de l'art exhaustif des modèles RBM jusqu'en 2006 a été fait par Hlavacek *et al.* et de 2007 à 2013 par Chylek *et al.* [120, 49]. Une comparaison d'utilisation d'ODEs et de RBM pour l'analyse de la cascade de signalisation de DARP-32 montre l'avantage de la flexibilité des modèles RBM par rapport aux ODEs. Dans cet article est aussi fait un rapide état de l'art des modèles RBM depuis 2015 [282]. Les deux langages les plus utilisés en RBM sont Kappa [59, 58] et BNGL (BioNetGen Language) [115].

L'intérêt des approches RBM est que la formalisation mathématique est cachée à l'utilisateur. Il est possible d'écrire des règles. La définition mathématique du modèle est définie par la sémantique du langage. Le langage et le simulateur sont séparés, par exemple avec BNGL, il est possible de choisir une simulation déterministe avec des ODEs ou stochastique avec SSA. Il en est de même pour Kappa avec son simulateur stochastique KaSim et son générateur d'ODEs : KaDE [41]. L'intérêt des RBM repose donc sur leur sémantique qui permet un gain de temps et de compréhension dans l'élaboration d'un modèle. La description d'espèces complexes comme des cellules ou protéines possédant de nombreux sites et états différents se retrouve simplifiée avec l'approche RBM comparativement à une modélisation directe sous forme de ODEs ou par une approche ABM [61].

Kappa est un langage développé en 2004 par V. Danos *et al.* [59, 58] et constitue, avec BNGL, l'un des plus anciens langages basés sur les règles utilisés actuellement. Il s'agit d'un langage de réécriture de graphes [58, 86, 5, 136], qui par une syntaxe inspirée de la chimie permet de décrire de manière simple les interactions entre les éléments du modèle. Cette syntaxe est celle des langages RBM, ainsi chacun des éléments du modèle est défini comme un *agent*. Ces agents sont définis par des *sites* qui permettent aux agents d'interagir entre eux. Ces *sites* peuvent être liés, ou non, et porter des propriétés, appelées *états*.

Dans l'exemple Figure 10, le graphe décrit un complexe composé de quatre agents

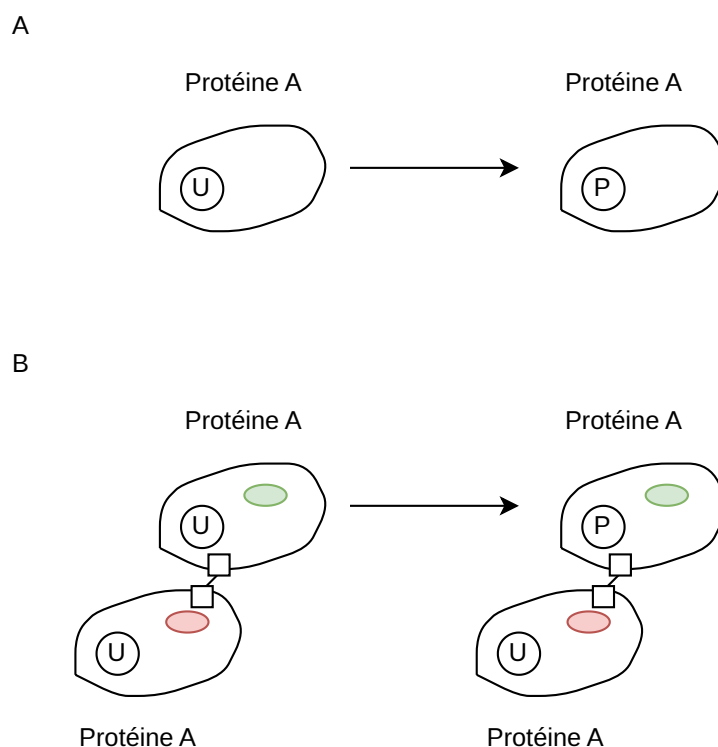


FIGURE 9 – Schéma de l’application d’une règle. **A)** est la règle, la valeur dans le cercle de la protéine A passe de U à P. **B)** Application de la règle sur une protéine A dont la valeur dans le cercle est U, quel que soit l’état des autres sites de A.

différents, A, B, C et D, qui interagissent via leur site respectif. Le langage Kappa permet au travers de ces *sites* de former des interactions entre agents qui peuvent ainsi former des complexes. Le langage Kappa considère l’état du système qu’il décrit comme un graphe dans lequel les agents (p.e., instances de protéines, instances de molécules) sont des nœuds et les interactions des arêtes. Ainsi, chaque règle va modifier une partie du graphe en créant ou en supprimant par exemples des arêtes ou des nœuds.

2.4.2 La syntaxe de Kappa

Pour expliquer la syntaxe de Kappa, nous utiliserons un modèle jeu correspondant à celui décrit sur le site web dédié au formalisme Kappa² et dans le manuel d’utilisation [31]. Ce modèle décrit les interactions entre trois agents (protéines) A, B, C correspondant à

2. Site web de Kappa

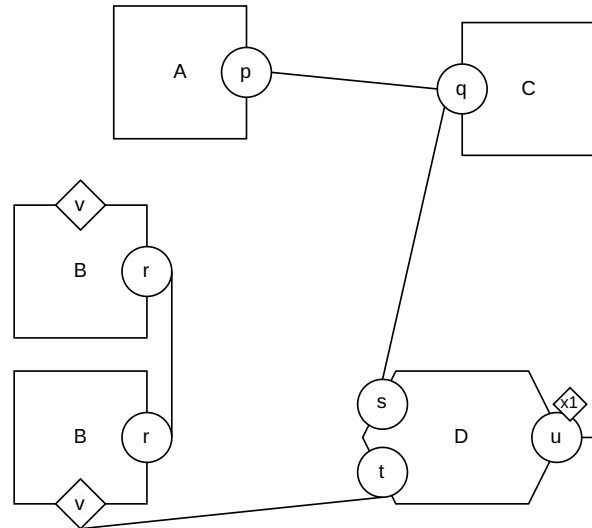


FIGURE 10 – Graphe de site. Un graphe de sites est constitué de nœuds qui indiquent les occurrences de certains composants, ici A , B , C et D . Ces composants sont associés à une liste de sites d'interaction, ici p , q , r , s , t , u et v . Les sites peuvent être libres ou liés par paire. Ces sites peuvent posséder des propriétés, plus communément appelé *état*, comme ici le site u de C qui possède un état $x1$.

un modèle biologique où les agents A (kinases) induisent la phosphorylation des agents C sur deux sites différents, ces réactions étant dépendantes des interactions avec les agents B . L'agent C non phosphorylé ne peut lier un agent A que dans un complexe avec B . Une fois phosphorylé, l'agent C peut lier un agent A , qui phosphoryle alors l'agent C sur un deuxième site. Nous allons détailler les différentes déclarations en syntaxe Kappa nécessaires pour élaborer ce modèle (voir encarts gris) ; ces déclarations concernent les agents, les variables, les observables et les règles dont sept décrivent les interactions entre agents et sont illustrées dans la Figure 11 et 5 concernent l'utilisation des *tokens* (jetons) et des *counters* (compteurs).

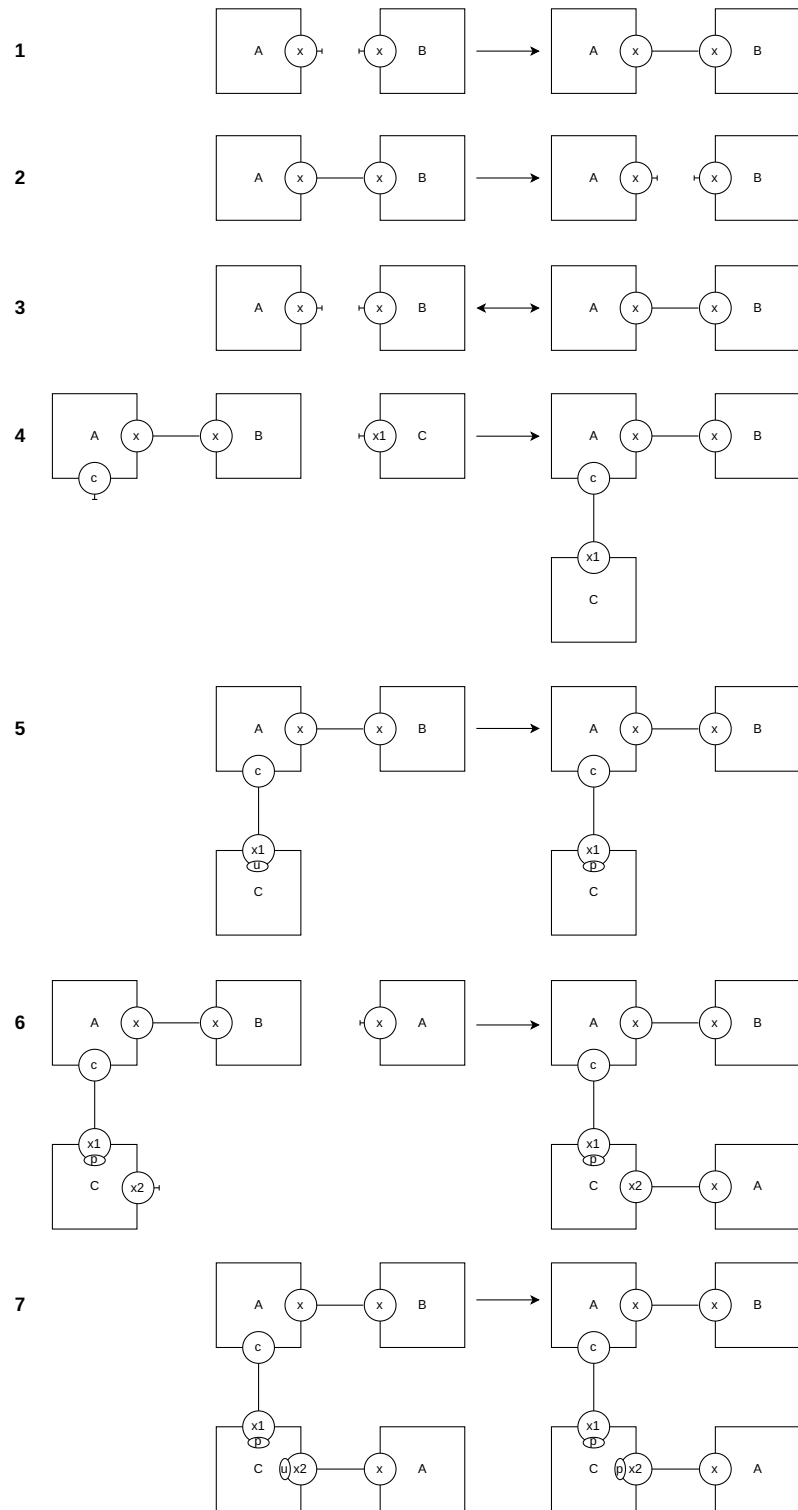


FIGURE 11 – Les 7 règles du modèle jeu décrivant la formation d'un complexe entre deux agents A et B , et la modification des sites d'un agent C . 1) A et B se lit sur leur site x ; 2) le complexe AB se dissocie; 3) Fusion des règles 1 et 2 décrivant la fixation de A et B ainsi que leur dissociation; 4) le complexe AB se lie à C sur le site c de A et le site $x1$ de C ; 5) le complexe ABC a le site $x1$ de C qui passe de u à p ; 6) un nouvel agent A se lie sur le site $x2$ de C ; 7) le complexe $ABCA$ a le site $x2$ de C qui passe de u à p .

La déclaration des Agents.

La syntaxe de Kappa commence par la déclaration des agents qui se fait de la manière suivant (Déclaration 1), ici trois agents A , B et C sont déclarés. L'agent A possède deux sites x et c , le site x pouvant se lier au site x des occurrences de B comme spécifié dans la déclaration de l'agent A (notez que cette déclaration des sites d'interaction entre deux agents n'est pas obligatoire). Les sites de l'agent C sont un peu différents, en Kappa l'état d'un *site* est définie entre accolades {}, ici les sites $x1$ et $x2$ peuvent avoir un état u ou p .

Déclaration 1

Déclaration des agents

```
%agent : A(x[x.B], c) // Déclaration de l'agent A
```

```
%agent : B(x[x.A]) // Déclaration de B
```

```
%agent : C(x1{u p},x2{u p}) // Déclaration de C avec deux sites modifiables
```

La déclaration de constantes et de variables.

Elle se fait par l'utilisation de `%var` : '*Nom variable*' *valeur*. Il est possible de définir des constantes (et variables) à partir d'autres constantes ou de concentrations (voir Déclaration 2), par exemple la variable `Concentration_C` qui est définie par la concentration de C . L'utilisation de `%obs` : permet de définir les observables à suivre durant la simulation, il peut s'agir d'agents ou de variables (voir Déclaration 3).

Déclaration 2

Déclaration des constantes et variables

```
%var : 'on_rate' 1.0E-3 // par molécule par seconde
```

```
%var : 'off_rate' 0.1 // par seconde
```

```
%var : 'mod_rate' 1 // par seconde
```

Déclaration 3**Déclaration des observables**

```

%obs : 'AB' |A(x[x.B])|
%obs : 'Cuu' |C(x1u,x2u)|
%obs : 'Cpu' |C(x1p,x2u)|
%obs : 'Cpp' |C(x1p,x2p)|
%var : 'n_ab' 1000
%obs : 'n_c' 0
%var : 'Concentration_C' |C()|

```

La déclaration des règles.

Les règles sont au cœur du modèle, elles décrivent les processus qui peuvent se réaliser en spécifiant les conditions dans lequel le graphe doit être pour que les processus se réalisent. Une règle fonctionne avec le principe *Si ... alors ...*, par exemple la première règle (Déclaration 4 règle 1) dit que *Si* les agents A et B sont libres sur leur site x (ce qui est spécifié par la notation $x[.]$) *alors* (symbolisé par \rightarrow) A et B peuvent se lier entre eux sur le site x . La seconde décrit la dissociation et présente une autre sémantique pour décrire la formation et la disparition de liaison entre agents (Déclaration 4 règle 2).

Le symbole $@$ permet de définir le taux d'une règle. Ce taux décrit la probabilité d'application de la règle pour chaque cas d'application potentiel dans un intervalle de temps infinitésimal. Dans le cas d'une réaction réversible, comme pour la liaison et la dissociation de A et B , il est possible de condenser la représentation des deux règles en utilisant la notation \Leftrightarrow et en donnant les deux taux d'applications correspondants (Déclaration 4 règle 3). Le langage Kappa permet de préciser quels sites de quels agents sont liés entre eux par l'utilisation d'indices entiers, dans notre exemple, il s'agit simplement de A et de B , mais parfois la formation de complexe nécessite plusieurs liaisons entre plusieurs agents et l'utilisation de ces indices permet de facilement lier les sites entre eux.

La règle suivante décrit la liaison entre le complexe AB et l'agent C . Pour cela, il suffit de préciser que l'agent A doit être lié sur son site x (ce qui est spécifié par la notation $x[.]$). Le modèle ne précisant qu'une seule interaction possible pour le site x de A cela correspond à définir la liaison du complexe AB avec C . En parallèle, l'agent C doit avoir son site $x1$ libre et dans l'état u pour pouvoir former la liaison. Cette liaison se fera sur le site c des agents A et C (Déclaration 4 règle 4). Une fois lié à C , le complexe AB va

pouvoir modifier l'état du site $x1$ pour le faire passer de u à p (Déclaration 4 règle 5). La règle 6 (Déclaration 4 règle 6) décrit la liaison entre un agent A , libre cette fois et un agent C sur le $x2$ de ce dernier, si le site $x1$ est dans un état p . La dernière règle, (Déclaration 4 règle 7), décrit la modification du site $x2$ d'un agent C , passant de u à p lorsque ce C est lié à un agent A . Ces règles décrivent donc la formation d'un complexe entre deux agents A et B qui iront phosphoryler le site d'un agent C , permettant à l'agent A de fixer un nouveau site de C et de la phosphoryler à son tour.

Déclaration 4

Déclaration des règles

Règle 1 : A se lie à B :

$$'a.b' A(x[.]), B(x[.]) \rightarrow A(x[1]), B(x[1]) @ 'on_rate'$$

Règle 2 : Dissociation de AB :

$$'a..b' A(x[1/.]), B(x[1/.]) @ 'off_rate'$$

Règle 3 : Association et dissociation de AB :

$$'a.b' A(x[.]), B(x[.]) \leftrightarrow A(x[1]), B(x[1]) @ 'on_rate' 'off_rate'$$

Règle 4 : AB se lie à C sur x1 :

$$'ab.c' A(x[_], c[.]), C(x1\{u\}[.]) \rightarrow A(x[_], c[2]), C(x1\{u\}[2]) @ 'on_rate'$$

Règle 5 : AB modifie x1 :

$$mod\ x1' C(x1\{u\}[1]), A(c[1]) \rightarrow C(x1[.]\{p\}), A(c[.]) @ 'mod_rate'$$

Règle 6 : A se lie à C sur x2 :

$$'a.c' A(x[.], c[.]), C(x1\{p\}[.], x2[.]\{u\}) \rightarrow A(x[.], c[1]), C(x1\{p\}[.], x2[1]u) @ 'on_rate'$$

Règle 7 : A modifie x2

$$'mod\ x2' A(x[.], c[1]), C(x1\{p\}, x2\{u\}[1]) \rightarrow \\ A(x[.], c[.]), C(x1\{p\}, x2\{p\}[.]) @ 'mod_rate'$$

La modulation de la simulation.

La dernière partie permet d'interagir directement avec la simulation, l'utilisation de `%init` : permet de définir les quantités d'espèces présentes au début de la simulation (voir Déclaration 5). Enfin, l'utilisation de `%mod` : permet d'appliquer des interventions durant la simulation. Par exemple, la première commande permet d'ajouter 10 000 agents *C* lorsque la simulation atteint le temps 10. La seconde quant à elle indique le moment d'arrêter la simulation.

Déclaration 5

Modulation de la simulation

```
%init : 'n_ab' A(), B()
%init : 'n_c' C()
%mod : alarm 10.0 do $ADD 10000 C();
%mod : alarm 100 do $STOP;
```

La déclaration des *Tokens* et règle associée.

Ce qui vient d'être rapidement présenté est l'utilisation *classique* de Kappa, il est possible de travailler avec des jetons appelés *tokens* qui fonctionnent comme des concentrations et qui peuvent être mis à jour durant l'application des règles. D'un point de vue purement informatique, le stockage en mémoire des agents et des *tokens* est différent. Chaque occurrence d'un agent sera stockée en mémoire alors que ce sera la valeur des *tokens*, indivisibles, qui sera stockée en mémoire. Par exemple, si un *tokens* a une valeur de 10 à l'instant t et de 10 000 à l'instant $t + n$, il sera stocké sur la même case mémoire, avec une simple mise à jour de sa valeur. Dans le cas d'un agent, si le nombre de ses occurrences passe de 10 à 10 000, le nombre d'éléments stockés passera lui aussi de 10 à 10 000.

Dans notre exemple, la phosphorylation de l'agent *C* nécessite des molécules d'adénosine triphosphate (ATP) (l'ATP est clivée, par hydrolyse, en adénosine diphosphate (ADP) et en phosphate). Nous déclarons les molécules d'ATP et ADP comme des *tokens*; $1 \cdot 10^{-6}$ d'ATP est consommée, et la même quantité d'ADP est produite. Pour modifier la valeur d'un *tokens* en fonction d'une règle, il faut insérer le symbole `|` après la déclaration des agents, et spécifier entre ce symbole et `@` comment modifier les *tokens*.

Déclaration 6**Déclaration des *tokens* et des règles associées**

%token : ATP

%token : ADP

%init : 9.6 ATP

%init : 0.55 ADP

Règle 8 : Déclaration d'une règle utilisant les *counters*

$$\begin{aligned} & \text{'hybrid rule a' } C(x1\{u\}[1]), A(c[1]) \rightarrow C(x1[.]\{p\}), A(c[.]) \\ & \quad | - 1E - 6 \text{ ATP } + 1E - 6 \text{ ADP } @ \text{'mod_rate'} \end{aligned}$$
La déclaration des *Counters* et règles associées.

La sémantique Kappa utilise aussi des *counters*. Implémentés en 2018, ils ajoutent un nouveau type d'*état* aux *sites* des agents [28]. Habituellement dans le langage Kappa, les *sites* des agents sont décrits par un *état* qui est défini par une chaîne de caractères (*u* et *p* dans notre exemple). Pour chaque agent, il est possible de déclarer un *counter* décrivant un *site* avec un *état* évoluant sur un intervalle discret (décrit par des entiers) et de préciser son intervalle (Déclaration 7). Nous pouvons définir une condition basée sur la valeur de ce *counter* dans les règles de trois manières différentes :

- en exigeant une valeur précise (Déclaration 7 règle 9) ;
- en exigeant une valeur supérieure (Déclaration 7 règle 10 et 11) ;
- en définissant la constante d'application de la règle selon la valeur du *counter* (Déclaration 7 règle 12)

La valeur de chaque *counter* peut être incrémentée ou décrémentée. Jusqu'ici, les règles Kappa étaient de la forme *Si j'ai ...alors, je peux obtenir ...*. Les *counters* permettent d'introduire des contraintes de la forme *Tant que je n'ai pas ...alors ...*, en incrémentant ou décrémentant la valeur du *counter* tant que la valeur souhaitée n'est pas atteinte. Sans les *counters*, les états intermédiaires d'un processus biologique complexe seraient déclarés individuellement sous forme d'*états* (chaîne de caractères) et une règle de type *Si j'ai 'chaîne de caractère' alors, je peux obtenir* devrait être définie pour chaque transition entre deux *états*. Avec les *counters*, les états intermédiaires peuvent être représentés par une liste d'entiers, dont le passage d'une valeur à une autre peut être défini par une seule règle de type *Tant que je n'ai pas 'une valeur seuil' alors, j'incrémente (ou décrémte)*.

Cela permet de réduire le nombre de règles proportionnellement aux nombres d'états intermédiaires. De plus, ces états intermédiaires peuvent avoir des dynamiques différentes, dans ce cas, les *counters* permettent de moduler le taux d'application de la règle selon la valeur des états intermédiaires.

Déclaration 7

Déclaration des *counters* et des règles associées

Déclaration :

$$\%agent : A(x, c, y\{= 2 / + = 6\})$$

Règle 9 : Test d'égalité :

$$A(x[.], y\{= 5\}), B(x[.]) \rightarrow A(x[1]), B(x[1]) @ 'on_rate'$$

Règle 10 : Test d'inégalité :

$$A(x[.], y\{>= 5\}), x(y[.]) \rightarrow A(x[1]), B(x[1]) @ 'on_rate'$$

Règle 11 : Test d'inégalité et modification du *counter* :

$$A(x[.], y\{>= 5\}), B(X[.]) \rightarrow A(x[1], y\{+ = 2\}), B(x[1]) @ 'on_rate'$$

Règle 12 : Règle dépendante de la valeur du *counter* :

$$A(x[.], y\{= var\}), B(x[.]) \rightarrow A(x[1]), B(x[1]) @ 'var' * 'on_rate'$$

Cette rapide présentation de la syntaxe de Kappa n'est certes pas exhaustive, mais peut être complétée par le guide d'utilisation de Kappa [31].

2.4.3 La plateforme de modélisation Kappa

2.4.3.1 Simulation : KaSim.

Comme expliqué précédemment, les langages RBM offrent une syntaxe qui peut s'interpréter dans différents formalismes mathématiques. La simulation stochastique est la plus souvent utilisée pour exécuter les modèles Kappa. Elle est rendue possible par le simulateur stochastique KaSim [63], soit en ligne de commande, soit via l'utilisation de l'interface utilisateur de Kappa [34]. Le simulateur KaSim [63], utilise l'algorithme de Gillespie [104]. Ainsi, l'ensemble des événements possibles est calculé dans chaque état atteint par le système. Chaque événement potentiel se caractérise par une occurrence de son membre gauche dans le graphe qui décrit l'état du système. La propension d'application

d'un tel événement est défini comme le taux de la règle qui a engendré cet événement, il s'agit de la probabilité qu'à l'évènement de survenir pendant un temps infinitésimal en suivant l'algorithme SSA de Gillespie (voir en Section 2.1.2.3). Il serait trop long et compliqué de recalculer la liste des événements potentiels à chaque changement de l'état du système. Cette liste peut être mise à jour dynamiquement en prenant en compte les événements qui, suite aux modifications apportées par le dernier évènement, ne sont plus applicables et ceux qui le sont devenus [63]. KaSim utilise les sous-motifs communs des motifs présents dans les membres gauches des règles pour faciliter la représentation de l'ensemble des événements et leur mise à jour [29].

2.4.3.2 Analyse statique : KaSa.

Une fois la simulation effectuée, il est important d'analyser celle-ci aussi bien en regardant les résultats bruts, mais aussi en comprenant comment les règles ont été appliquées et quelles ont été les relations entre les agents. Ainsi, l'analyseur statique KaSa [33], basé sur l'interprétation abstraite, permet de calculer des propriétés des modèles. Le but de cette analyse est de mieux comprendre les règles du modèle et donc d'avoir davantage confiance en leurs effets. Pour cela, il faut savoir si ce que l'on souhaite modéliser et ce que l'on a modélisé sont identiques et si ce n'est pas le cas, de trouver les erreurs dans la modélisation. En particulier, il permet de poser et de répondre aux questions suivantes :

- Existe-t-il une relation entre les états de différents sites dans les occurrences d'un agent ?
- Quand deux occurrences d'agents sont liées, existe-t-il une relation entre leurs états ?
- Un agent peut-il être doublement lié à un autre agent, est-ce qu'un agent peut être lié simultanément à plusieurs occurrences d'un même agent ?

La première question permet d'effectuer une analyse relationnelle classique. Cela permet par exemple de savoir si le site d'un agent ne peut être lié que si un de ses autres sites l'est déjà ou si le site d'un agent ne peut être lié que s'il est dans un état spécifique (phosphorylé par exemple). La seconde question est utile quand les états des sites permettent de représenter la localisation des agents. Il est alors attendu à ce que deux agents liés se trouvent toujours dans la même localisation (il est cependant tout à fait possible de considérer des liaisons transmembranaires). La dernière question concerne l'analyse de la formation de double liaisons.

2.4.3.3 Analyse de Causalité : KaStor.

Il est important de comprendre le comportement et les dynamiques des agents d'un modèle Kappa au cours du temps, autrement dit comment certains événements sont conditionnés par d'autres événements. L'analyse de Causalité avec l'outil Kastor répond à cette question. Dans un ensemble d'événements, certains peuvent être reliés par une relation de causalité, et former ce qui est appelée une *Trace Causale*. Une telle trace montre quels événements ont permis, suite à leurs applications, la réalisation des autres événements [60]. Dans l'exemple précédemment décrit, un agent C avait ses deux *sites* phosphorylés par un agent A et un complexe AB . Le processus décrit par exemple la fixation de A avec B pour former un complexe qui se fixera avec C pour induire la phosphorylation du site $x1$. La trace causale définit l'ensemble des événements indispensables à la réalisation d'un scénario donné, au cours de chaque simulation. Il existe de multiples traces qui sont obtenues à partir de la simulation en identifiant toutes les occurrences d'un événement d'intérêt. Dans notre exemple, la liaison de A et B est essentielle, car elle conditionne la fixation de C au complexe et la modification du premier *site* qui permettra la fixation d'un autre agent A qui modifiera le second *site* de C . Cela semble trivial sur un modèle de cette taille, mais lorsque celui-ci est composé de dizaines de règles, il est important de connaître la liste d'événements indispensables à la réalisation d'un processus, car elle n'est pas toujours connue au moment de l'implémentation du modèle. Les traces causales permettent d'avoir une vision des voies privilégiées pour l'acquisition d'une réponse, elle ne remplace pas les règles d'un modèle, car ces traces sont obtenues à partir de la simulation générée à partir de ces règles. Ainsi, il est possible de savoir quelles pré-conditions doivent être remplies pour qu'une règle soit appliquée, permettant de savoir si deux événements sont causalement liés ou non. Cet outil explique comment un événement précis a pu être atteint sans pour autant permettre de comprendre globalement comment le comportement collectif du modèle émerge des interactions. En revanche, il permet de résumer un grand nombre de traces classiques en une représentation synthétique.

2.4.3.4 Équations différentielles ordinaires et réduction de modèles : KaDe.

Simuler le modèle via l'algorithme SSA peut s'avérer trop coûteux et le passage à une autre approche mathématique est parfois nécessaire. Avec l'outil KaDe [41] il est possible d'exporter le modèle Kappa en ODE qui pourront être analysées avec les outils MAPLE [182], MATHEMATICA [280], MATLAB [128], ou OCTAVE [78]. Le modèle Kappa peut

aussi être exporté au format SBML [126] et au format DOTNET, celui-ci étant supporté par BioNetGen [85, 24].

Lors de la modélisation, il peut arriver que certains éléments du modèle aient un impact limité sur celui-ci. Conserver ces événements engendre donc un coût supplémentaire durant la simulation, sans fournir d'impact justifiant ce coût. Il est dans ce cas nécessaire de réduire le modèle, chose qu'il est possible de faire en Kappa. La réduction du modèle se fera par analyse du flot d'information entre les différents sites des composants biologiques [92] et par la détection des symétries [88, 25], ceci permettant d'identifier les variables qui fluctuent et celles qui restent stables au niveau de la structure du graphe.

2.4.4 Les exemples d'utilisation du formalisme Kappa dans la littérature

Les premières utilisations et les plus répandues du langage Kappa ont consisté à modéliser les voies de signalisations cellulaires impliquant de multiples interactions entre ligand, récepteur et cascade de phosphorylation intracellulaires. En 2007, Danos & al. [61] ont démontré pour la première fois l'applicabilité du langage Kappa pour décrire la signalisation cellulaire induite par l'interaction du facteur de croissance EGF avec son récepteur [61]. Suderman & al.[244] ont développé par la suite un modèle décrivant la voie de signalisation des kinases MAPK (Mitogen-activated protein kinases) induite par les phéromones chez la levure. En 2015, dans le cadre du projet DARPA Big Mechanism [53] qui avait pour but de générer automatiquement à partir de données biologiques des modèles causaux, Bachman et Gyori ont développé un modèle sur la signalisation cellulaire induite par la protéine protooncogénique Ras [109]. Au sein de ce même projet, Héctor F. Medina Abarca a modélisé la voie de signalisation induite par la protéine Wnt et a poursuivi ce travail durant sa thèse et post-doctorat [179].

Toujours dans les mécanismes de signalisation, Kohler *et al.* [150] ont développé un modèle sur la voie de réparation de l'ADN par excision de bases. À cette occasion, ils ont développé des règles vérifiant si des événements non souhaités apparaissaient et sur lesquels une analyse causale a été réalisée pour comprendre leurs origines et éviter toutes erreurs dans le modèle. Plus récemment, notre équipe a développé un modèle de régulation de l'activation du facteur de croissance *TGFB1* par le réseau d'interactions des composants de la matrice extracellulaire [251].

A noté qu'en 2010, Palominos *et al.*[199] ont développé plusieurs petits modèles (cycle

circadien des mammifères, transport Nucléocytoplasmique, dynamique prédateur-Proie, formation de lien Bi-trivalent pour intégrer des concepts d'espace). Les deux premiers modèles avaient pour objectif de travailler sur la compartimentation et le transport des agents, Le troisième abordait la notion de déplacement dans un espace restreint en 3D, le dernier modèle observait la notion d'encombrement induite par le cluster de récepteurs et son impact sur l'interaction récepteur-ligand (Ce phénomène a été également étudié en Kappa par Andrews *et al.* [6]).

Enfin, l'usage de la modélisation Kappa n'est pas restreinte aux interactions moléculaires et le langage Kappa a été très récemment utilisé dans d'autre champ de recherche ; comme l'économie ou pour mesurer l'impact épidémiologique du comportement d'individu durant une pandémie [266, 193]. À travers leur étude, Waites *et al.* [266] ont mesuré l'impact de différents facteurs sur la transmission de virus (port du masque, lavage de mains, dépistage, suivi des contacts entre individus, vecteurs de transmission).

Aujourd'hui, au sein de l'équipe de W. Fontana, l'utilité de Kappa dans la modélisation de larges modèles en prenant en compte des aspects thermodynamiques, tel que le non-équilibre ou les mécaniques statistiques, sont aussi étudiées. Il s'agit, en effet, de notions rarement introduites durant l'élaboration de modèles et qui permettent notamment d'étendre l'analyse de la Causalité aux systèmes chimiques. Par ailleurs, cette équipe s'intéresse à l'application de la modélisation Kappa pour des problématiques liées à l'évolution, comprendre comment les mutations définissent les phénotypes et comment celles-ci apparaissent aléatoirement au cours de l'évolution.

Les outils utilisant Kappa.

L'utilisation de Kappa et Kasim, s'est aussi imposé à travers son utilisation dans des outils plus larges pouvant aussi bien être utilisés pour des simulations que pour des analyses de causalité. Dans cette partie, nous verrons quelques outils qui ont incorporé Kappa dans leur pipeline.

- Atlas est un outil python de conversion en modèles dynamiques pouvant prendre en entrée des graphes de régulation génétiques ou encore des réseaux interactions moléculaires et des réseaux métaboliques [219]. L'objectif est de générer des modèles RBMs à partir de ces données. L'analyse de Causalité proposée par Atlas est réalisée à travers l'utilisation de KaSA.
- Sterope est un outil permettant le calcul de la sensibilité globale des paramètres de modèles basés sur des règles [199]. Cette analyse permet de quantifier l'impact des

incertitudes des paramètres fournis en entrée d'un modèle sur les sorties simulées. La mesure de la sensibilité globale est effectuée sur des modèles simulés avec KaSim. Cela leur permet d'obtenir le nombre et l'influence des règles du modèle pour estimer les interactions et les effets totaux de sensibilité.

- KAMI (Knowledge Aggregator and Model Instantiator) est un outil de Biocuration et de modélisation qui permet de générer des modèles RBM en Kappa à partir des données d'interactions moléculaires [159].

Il existe aussi d'autres outils qui intègrent différents formalismes pour proposer une syntaxe commune permettant comme PySB qui est un outil de programmation en python permettant de programmer un modèle dans différents formalismes et langages dont font partie Kappa et BNGL [165].

2.4.5 Les limites et extensions de Kappa

Actuellement, Kappa souffre encore de quelques limitations identifiées par la communauté, à savoir la déclaration des sites d'interactions, l'intégration des notions d'espace et certains manques de l'interface utilisateur.

2.4.5.1 Les sites d'interaction.

Dans la syntaxe Kappa, il existe encore des limitations. Ainsi, les sites d'interactions d'un agent ne peuvent porter deux fois le même nom et nécessitent de créer de nouveaux identifiants. Ce choix est cependant justifié, car il permet de réduire grandement le coût de recherche des occurrences possédant le motif décrit à gauche des règles [90]. Par ailleurs, W. Waites [265] et Urbanovský [259] considèrent l'obligation de spécifier les sites interagissant ensemble entre deux agents comme une limitation, car dans certains cas l'interaction en elle-même est plus importante que les sites concernés par celle-ci.

2.4.5.2 Les notions d'espace en Kappa.

Une des principales limitations du langage Kappa est l'absence de notion d'espace pour les agents d'un modèle, l'hypothèse étant que l'ensemble des agents est parfaitement distribué dans l'espace. Bien qu'il soit théoriquement possible d'implémenter cette notion de distribution spatiale des agents avec le langage Kappa, cela reste long et fastidieux. En effet, cela nécessite l'implémentation d'une grille représentant de manière discrète les coordonnées 3D des positions potentielles des agents. Le déplacement des agents à

travers cette grille se ferait *via* l'implémentation de très nombreuses règles modifiant ces coordonnées. Afin d'introduire cette notion de manière plus simple et facilitant la modélisation, des outils tels que Spatial Kappa [236] ou PISKaS [201] ont été développés. Chacune de ces extensions de Kappa, introduit la notion de compartiment et de transport, mais cette approche reste non-intuitive et complexe. Il est important de noter que ces deux outils sont développés par d'autres équipes que celles qui s'occupent du maintien et de l'amélioration de Kappa.

2.4.5.3 Les manques de l'interface utilisateur.

Dans son fonctionnement, le langage Kappa souffre aussi de limitation, notamment dans la gestion des annotations. Urbanovský [259] précise aussi que les annotations du script sont insuffisantes, alourdissant la lecture de celui-ci. W. Waites propose par ailleurs un outil KRDF [265] pour aider à l'annotation. Un autre outil, BSC (Biochemical Space), qui est un outil de notation semi-formelle, permet l'écriture et l'annotation d'un modèle dans une syntaxe simplifiée, inspirée de Kappa, qui sera ensuite traduite en Kappa pour être simulée [68]. Dans son article, J. Feret [90] relève que l'utilisation de réécriture de graphes à site induit des abstractions nécessaires pour définir les propriétés du comportement collectif des populations d'agents. Sans la topologie des liens, il est impossible de savoir le résultat d'une dissociation entre deux agents, car cela demande de connaître la distribution de ceux-ci de part et d'autres d'un lien.

APPORT DE LA BIOLOGIE COMPUTATIONNELLE DANS L'ÉTUDE DE LA FIBROSE ET SON DIAGNOSTIC

La « biologie computationnelle » ou « biologie numérique » se situe à l'interface des sciences biologiques, informatique et mathématiques, elle est par excellence une science interdisciplinaire. L'apport de la biologie computationnelle dans le domaine de l'hépatologie a considérablement fait évoluer la compréhension des pathologies hépatiques, que ce soit dans le traitement des données d'imagerie, cliniques, cellulaires et moléculaires [46, 9, 174, 20, 131] et plus récemment les approches liées à l'intelligence artificielle (IA) qui se sont développées en matière de diagnostic [190]. Dans le domaine de la modélisation à partir d'images, on peut citer la CFD (Computational fluid dynamic) qui utilise des données d'imagerie de patients (CT, MR ...) pour reconstruire la vascularisation et la perfusion du foie. Comprendre comment le flux est perturbé à différentes échelles (foie complet ou sinusoides) permet de comprendre l'architecture et la perfusion des tissus hépatiques, que ce soit pour des patients sains ou atteints de maladies du foie. Un état de l'art complet des nombreuses études de l'impact de la fibrose sur la vascularisation et la perfusion hépatique par des approches computationnelles est proposé dans la thèse de Charlotte Debbaut [67]. Dans ce chapitre, nous ferons un état de l'art des approches de modélisation multi-échelles liées à la fibrose hépatique et sa régulation, à la fibrose dans d'autres organes et nous évoquerons de façon non exhaustive les avancées en IA.

3.1 Modéliser la fibrose hépatique

Le développement de la fibrose est associé aux maladies chroniques du foie quelle que soit leur étiologie. Elle s'apparente à un défaut de cicatrisation en lien avec une inflammation chronique et conduisant à un dépôt excessif de matrice extracellulaire, principa-

lement constituée de collagène de type I. Les approches de modélisation dynamique de la fibrose hépatique restent limitées et consistent en modèles multi-échelles qui intègrent principalement les interactions entre les cellules produisant le collagène, les cellules de l'inflammation et les cytokines responsables de la communication entre les cellules. Ce sont des modèles de type PDEs ou basés sur des Agents. Ces deux types de modèles permettent l'intégration de déplacements spatiaux des éléments, utiles pour décrire, par exemple, le chimiotactisme cellulaire.

Le premier modèle ABM développé par Dutta-Moscato *et al* [77] est constitué d'agents représentant les cellules parenchymateuses (hépatocytes, vivants et morts), les cellules inflammatoires (cellules de Kupffer), les cellules productrices de collagène (CEF et fibroblastes portaux), et des éléments structuraux qui définissent les lobules (voir Figure 12). Les agents peuvent produire des facteurs diffusibles ($\text{TNF-}\alpha$, TGFB1, HMGB1) qui sont soumis à la dégradation et peuvent s'affecter mutuellement ou affecter d'autres agents. Les interactions entre agents sont gouvernées par quatre types de règles issues d'observations de la littérature; "transforms into", "releases", "recruits/activates", "inhibits". Contrairement à la littérature, le TGFB1 n'interagit pas avec les cellules étoilées, mais inhibe le $\text{TNF-}\alpha$ qui joue le rôle de cytokine pro-inflammatoire. Les paramètres de durée de vie et de taux de prolifération ont été calibrés de manière à ce que, en l'absence de perturbation externe, l'équilibre de base soit maintenu. Les données expérimentales sont obtenues dans un modèle de fibrose induite par le CCl4 chez la souris et la progression de la fibrose a été suivie par une coloration au rouge Sirius et une mesure de l'élasticité. Le modèle n'a pas été calibré sur une évolution spécifique des lésions du foie, même si les simulations correspondaient à la progression générale des changements histologiques induits par le CCL4.

Ce modèle a été par la suite modifié à deux reprises. Dans une première étude, Wang *et al.* [271] ont complété le modèle en introduisant le développement de la stéatose et l'accumulation de gouttelettes lipidiques selon le schéma proposé par Schleicher *et al.* [221]. L'idée étant que l'accumulation de lipides dans les hépatocytes conduit à une diminution du volume des sinusoides et une diminution de la circulation, affectant ainsi l'apport en oxygène et l'oxydation des lipides qui s'accumule en gouttelettes. La connexion entre ce petit réseau métabolique et le réseau de la fibrose induite par le CCl4 de Dutta-Moschetto se fait via le $\text{TNF-}\alpha$ comme le montre la Figure 13. La présence de gouttelettes lipidiques induit une augmentation du $\text{TNF-}\alpha$ qui est connu pour participer activement au métabolisme énergétique, en particulier à l'homéostasie lipidique. Cet ajout vise à

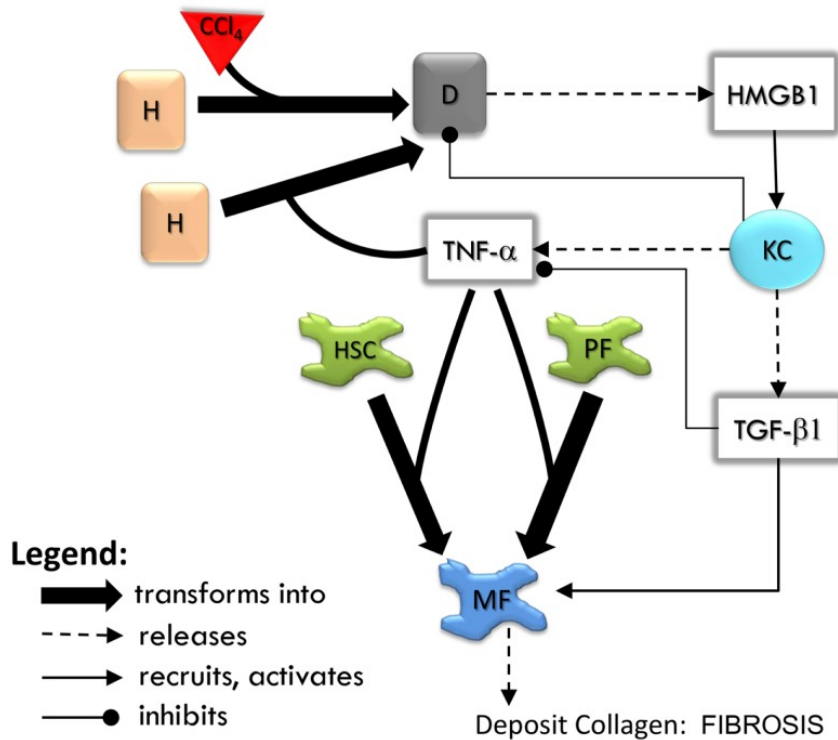


FIGURE 12 – Modèle de Dutta-Moschetto *et al* [77].

Représentation schématisée des interactions entre les agents dans le modèle.

H, hépatocyte ; D, hépatocyte mort ; KC, cellule de Kupffer ; HSC, cellule étoilée du foie ; PF, fibroblaste portal ; MF, myofibroblaste. Facteurs diffusibles $TNF-\alpha$, tumor necrosis factor alpha ; $TGF-\beta 1$, transforming growth factor beta 1 ; HMGB1, high mobility group box protein 1.

obtenir un modèle de développement de la fibrose en présence ou absence de stéatose. Les auteurs ont ensuite converti les composants (p.e. cellules) en grille de calcul (échelle 50–100 μm) pour faire des simulations d'ondes acoustiques avec pour objectif de prédire les mesures d'ondes expérimentales (c'est-à-dire la vitesse du son [SOS] et la vitesse de l'onde de cisaillement [SWS]) dans le contexte de fibrose avec ou sans stéatose. Si les résultats obtenus semblent être comparables avec les données de la littérature, les auteurs mentionnent l'aspect simpliste du modèle et la nécessité de le compléter.

Plus récemment, le modèle Dutta-Moscato *et al* [77] a été affiné en modifiant la description du processus inflammatoire ainsi que la diffusion des cellules dans l'espace [285]. Dans ce nouveau modèle, représenté Figure 14, les auteurs ont implémenté deux types de collagènes produits par les myofibroblastes et les fibroblastes portaux alors que dans le modèle de Dutta-Moscato, il n'y avait qu'un seul type de collagène. Les cellules étoilées

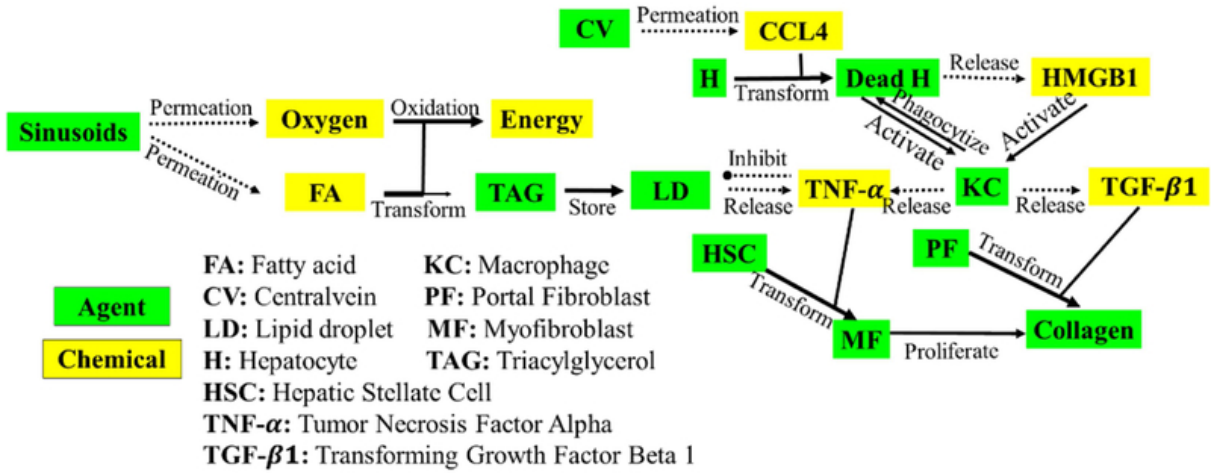


FIGURE 13 – Modèle de Wang *et al* [271].

Représentation schématique des interactions entre les agents du modèle de Dutta-Moschetto *et al* [77]) et les éléments moléculaires liés au développement de la stéatose.

et les myofibroblastes ont été paramétrés pour ne pas migrer de leur position initiale alors que dans le modèle de Dutta-Moscatto les cellules se déplaçaient de manière aléatoire en produisant du collagène lorsqu'elles atteignaient la zone périportale. Ce nouveau modèle reproduit les observations du modèle initial et les analyses de sensibilité ont identifié le seuil d'hépatocytes "morts" suite à l'inflammation et la balance entre le nombre de cellules de Kupffer et le nombre de CEF comme deux éléments critiques de l'initiation et de la progression de la fibrose hépatique.

En parallèle de ces approches ABM, l'équipe de Friedman a développé un modèle PDEs de fibrose rénale [113] qu'elle a adapté à la fibrose hépatique [94] en ajoutant les cellules étoilées et les lymphocytes T. Ce modèle est représenté par un système de 24 PDEs pour les concentrations cellulaires et les cytokines. Les cellules sont les cellules étoilées, les fibroblastes, les myofibroblastes, les macrophages M1 et M2, les cellules T (Th1 and Th2) et les cellules épithéliales (voir Figure 15). Ce modèle PDE permet de décrire l'accumulation de collagène en intégrant : le rôle des cellules de l'inflammation, du remodelage matriciel médié par celles-ci, l'accumulation de collagène sous forme de cicatrice ou de dépôt de fibrose. Les auteurs ont exploré l'efficacité de médicaments potentiels et ceux actuellement utilisés pour bloquer la progression de la fibrose. Ils ont par ailleurs quantifié la dépendance de la densité cicatricielle dans le foie en fonction des concentrations de deux biomarqueurs connus (TIMPs et l'acide hyaluronique) et ont montré que ces concentrations sont plus élevées dans le sérum des patients atteints de fibrose. Ce type de

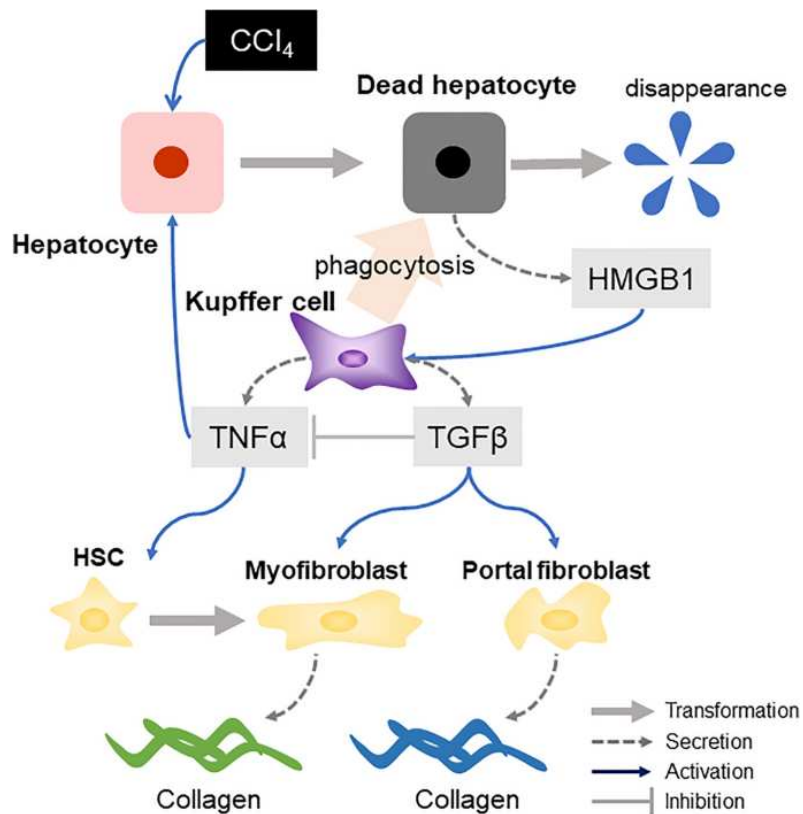


FIGURE 14 – Modèle de Yoshizawa *et al* [285]
Représentation schématique des interactions entre les agents, adapté du modèle de Dutta-Moschetto *et al* [77].

modèle reste très discutable dans le choix même de sa construction. Issue d'un modèle de fibrose rénale, il n'y a que très peu de paramètres hépatiques et très curieusement, il n'y a pas d'activation des cellules étoilées en myofibroblastes mais une induction de l'activation de fibroblastes par l'acide hyaluronique (HA) libéré par les cellules étoilées. De même, les aspects de remodelage, à savoir la production de MMPs et de TIMPs sont uniquement associés aux macrophages dans ce modèle alors que les cellules étoilées activées sont le moteur de ce remodelage au cours de la progression de la fibrose [23].

Ce modèle de fibrose hépatique a été ensuite modifié pour introduire l'infection par le virus de l'hépatite B comme cause de l'inflammation conduisant à la fibrose [95]. Pour cela, les auteurs ont complété le modèle en ajoutant des informations issues de la littérature, à savoir : des macrophages (M1 et M2), des cellules étoilées infectées par le virus B, la sécrétion de la Chemokine (C-C motif) ligand 3 (CCL3) par les cellules infectées qui va promouvoir l'attraction des macrophages et l'inflammation, et la réplication du virus

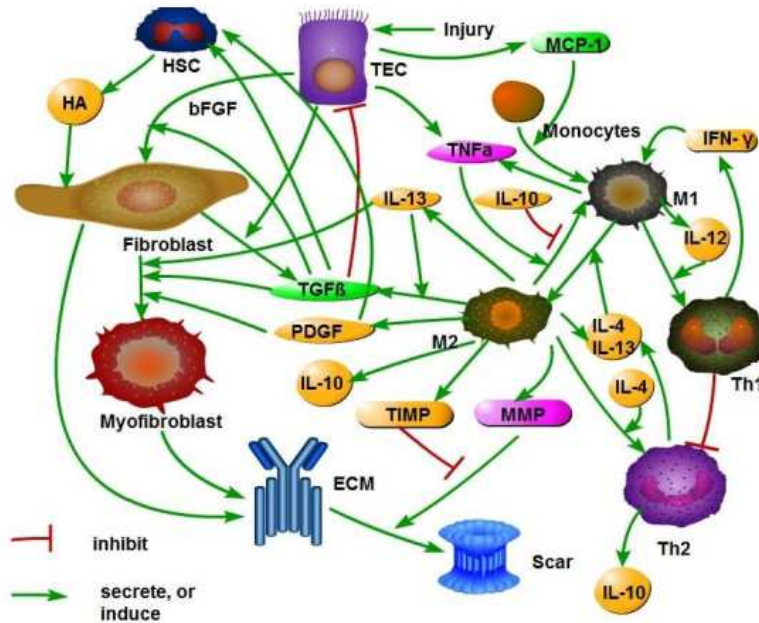


FIGURE 15 – Représentation schématique du modèle de fibrose hépatique développé par Friedman *et al* ([94].)

qui est bloquée par les cytokines ($\text{IFN-}\alpha$, $\text{TNF-}\alpha$, $\text{TGF}\beta 1$ and $\text{IFN-}\gamma$). Les simulations sont effectuées sur 180 jours et montrent bien une association entre l'infection et l'augmentation des paramètres liés à la fibrose. Le principal apport de ce modèle est l'étude de l'impact de deux traitements utilisés en clinique ($\text{IFN-}\alpha$ et adefovir) sur la fibrose, l'idée étant d'identifier les concentrations de ces deux traitements susceptibles de produire un effet maximal sur les paramètres de fibrose. La difficulté vient de la définition du score d'efficacité du traitement à partir des observables. Les auteurs concluent quant à l'effet plus bénéfique d'un traitement par un seul agent ($\text{IFN-}\alpha$ ou adefovir) plutôt que la combinaison de ces deux agents. Cependant, comme ils le soulignent eux-mêmes, leurs résultats sont en contradiction avec ceux d'une méta-analyse comparant l'efficacité de l'association $\text{IFN-}\alpha$ -adefovir par rapport à l' $\text{IFN-}\alpha$ en monothérapie [124] chez les patients traités et qui conclue sur la supériorité de la combinaison $\text{IFN-}\alpha$ plus adefovir.

3.2 Modéliser la fibrose dans d'autres organes

Parce que la fibrose est un processus lié à une défaillance de réparation/cicatrisation des tissus, elle se caractérise par des mécanismes communs à différents organes, à savoir

une réaction inflammatoire, une activation de cellules fibroblastiques et un remodelage du micro-environnement matriciel, l'association de ces composantes conduisant *de facto* au développement de modèles multi-échelles [4]. En ce sens, l'équipe de Avner Friedman a développé un modèle PDEs qu'elle a appliqué à la fibrose rénale [113], pulmonaire [112], pancréatique [114] et hépatique [94]. Tous ces modèles sont basés sur des interactions entre cellules (fibroblastes, cellules endothéliales, cellules de l'inflammation) et cytokines pour conduire à la cicatrisation ou la fibrose ; la différence, entre organe, résidant dans certain type de cellule et d'éléments diffusibles impliqués. S'inspirant des travaux de Avner Friedman, Adler *et al.* [2] ont proposé des modèles de circuits cellulaires (myofibroblastes/macrophages) conduisant à trois états : réparation, "cold fibrosis" (absence de macrophage) et "hot fibrosis". Si ces approches décrivent des mécanismes communs, ils ne prennent pas en compte les contraintes environnementales (moléculaires, cellulaires, physico-mécaniques etc) spécifiques à chaque organe et qui conditionnent différemment le développement de la fibrose et sa résolution.

Modèle de fibrose pulmonaire

Les approches de modélisation multi-échelles de la fibrose pulmonaire ont fait récemment l'objet d'une revue [160]. En lien avec les contraintes contextuelles évoquées ci-dessus, des modèles ont été développés, par exemple pour la fibrose pulmonaire idiopathique et la fibrose chez les patients tuberculeux, prenant en compte les granulomes¹. L'exposition à des particules (p.e. cigarette, polluants), qui est étroitement associée au développement de la fibrose pulmonaire, a été modélisée à partir des relations entre différents agents incluant les macrophages, les fibroblastes, une cytokine pro-inflammatoire (TNF- α), une cytokine anti-inflammatoire (TGFB1), le collagène et les "dégâts tissulaires". L'entrée de ce modèle (de type ABM) consiste en particules distribuées de manière aléatoires et qui vont activer les macrophages [36]. Le modèle prédit trois états : (1) une résolution, (2) une inflammation locale et une fibrose, (3) une fibrose et un dégât persistant des tissus. Le modèle est sensible à la durée et à la chronicité de l'exposition. Cependant, il aurait été intéressant d'avoir plus de conditions d'exposition aux particules pour identifier le seuil à partir duquel la fibrose se met en place et identifier la possibilité d'une réversion. Hao *et al.* [112] ont développé un modèle basé sur des équations aux dérivées partielles. Ce modèle plus complexe intègre quatre types cellulaires (cellule

1. granulome tuberculeux : Granulome spécifique constitué de lymphocytes et d'histiocytes d'un type spécial, les cellules épithélioïdes, plus ou moins groupées en follicules ou granulomes épithélioïdes [74]

épithéliale alvéolaire, macrophage, fibroblaste et MFB), six types moléculaires (TGFB1, $\text{TNF}\alpha$, PDGF, MCP-1, TIMP, MMP) et deux complexes (ECM, cicatrice). Les simulations du modèle sont validées par des observations chez 16 patients et par des données de la littérature, notamment sur les effets bloquants d'anticorps dirigés contre les cytokines. Cependant, le choix des molécules et des interactions décrites sont discutables, par exemple le choix de MMP et TIMP qui ne sont associés qu'aux macrophages alors que les fibroblastes sont aussi sources de ces molécules et agissent sur le remodelage.

Modèle de fibrose cardiaque

Les approches de modélisation ont largement contribué à mieux comprendre le rôle particulier du fibroblaste cardiaque dans la conduction électrique et le remodelage tissulaire au niveau cardiaque [286]. Dans le contexte d'une lésion cardiaque (infarctus, maladies infectieuses, diabètes), les cytokines inflammatoires conduisent à l'activation des fibroblastes en myofibroblastes produisant du collagène cicatriciel aboutissant à la réparation du tissu. S'il y a un déséquilibre dans le remodelage tissulaire, une fibrose interstitielle s'installe, même dans les zones éloignées de l'infarctus. Il n'existe que peu de modèles multi-échelles incluant l'ensemble des acteurs. En 2014, Rouillard et Holmes ont étendu leur précédent modèle ABM de la croissance et du remodelage de l'infarctus du myocarde [216] en l'associant à une méthode basée sur des éléments finis de la mécanique de l'infarctus. Cette approche a permis de mettre à jour de manière itérative à la fois les structures cicatricielles locales et les déformations au niveau des tissus [215]. Plus récemment, le modèle ABM initial du remodelage dans l'infarctus de [216] a été couplé à un modèle d'équations différentielles logiques (LDE) de la signalisation intracellulaire dans des fibroblastes cardiaques [286]. Comme décrit dans la Figure 16, ce nouveau modèle de réseau à grande échelle (91 nœuds, 142 réactions) est capable de prédire les modifications du tissu (fibrose) et des informations sur le micro-environnement qui régulent la signalisation intracellulaire des cellules [213]. Dans ce modèle hybride, l'ABM calcule la dynamique sur les niveaux de cytokines et de facteurs de croissance dans l'environnement, notamment TGFB1, $\text{TNF}\alpha$, IL-1b et IL-6, qui sont transmises au modèle LDE. Le modèle LDE calcule ensuite l'état de la signalisation intracellulaire des fibroblastes cardiaques dans le modèle ABM. En fonction de l'état du réseau, les fibroblastes font varier la sécrétion des cytokines, la production et la dégradation du collagène. Ce modèle est le premier modèle multi-échelles de la fibrose cardiaque, mais qui a vocation à être complété pour comprendre les phénomènes émergents qui résultent de l'interaction dynamique entre le

comportement des cellules, la signalisation moléculaire, la composition de l'ECM et la mécanique des tissus.

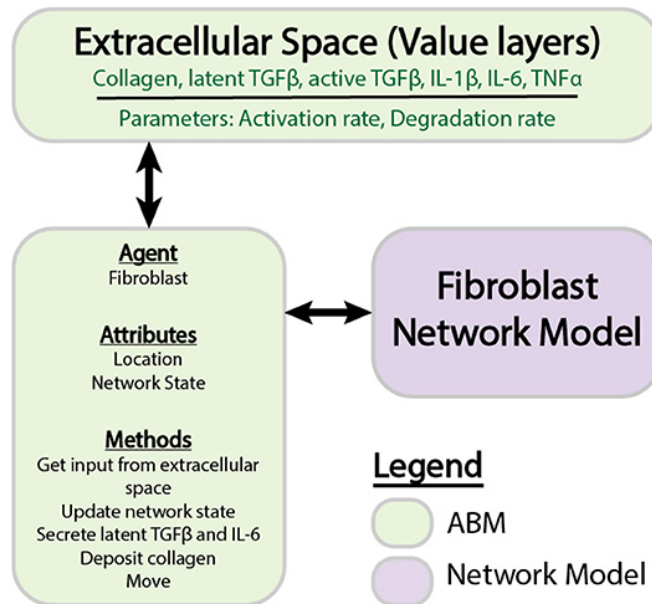


FIGURE 16 – Représentation schématique du modèle développé par Rikard *et al* [213]. Le modèle ABM est composée d'agents qui stockent des informations sur les attributs et exécutent des méthodes. Chaque agent stocke un état du réseau, qui est mis à jour par le modèle de réseau de fibroblastes.

Modèle de fibrose rénale

Le modèle de référence est celui de l'équipe de Avner Friedman, développé dans le cadre de la La néphrite lupique qui est associée au développement d'une fibrose interstitielle [113]. Ce modèle de type PDE inclut cinq types cellulaires (cellules épithéliales tubulaires (TEC), les fibroblastes, les myofibroblastes, les monocytes et les macrophages), cinq composants moléculaires diffusibles comme la protéine proinflammatoire chemoattractante MCP1 (monocyte chemoattractant protein 1), les facteurs de croissances TGFβ1 et PDGF, et les facteurs du remodelage matricielle, MMP (matrix metalloproteinase) et TIMP (tissue inhibitor of metalloproteinase) sécrétés par les macrophages activés. Le modèle contient aussi deux structures complexes (MEC et cicatrice). La dynamique d'interaction entre ces éléments est régie par des équations aux dérivées partielles et le modèle définit un espace autour de la zone lésée dans laquelle les cellules et les cytokines diffusent librement. Le modèle a été validé avec des données de patients (n=47) ayant des fibroses

de différents grades et pour lesquels des mesures des cytokines ont été faites dans les urines. Par conséquent, à partir d'échantillons urinaires, faciles à obtenir en routine, ce modèle permet de prédire l'avancée de la fibrose qui est très difficilement évaluable en clinique. Par ailleurs, ce modèle constitue un outil pour tester les effets de molécules thérapeutiques, cependant il ne prend pas en compte les modalités de délivrance, d'action et d'effet collatéraux de ces molécules. À noter que ce modèle a été repris pour développer des méthodes de calibrations de modèles biologiques [139]. Ces méthodes sont basées sur l'optimisation globale bayésienne (BGO), la sensibilité globale et l'analyse de l'environnement, et l'émulation du processus gaussien (GP). Le résultat de cette étude a permis de montrer que MCP1 et l'induction de production de TEC par le TGFB1 constitue les éléments clés dans la progression de la fibrose interstitielle. Plus récemment, Setten *et al.* [228] ont utilisé le modèle générique proposé par Adler *et al.* [2] pour montrer que le développement d'une fibrose irréversible est associé à des combinaisons spécifiques de processus métabolique et d'inflammatoire. Ce modèle mathématique de type ODE décrit l'interaction entre macrophages et myofibroblastes qui constitue un circuit Multi-Stable et reflète les effets fonctionnels de l'inflammation, de l'hypoxie et du contexte immunitaire adaptatif.

3.3 Modéliser le signal TGFB1, moteur de la fibrose

Le rôle du TGFB1 dans l'induction et la promotion de la fibrose n'est plus à démontrer [37, 72] et le signal TGFB1 constitue un véritable système complexe qu'il convient de présenter. Le TGFB1 est synthétisé sous la forme d'un précurseur homodimérique inactif, composé d'un propeptide auto-inhibiteur, de la protéine associée à la latence (LAP) et du peptide actif du TGFB1 lié de manière covalente. Le pro-TGFB1 est ensuite clivé intracellulairement par des enzymes de type furine pour générer le TGFB1 mature, qui reste associé de façon non covalente à la LAP sous la forme d'un petit complexe latent (SLC). le dimère est lié de façon covalente à une protéine de liaison du TGFB1 latent (LTBP) pour former le grand complexe latent (LLC) qui est séquestré dans la matrice extracellulaire (MEC). Le processus d'activation du TGFB1, schématisé sur la Figure 17, nécessite la dissociation du TGFB1 de la LAP liée à la MEC et implique de nombreuses molécules comme les intégrines et les protéases. Lorsqu'il est libéré, le TGFB1 se lie à des récepteurs spécifiques pour induire une variété de voies de signalisations intracellulaires. Les récepteurs du TGFB1 sont des sérines/thréonines kinases transmembranaires, qui

comprennent le type I (TGFB1) et de type II (TGFB2). La fixation du TGFB1 sur TGFB2 induit le recrutement du TGFB1 qui phosphoryle des protéines Smads. Ces Smads forment des complexes qui sont transportés vers le noyau, où elles activent la transcription des gènes cibles du TGFB1. À côté de cette voie dite canonique, le TGFB1 induit des voies non-Smad qui entraînent l'activation d'autres voies de signalisation comme les MAP kinases, GTPase Rho-like et des voies de la PI3K/AKT (Zhang 2017).

De très nombreux modèles ont été développés pour décrire la dynamique de signalisation du TGFB1 et ces modèles ont fait l'objet d'une revue récente réalisée par notre équipe [251] et dont la liste récapitulative est donnée Tableau 2.

Plusieurs modèles rendent compte du trafic des récepteurs qui contrôlent la réponse, transitoire ou permanente, dépendante du TGFB1 (Vilar *et al.* 2006b). Des modèles intégratifs ont ensuite couplé le trafic des récepteurs aux voies dépendantes des protéines Smad (Chung *et al.* 2009 ; Zi *et al.* 2011a ; Wegner *et al.* 2012, Nicklas et Saiz 2013 ; Shankaran et Wiley 2008). Les modèles décrivant le comportement de la voie canonique dépendante des Smad sont nombreuses [292, 291] et utilisent des réseaux de réactions chimiques (CRN) et des équations différentielles ordinaires (ODE). Elles incluent la dynamique de phosphorylation des Smad (Clarke *et al.* 2006b), du trafic des récepteurs (Vilar *et al.* 2006b) et des Smad (Melke *et al.* 2006b, Schmierer *et al.* 2008a) et de l'oligo-dimérisation des Smad (Nakabayashi et Sasaki 2009a). L'utilisation de ces modèles a permis de montrer la sensibilité de la réponse au TGFB1 i) à la quantité de celui-ci et à la durée d'exposition et, ii) au trafic d'internalisation et de recyclage du récepteur.

En tirant profit de ces modèles basés sur les ODE, notre équipe a développé ses propres modèles pour étudier le rôle du biomarqueur tumoral ADAM12 (Gruel *et al.* 2009) et du suppresseur de tumeur TIF1 γ (Andrieux *et al.* 2012), démontrant ainsi que des petits modèles différentiels numériques peuvent être de bons outils pour étudier le rôle de nouveaux composants régulateurs de la voie de signalisation canonique du TGFB1. L'explosion combinatoire des variables et des paramètres empêche l'utilisation d'approches ODE pour intégrer toutes les voies dépendantes du TGFB1, y compris la voie Smad, les voies non dépendantes de la voie Smad et le cross-talk avec d'autres voies. Pour dépasser cette limitation, notre équipe a utilisé un formalisme discret pour développer un modèle à grande échelle de signalisation dépendante du TGFB1 (Andrieux *et al.* 2014). Dans ce formalisme, les espèces et les complexes moléculaires sont représentés comme des variables booléennes (voir méthode BAN, Tableau 2) placées dans les nœuds d'un réseau et reliées par des "transitions gardées", c'est-à-dire des transformations mono-moléculaires qui ont

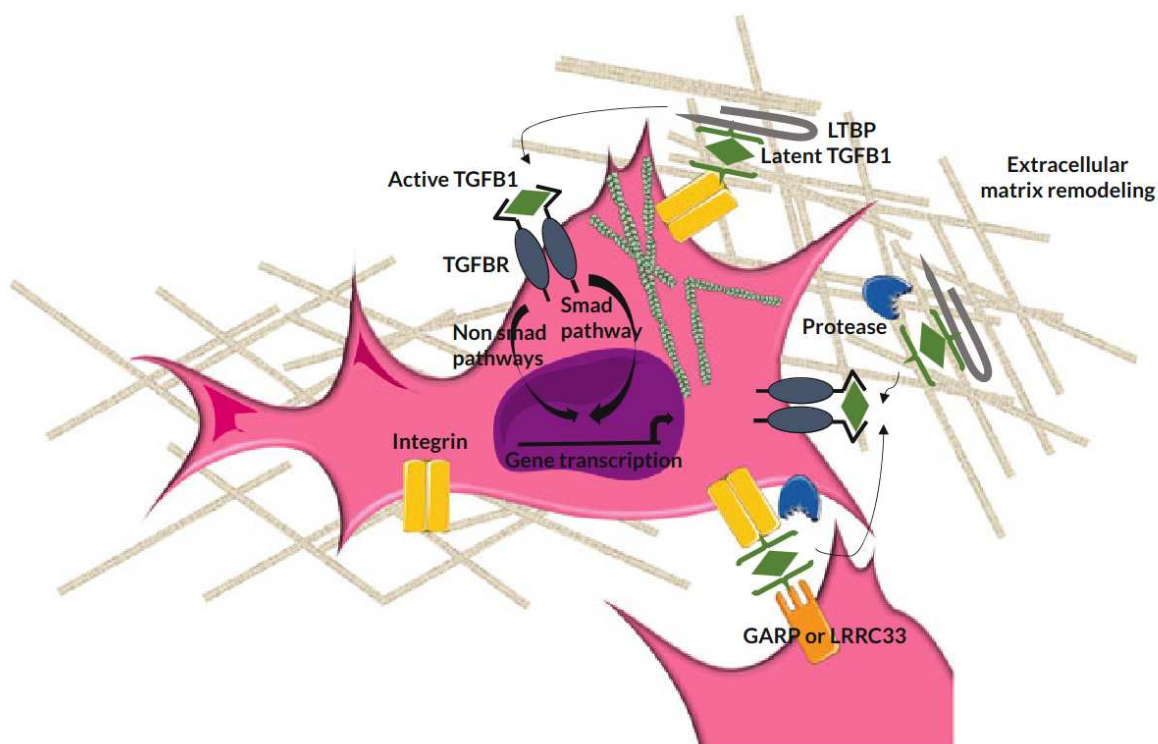


FIGURE 17 – Représentation schématique de l'activation du TGFB1 (Théret *et al*). La forme latente du TGFB1 (Latent TGFB1) est séquestrée dans la matrice extracellulaire (MEC) sous la forme d'un grand complexe latent associé à la LTBP qui se lie à la MEC. La libération du peptide actif du TGFB1 (losange vert) implique des mécanismes mécaniques et non mécaniques dépendant ou non des activités des protéases. Les intégrines (rectangle jaune) ou d'autres récepteurs de la surface cellulaire tels que GARP et LRRC33 lient le TGFB1 latent. Les contraintes de force entre les intégrines liées au cytosquelette et la MEC liée au LTBP induisent la libération du peptide TGFB1 actif. Les activités protéasiques sont impliquées dans l'activation du TGFB1 latent lié aux récepteurs de la surface cellulaire et contribuent à la libération du TGFB1 en agissant sur le remodelage de la matrice extracellulaire. Le TGFB1 actif interagit avec les récepteurs (TGFB1R) et l'activation des voies smad et non smad-dépendantes conduisant à la régulation transcriptionnelle des gènes cibles du TGFB1.

lieu si les conditions logiques sur les régulateurs et les événements définissant l'ordre de déclenchement sont satisfaites. Grâce aux événements, il est possible de simuler des dynamiques synchrones et asynchrones. Dans ce cas, une trajectoire du réseau représente une séquence de ces transitions. Sur la base de ce formalisme, un réseau TGFB1 composé de plus de 9 000 nœuds extraits de la base de donnée PID (Pathway Interaction Database)

Refs.	Pubmed/Biomodels Id	vars	Method	ECM
Andrieux <i>et al.</i> 2012 [8]	22461896	21	ODE	No
Andrieux <i>et al.</i> 2014 [7]	24618419	9000	BAN	Yes
Ascolani and Liò 2014 [12]	24586338	13	DDE	No
Cellière <i>et al.</i> 2011 [42]	22051045	18	ODE	No
Chung <i>et al.</i> 2009 [47]	19254534	17	ODE	No
Clarke <i>et al.</i> 2006[51]	17186703	10	ODE	No
Khatibi <i>et al.</i> 2017 [141]	28407804	11	DDE	No
Li <i>et al.</i> 2017 [161]	29322934	14	ODE	Yes
Lucarelli <i>et al.</i> 2018 [168]	29248373	13	ODE	No
Melke <i>et al.</i> 2006 [181]	17012329	17	ODE	No
Musters and van Riel 2004 [188]	17270884	5	PL	Yes
Nakabayashi and Sasaki 2009 [189]	19358856	7	ODE	No
Nicklas and Saiz 2013 [192]	23804438	38	ODE	No
Proctor and Gartland 2016 [205]	27379013	37	ODE	No
Schmierer <i>et al.</i> 2008 [222]	18443295	26	ODE	No
Shankaran and Wiley 2008 [229]	18780891	15	ODE	No
Steinway <i>et al.</i> 2014 [238]	25189528	65	BAN	No
Steinway <i>et al.</i> 2015 [237]	28725463	69	BAN	No
Strasen <i>et al.</i> 2018 [239]	29371237	26	ODE	No
Tortolina <i>et al.</i> 2015 [252]	25671297	460	ODE	NO
Venkatraman <i>et al.</i> 2012 [260]	23009856	13	ODE	Yes
Vilar <i>et al.</i> 2006 [262]	16446785	6	ODE	No
Vilar and Saiz 2011 [263]	22098729	12	ODE	No
Vizán <i>et al.</i> 2013 [264]	24327760	26	ODE	No
Wang <i>et al.</i> 2014 [270]	24901250	27	ODE	No
Warsinske <i>et al.</i> 2015 [272]	26384829	10	ODE	Yes
Wegner <i>et al.</i> 2012 [273]	22284904	53	ODE	No
Zhang <i>et al.</i> 2018 [287]	29872541	7	ODE	No
Zi and Klipp 2007 [290]	17895977	16	ODE	No
Zi <i>et al.</i> 2011 [293]	21613981	21	ODE	No

TABLEAU 2 – Modèles de signalisation du TGF β 1 [251]

ODE : équations différentielles ordinaires, DDE : équations différentielles à retardement, BAN : réseaux d'automates booléens, PL : hybrides , équations différentielles linéaires par morceaux ODE.

désormais disponible à l'adresse² a été générée et a permis d'explorer 15 934 trajectoires impliquant 145 gènes cibles du TGF β 1 (Andrieux *et al.* 2014 [55]).

2. <https://www.pathwaycommons.org>

Cependant, tous ces modèles ne prennent pas en compte les composantes extracellulaires qui régulent la biodisponibilité du TGFB1. La difficulté à ce stade est non seulement d'acquérir les données biologiques, mais aussi de trouver un langage adapté à la description de cette dynamique moléculaire extracellulaire. Pour répondre à ces contraintes, notre équipe a utilisé le formalisme Kappa pour intégrer les connaissances issues de la littérature décrivant les interactions entre protéines extracellulaires impliquées dans l'activation du TGFB1 (Coquet *et al.* 2017; Boutillier *et al.*, 2018). Les interactions font intervenir de multiples sites de fixation au sein de ces protéines, facilement modélisables par le langage Kappa (voir carte de contact Figure 18). Ce premier modèle contient une trentaine de molécules et environ 300 règles d'interaction qui sont librement accessibles sur le web (Théret *et al.*, 2020.³). Parmi ces molécules figurent des Adamalysines dont le rôle de régulateur du signal TGFB1 dans les maladies chroniques hépatiques avaient été identifiés par l'équipe (Atfi *et al.*, 2007 [14]; Bourd-Boitin *et al.*, 2011 [27]; Ruff *et al.*, 2015 [217]). La conception de ce modèle a permis d'une part d'organiser les données de la littérature grâce à des règles d'interaction et d'autre part de faire des hypothèses sur des inconnues.

La biodisponibilité du TGFB1 est intrinsèquement liée à la plasticité de la matrice extracellulaire et des modèles ont été développés pour étudier l'effet de l'élasticité du micro-environnement sur l'activation du TGFB1 à partir du LLC. En retour, le TGFB1 induit l'expression de composants matriciels par les cellules α -SMA positive (myofibroblastes) qui contribuent à cette rigidité et induit l'expression de LLC. L'ensemble d'équations cinétiques de Cockerill *et al.* ([52]) décrit cette boucle de rétroaction dynamique qui est capable d'interpréter la mesure de la rigidité du substrat et de la convertir en changements morphologiques de la cellule en réponse à un tel substrat. Ce modèle, bien que très abstrait au niveau de la description cellulaire, a permis de montrer la stabilité du système dans des environnements de rigidité variés. L'impact du TGFB1 sur le système est majeur et s'accélère avec l'augmentation de la rigidité qui induit une plus forte activation de LLC.

3. https://gitlab.inria.fr/DYLISS/kappa_TGFB

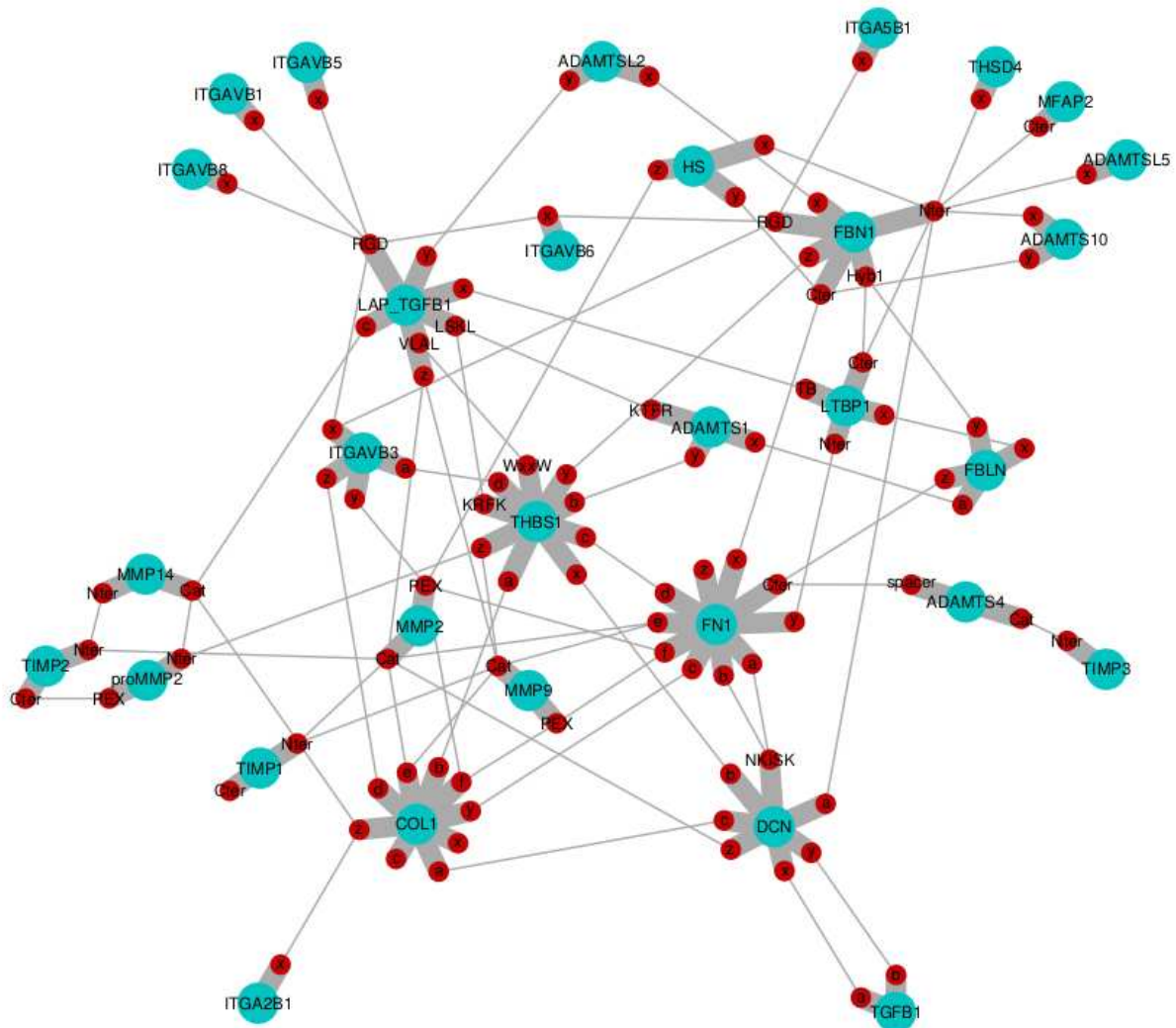


FIGURE 18 – Carte de contact du modèle Kappa d'activation du TGFβ1, Théret *et al* [251]). Les nœuds turquoise sont des protéines, les nœuds rouges sont les sites de fixation des protéines et les arêtes fines en gris décrivent les interactions protéiques.

HYPOTHÈSES DE TRAVAIL

Les maladies chroniques du foie sont associées au développement d'une fibrose caractérisée par des dépôts de matrice extracellulaire qui altèrent la fonction hépatique. Les CEFs sont les cellules à l'origine de la réparation des tissus et de la fibrose. Dans ce cadre, il est nécessaire de comprendre leur dynamique pour développer de nouvelles approches thérapeutiques. Il existe peu de modèles dynamiques de la fibrose hépatique [94, 95, 77, 271, 285]. Ces modèles multi-échelles intègrent principalement les interactions entre les cellules produisant le collagène, les cellules de l'inflammation et les cytokines. Ces modèles contiennent peu d'information concernant la dynamique des CEFs. Notre objectif est d'étudier plus spécifiquement la dynamique des CEFs au cours du développement et de la résolution de la fibrose. Cela implique l'utilisation d'un formalisme pouvant traiter des données quantitatives discrètes, et pouvant gérer une approche multi-échelles ainsi que des interactions multi-états. Il s'agit de décrire la dynamique d'une population de cellules possédant différents phénotypes et interagissant avec des éléments moléculaires. Le langage Kappa, issu du formalisme RBM, est particulièrement bien adapté à ces besoins (Tableau 1). En plus des CEFs, le TGFB1 est aussi connu comme étant un acteur clef du développement de la fibrose. Un modèle en langage Kappa décrivant la dynamique d'activation du TGFB1 dans la matrice extracellulaire a été développé par notre équipe [251]. Cependant, en raison de l'absence des cellules permettant la production de TGFB1 (i.e. les CEFs), ce modèle ne permet pas d'appréhender la dynamique d'activation du TGFB1 durant le développement de la fibrose.

Dans ce contexte, le développement d'un modèle de la dynamique des CEFs en langage Kappa a été initié par Octave Hazard (Stage de recherche, Institut Polytechnique de Paris, avril-août 2020) et a permis de mener une première réflexion sur l'élaboration de règles décrivant l'activation des cellules étoilées par le TGFB1. Le projet de ma thèse consistait d'une part à reprendre et développer ce premier modèle pour démontrer la possibilité d'utilisation du langage Kappa dans l'élaboration de modèles multi-échelles et d'autre part de modéliser la dynamique des CEFs durant le développement et la réversion de

la fibrose sous l'influence du TGFB1. Un objectif à plus long terme visait à fusionner les modèles d'activation du TGFB1 et de la dynamique des CEFs pour caractériser la régulation de l'activation du TGFB1 au cours de la fibrose afin d'identifier de potentielles cibles thérapeutiques.

Ces travaux ont conduit à l'implémentation d'un premier modèle dans lequel chaque espèce est définie par un agent et dont chaque occurrence correspond à un élément biologique. Autrement dit, un agent TGFB1 représente une protéine et un agent HSC représente une CEF. Ce modèle s'est révélé inutilisable, dû au stockage en mémoire d'un grand nombre de protéines par cellule (10 000 par agent cellulaire). Chaque simulation entraînait une saturation de la mémoire (Dell Inc. Latitude 7410 ; 8 x 1.8 GHz ; 16 GB RAM). Afin de surmonter cette limitation mémoire, nous avons choisi de considérer un agent TGFB1 comme un groupe de 10 000 protéines (correspondant au nombre de récepteurs TGFBR). Au final, nous avons développé un modèle contenant 52 règles, 33 paramètres et les abstractions choisies pour le TGFB1 nous a permis d'obtenir des premiers résultats présentés dans le chapitre de livre 5.

Cependant, considérer qu'un agent TGFB1 correspond à un ensemble indivisible de protéines rend impossible toute fusion avec le modèle de la dynamique d'activation du TGFB1, où est décrit le comportement individuel des protéines. Nous avons donc introduit des tokens, qui permettent de conserver cette notion de groupe de protéine pour interagir avec les cellules. Ces *tokens* permettent de décrire cet ensemble comme divisible, permettant aussi les interactions protéine-protéine. L'utilisation d'un seul agent cellulaire au lieu de deux, nous a permis par la suite de créer un historique d'activation et de différenciation des cellules. L'utilisation de counters a permis, en parallèle, de simplifier la représentation multi-états des CEFs et MFBs, réduisant ainsi le nombre de règles nécessaires pour décrire leur dynamique d'activation et différenciation (réactivation et re-différenciation) mais aussi de production de COL1.

Les simulations de ce premier modèle, bien qu'encourageantes (section 5.2), ne correspondaient cependant pas aux observations expérimentales décrites dans la littérature. Nous avons donc exploré un certain nombre de paramètres pouvant influencer notre modèle. Parmi ceux-ci :

- **Le nombre de cellules initiales.** Pour cela, nous avons calculé à partir des données de la littérature sur les volumes des espaces de Disse et des CEF, le nombre de cellules par espace de Disse et nous avons choisi de fixer le nombre de cellules à 10 000.

-
- **Le rôle de la matrice extracellulaire sur l'activation des CEFs.** Une donnée importante dans la régulation de la dynamique des CEFs vient de l'impact de la matrice sur l'activation des cellules. Compte tenu du manque de données quantitatives sur les contraintes mécaniques de la MEC, nous avons concentré nos efforts sur le suivi de la dynamique du COL1 (synthèse, dégradation) en lien avec les dynamiques des CEFs durant le développement et la réversion de la fibrose.
 - **La dynamique d'élimination des MFBs.** Nous avons aussi fait un travail de recherche sur la dynamique d'élimination des MFBs durant la réversion de la fibrose en nous intéressant aux trois voies connues : la sénescence, l'apoptose et l'inactivation. Le manque de données quantitatives (p.e. vitesse de réaction) sur la dynamique d'entrée en sénescence des CEFs dans la littérature nous a obligé à considérer l'apoptose et la sénescence comme une seule voie d'élimination dans notre modèle. Pour l'inactivation, les travaux de Kisseleva et al. [143] ont apporté beaucoup d'information sur cette dynamique tout en laissant des inconnues quant au devenir des cellules inactivées et des cellules réactivées. Nous avons donc émis deux hypothèses sur le devenir de ces cellules : (i) les cellules inactivées reversent vers un état quiescent, et (ii) les cellules réactivées peuvent être de nouveau inactivées.
 - **Le turn over du récepteur au TGFB1.** En parallèle, nous avons essayé de décrire plus en détails la dynamique du récepteur au TGFB1 par l'utilisation de counters. Cela nous a permis d'individualiser les récepteurs et d'en décrire 10 pour chaque cellule, ce qui a en conséquence permis de faire interagir des groupes plus réduits de protéines avec les cellules (1 000 par récepteur au lieu de 10 000). Malheureusement, l'augmentation exponentielle du coût de calcul (504,05 s pour 10 récepteurs, ~ 36 h pour 100 récepteurs; Dell Inc. Latitude 7410; 8 * 1.8 GHz; 16 GB RAM) nous a contraint à mettre de côté l'individualisation des récepteurs en attendant une meilleure implémentation des counters dans le langage Kappa.

L'évolution du premier modèle décrit dans le chapitre de livre vers le modèle final s'est traduit par :

- le passage du TGFB1 d'agent à token
- le passage de deux agents cellulaires à un seul
- le passage de site à counter pour représenter les états d'activation et différenciation des agents cellulaires
- l'ajout de la dynamique de COL1 représenté sous forme de 2 tokens
- l'ajout des hypothèses de travail sur le devenir des cellules inactivées et des cellules réactivées
- la re-calibration complète des paramètres du modèle par confrontation d'une centaine de simulations aux données expérimentales

Les contributions de nos travaux de thèse sont détaillées dans le chapitre II RÉSULTATS et se divisent en trois parties :

- Dans une première partie, nous avons démontré l'applicabilité du langage Kappa à la modélisation multi-échelles en utilisant la dynamique des CEFs comme modèle d'étude. Ce travail a fait l'objet d'un chapitre de livre [26].
- Dans une deuxième partie, nous avons introduit l'utilisation de *tokens* et de *counters* pour proposer une série de modèles permettant de réaliser des simulations. Cela nous a permis de maîtriser l'explosion du coup mémoire lié à la déclaration des molécules et des cellules en tant qu'agents.
- Dans la troisième partie de la thèse, nous avons d'une part validé nos modèles *in silico* par de nouvelles expériences biologiques dans un modèle expérimental de fibrose induite par le CCl4 chez la souris. D'autre part, nous avons validé les prédictions du modèle en les confrontant à des données chez des patients présentant une fibrose. Ces travaux font l'objet d'un article soumis pour publication.

DEUXIÈME PARTIE

Résultats

LE LANGAGE KAPPA POUR LA MODÉLISATION MULTI-ÉCHELLE, APPLICATION À LA DYNAMIQUE DES CELLULES ÉTOILÉES

La première contribution de nos travaux de thèse a été de démontrer l'applicabilité du langage Kappa à la modélisation de systèmes multi-échelles. Ces travaux ont fait l'objet d'un chapitre de livre intitulé "The Rule-based model approach. A Kappa model for hepatic stellate cells activation by TGF β 1." et qui a été publié dans *Systems Biology Modelling and Analysis : Formal Bioinformatics Methods and Tools. E. De Maria (Ed.)*. Wiley.© , 2022.

De ce chapitre, les parties redondantes avec l'introduction de cette thèse ont été retirées. Ainsi, vous est présenté la construction du modèle, les résultats fournis par celui-ci, au travers d'une étude statique et de simulations, ainsi que les pistes de réflexions apportées par ce modèle aussi bien sur le langage Kappa, mais aussi sur la modélisation de la dynamique des cellules étoilées du foie.

Dans un souci de respect aux contributions apportées par Nathalie Théret et Jérôme Feret, j'ai choisi de conserver le texte en anglais afin de respecter au mieux leurs propres termes. Mes contributions dans l'élaboration de ce modèle sont l'amélioration du modèle d'Octave Hazard par l'ajout de règles et l'apport de certaines corrections. Mon travail a aussi permis de pointer les limites de Kappa lorsqu'il s'agit de représenter, sous forme d'agent, la plus petite échelle d'un modèle multi-échelle (ici le TGF β 1).

5.1 Model of activation of stellate cells

Now we describe in Kappa a model of the behavior of a population of hepatic stellate cells.

5.1.1 Overview of model

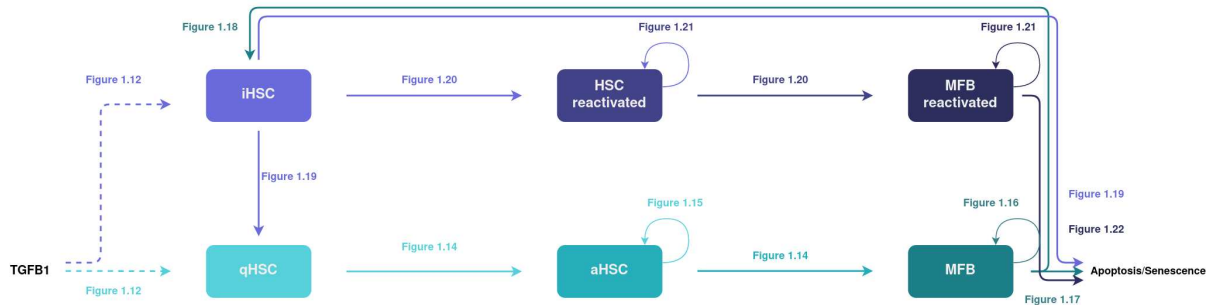


FIGURE 19 – The behavior of hepatic stellate cells. TGFβ1 protein induces the activation of quiescent hepatic stellate cells (qHSC) (green dashed arrow). The activated hepatic stellate cells (aHSC) differentiate to myofibroblasts (MFB) (light green arrow). MFB have two potential behaviors (green arrow) : either they enter in the apoptosis/senescence pathway or are inactivated to inactivated hepatic stellate cells (iHSC). iHSC have three potential behaviors : either they reverse into qHSC (light purple arrow), or they enter in the apoptosis/senescence pathway (light purple arrow), or they are activated by the protein TGFβ1 (purple dashed arrow), leading to reactivated hepatic stellate cells (HSC reactivated) (light purple) that may differentiate to reactivated myofibroblasts (MFB reactivated) (purple). The reactivated MFB have only one behavior, that is the apoptosis/senescence pathway (purple).

In Fig. 19 is sketched the potential behavior of an occurrence of hepatic stellate cell. This diagram itemizes the different transformation processes between the different forms of hepatic stellate cells. In particular, it describes how hepatic stellate cells are activated by the TGFβ1, and differentiate into myofibroblasts which undergo different processes. An important point is the inactivation pathway leading to the generation of inactivated hepatic stellate cells which differ from the quiescent phenotype. Inactivated cells are more quickly reactivated by TGFβ1 than quiescent ones, thereby amplifying cell response in chronic diseases. The apoptotic/senescence pathway allows for the feedback control (to keep the explanation simple) of the activation pathway.

5.1.2 Some elements of biochemistry

Before describing the interaction rules, we give some reminders of basic elements of biochemistry. Our goal is to explain how rate constants are defined and computed. As often as possible, we try to define them with respect to the reaction half-time, that is the time after which on average, half of the reactants of a reaction have been consumed.

5.1.2.1 Reaction half-time

Let us consider a first-order reaction of the form $A \rightarrow B$. We assume that the time when each instance of the reaction is applied is drawn randomly according to an exponential law with parameter k . That is to say that given an occurrence of the component A , the probability that this occurrence has turned into an occurrence of the component B after time t is equal to $(1 - e^{-k \cdot t})$.

Assuming that A is in large quantity in the system, the expectation $E[A](t)$ of the quantity of A remaining in the system at time t , is defined by the following equation :

$$E[A](t) = E[A](0) \cdot e^{-k \cdot t}$$

where $E[A](0)$ denotes the initial quantity of A (we write it as an expectation for the sake of homogeneousness).

The half-time of the reaction, that is written $t_{1/2}$, is then the time so that the expectation of the quantity of the component A that has been consumed at time $t_{1/2}$ is equal to half of the initial quantity of this component. That is to say that :

$$\frac{E[A](0)}{2} = E[A](0) \cdot e^{-k \cdot t_{1/2}}.$$

It follows that :

$$k = \frac{\ln 2}{t_{1/2}}.$$

In the case of a degradation reaction, that is to say, when the reaction has no product, the half-time of a reaction is also called its half-life time.

5.1.2.2 Conversion

In practice, reaction time is documented in the literature in various forms. Sometimes, it is documented as the time taken to transform all the occurrences of the component A

into occurrences of the component B . Since there always remains a residual quantity, we interpret this as the time $t_{99\%}$ so that the expectation of the quantity of the component A that has been consumed by the reaction at time $t_{99\%}$ is equal to 99% of the initial quantity of this component.

The conversion from $t_{99\%}$ to $t_{1/2}$ can be made thanks to the following reasoning. Given q between 0 and 1, we can define t_q the time so that the expectation of the component A that has been consumed is equal to the fraction q of the initial quantity of this component. The duration t_q is defined by the following equation :

$$(1 - q) \cdot E[A](0) = E[A](0) \cdot e^{-k \cdot t_q}.$$

It follows that :

$$\frac{\ln 2}{t_{1/2}} = \frac{-\ln(1 - q)}{t_q}.$$

Thus,

$$t_{1/2} = \frac{\ln 2}{\ln \frac{1}{1-q}} \cdot t_q.$$

We can conclude that :

$$t_{1/2} = \frac{\ln 2}{\ln 100} \cdot t_{99\%}.$$

5.1.2.3 Production equilibrium

It often happens that a degradation rule is counter-balanced by a synthesis rule in order to maintain an expected average amount of components in stationary regime.

Let us consider two reactions, a degradation reaction $A \rightarrow \cdot$ with a half-life time $t_{1/2}$ and a synthesis reaction $\cdot \rightarrow A$ at a rate k . The goal is to set the rate constant k so that the expected average amount of the component A is equal to $E[A]_{eq}$ in stationary regime. This means that, when the quantity of A is equal to $E[A]_{eq}$, the overall propensity of the degradation rule shall be equal to the one of the synthesis rule. That is to say that :

$$k = \frac{\ln 2}{t_{1/2}} \cdot E[A]_{eq}.$$

It is worth noting that this parameterization does not enforce a rigid equilibrium. This ensures only the eventual behavior of the system in the absence of other mechanisms that could modify the quantity of the component A .

5.1.2.4 Erlang distributions

The exponential law is defined by only one parameter. As a consequence, the standard deviation of the time a given event may take, is fully defined by its average time. This is not always satisfying from a modeling point of view, since some processes may require time distributions with different standard deviations.

A solution to this issue consists in decomposing a given interaction into several intermediary steps, each of these being executed according to an exponential distribution. The resulting composite process satisfies a so called Erlang-distribution [83] that is defined by two parameters (the average time of each intermediary step and the number of these steps). For instance, we may consider the sequential composition of two steps the duration of each of which being defined by an exponential law with a same parameter, and compare it with a single process the duration of which is defined by an exponential law, twice as slow as each step of the composite process. Then the standard deviation for the time so that half of the quantity of the initial component has completed the two intermediary steps is less than the standard deviation of the time so that half of the quantity of the initial component has completed the single step process. Moreover, the time so that the expectation of the quantity of the initial component that has completed the two-step process is equal to 99% is shorter than the time that is defined the same way for the single step process.

We do not know how to define analytically the reaction half-time of the intermediary processes with respect to the overall completion time of the process. Instead, we fit these values empirically, by simulating the behavior of the intermediary interactions (without considering the rest of the model).

5.1.3 Interaction rules

We now itemize the interaction rules of our model. The rate constants are parameterized with some values essentially found in the literature. The value of these parameters is given after the description of the rules in Fig. 3 on page 96.

5.1.3.1 The behavior of TGFB1 proteins

In Fig. 20 is specified the behavior of TGFB1 proteins.

The degradation of the protein TGFB1 is described in Fig. 5.20(a). In general, the occurrences of the protein TGFB1 are spontaneously degraded only in their active form.

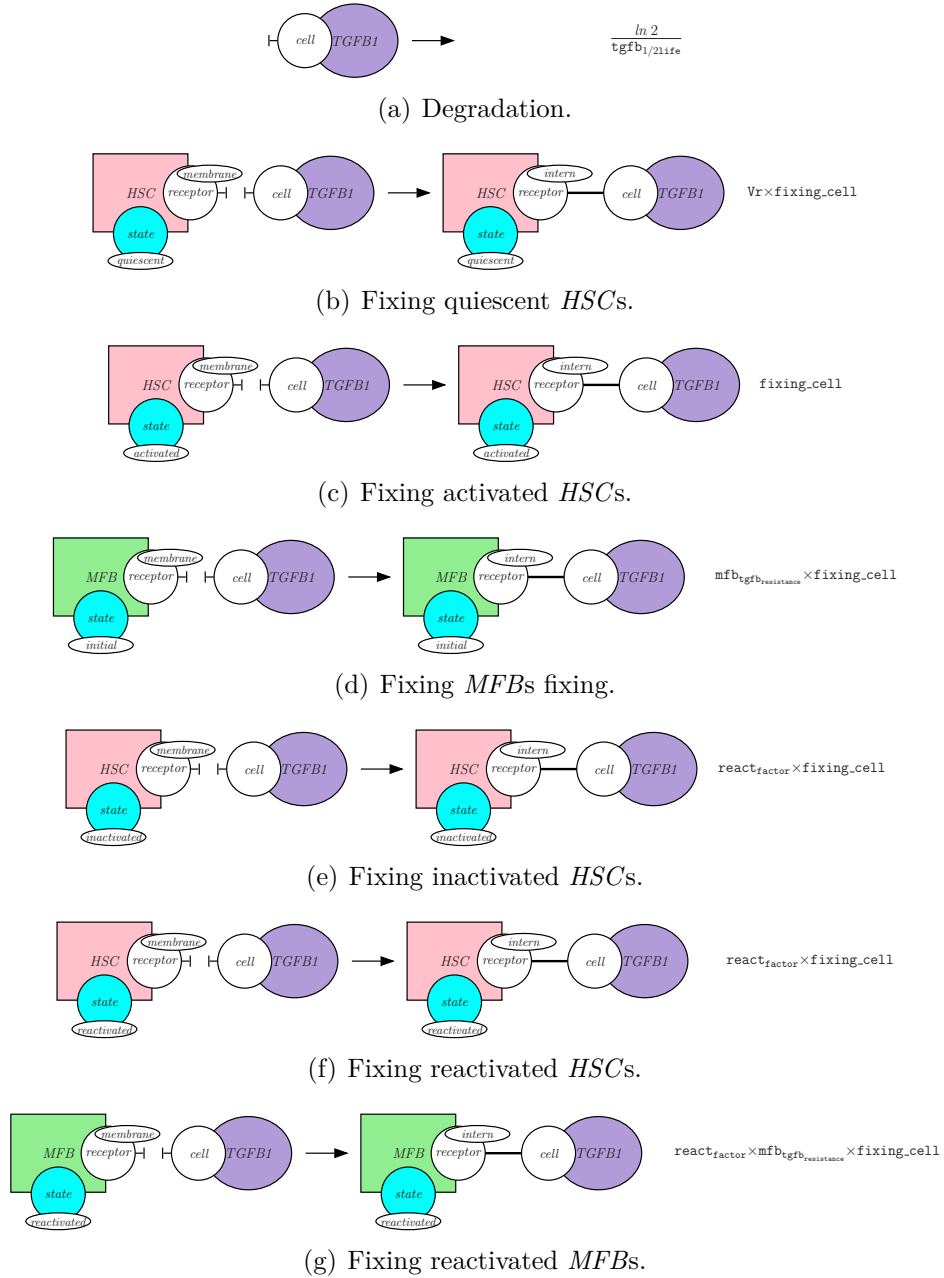


FIGURE 20 – The behavior of the protein *TGFB1*.



FIGURE 21 – The renewal of quiescent hepatic stellate cells.

Yet, in this model, only the active form of TGFB1 is involved, this is why the activation state of the occurrences of the protein TGFB1 is omitted. Inactive form of the protein TGFB1 plays an important role in the extracellular matrix. Additionally, this rule specifies that only the occurrences of the agent *TGFB1* that are not linked to any other agent may be degraded. The degradation rate is set according to the half-time of the protein.

The rules in Figs. 5.20(b), 5.20(c), 5.20(d), 5.20(e), 5.20(f), and 5.20(g) describe the interactions between TGFB1 and cells, the latter being in different states. We use several rules in order to give them different rates. Interactions with *TGFB1* agents are possible only when cell agents have a free *receptor* at their *membrane* and *TGFB1* agents are in their *active state* (which is omitted in our model) and not bound to any other agents. The receptors that bind TGFB1 are internalized leading to the state called *intern*. In order to preserve a pool of qHSC for cell renewal, we modulate the rule rate from (Fig. 5.20(b)) with a variable V_r that stands for the degree of renewal. This variable ranges between 0 and 1 according to the amount of cells that are currently bound to TGFB1 proteins. The more cells are bound to some TGFB1, the lower V_r will be, till reaching the value 0. Compared to hepatic stellate cells, myofibroblasts are less sensitive to the protein TGFB1 [224], therefore we reduced the rate of transformation induced by TGFB1 binding by 41,6% ($41,6\% = \text{HSC_proliferation} / \text{MFB_proliferation}$), whatever the states of the *MFB* (*initial* or *reactivated*) are. Compared to the cells in the state qHSC and MFB, the cells in the state iHSC, HSC reactivated, and MFB reactivated are more sensitive to TGFB1 [224], that is why we increased the corresponding rate constant (4 fold time).

5.1.3.2 Renewal of quiescent HSCs

The quantity of quiescent HSC in the liver results from an equilibrium between the production and the degradation of these cells. While qHSCs have mesenchymal stem cell origin [204, 117], there is no information about the qHSC production and their renewal in normal and pathological livers. Upon TGFB1 stimulation, qHSC are transformed to-

wards aHSC and the pool of qHSC is likely consumed. We introduced a variable V_r to preserve a residual pool of qHSC (see page 82). The degradation and production rules for hepatic stellate cells in their quiescent form are described in Fig. 21. The degradation of qHSC occurs when they have a *null* activation and no links with any other agents (see Fig. 5.21(b)). In Fig. 5.21(a), the state of the quiescent hepatic stellate cells which are synthesized has to be fully specified. It is written that their activation state is *null* and their receptors are free and on the membrane.

The rates of both rules are set according to the explanations that were given in Sec. 5.1.2 on page 80 so as to ensure the renewal rate of the hepatic stellate cells in their quiescent form and also the overall amount of them in regular regime.

5.1.3.3 Activation and differentiation

The activation of an occurrence of a hepatic stellate cell may happen when the three following conditions are satisfied : it shall be in a quiescent form (which is written as *quiescent*), its activation level shall be null (which is written as *null*), and it shall be activated by some TGF β 1 proteins (Fig. 5.22(a)). This activation modifies the conformation of the hepatic stellate cell, which is now an activated one (which is written *activated*) in the first stage of activation (which is written as *stage0*).

Activation is a gradual process (Figs. 5.22(b) and 5.22(c)), the hepatic stellate cells enter in their second activation stage (which is written as *stage1*), and their third and last one (which is written as *complete*). During this process, they remain in their activated form.

Once in the last stage of activation, hepatic stellate cells may differentiate into myofibroblasts. This process is done in three stages. The first step consists in replacing a fully activated hepatic stellate cell into a myofibroblast (*MFB* agent) in initial state (which is written as *initial*) and in the first differentiation stage (which is written as *stage0*). The description of this first step requires four rules (Fig. 5.22(d)). The main reason is that the state of the cell receptors shall be maintained. Yet since *HSC* and *MFB* are two different agents, it is not possible to inherit information from the *HSC* agents that are consumed to the *MFB* agents that are created. Thus, the solution is to write one rule for each potential state of the interaction site *receptor*, and there are four of them.

After differentiation of HSC towards MFB, the latter undergoes two other steps of differentiation. The rule in Fig. 5.22(e) describes the passage from the initial stage into the second one (which is written *stage1*). Then, the rule in Fig. 5.22(f) describes the

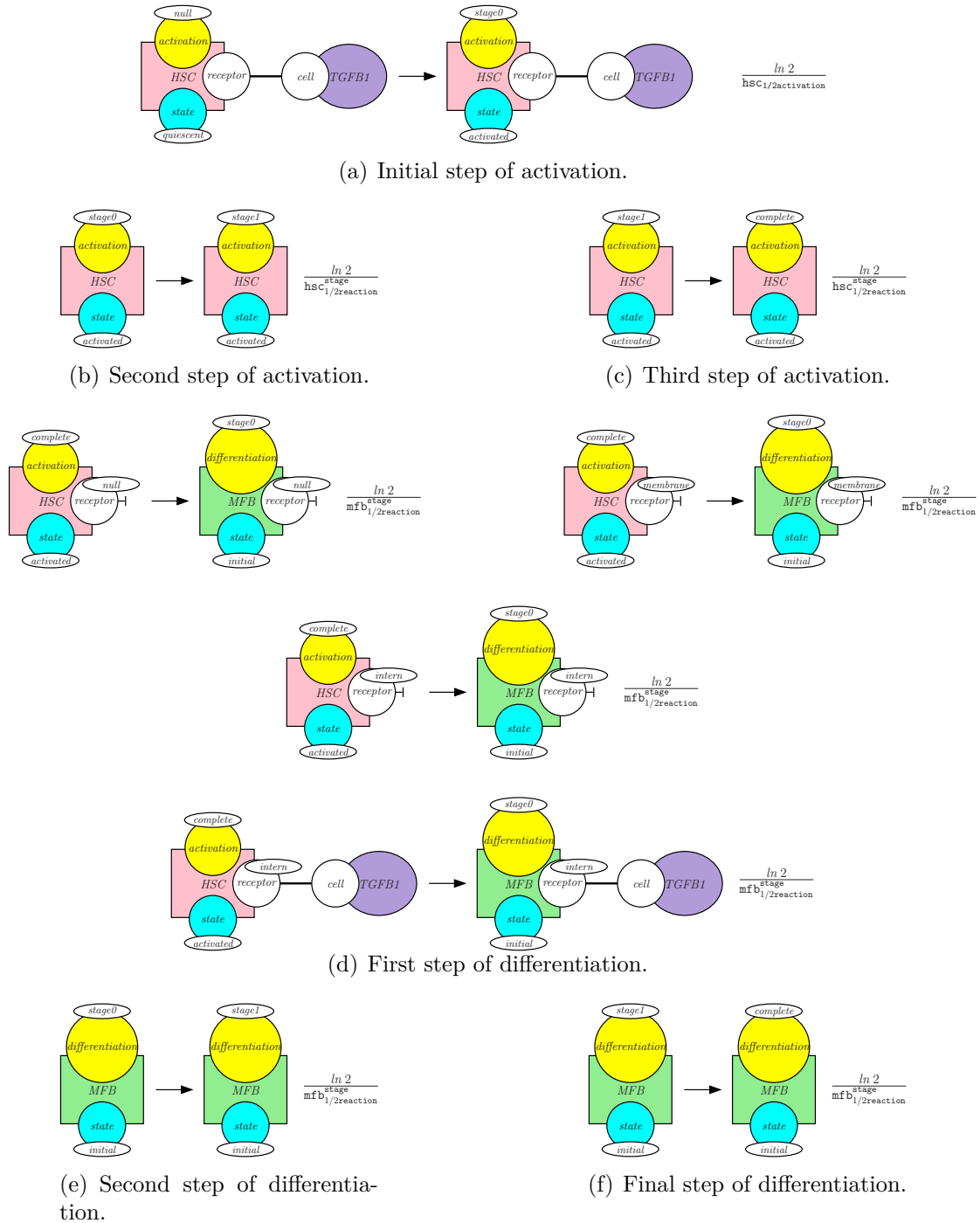


FIGURE 22 – Activation of hepatic stellate cells and formation of myofibroblasts.

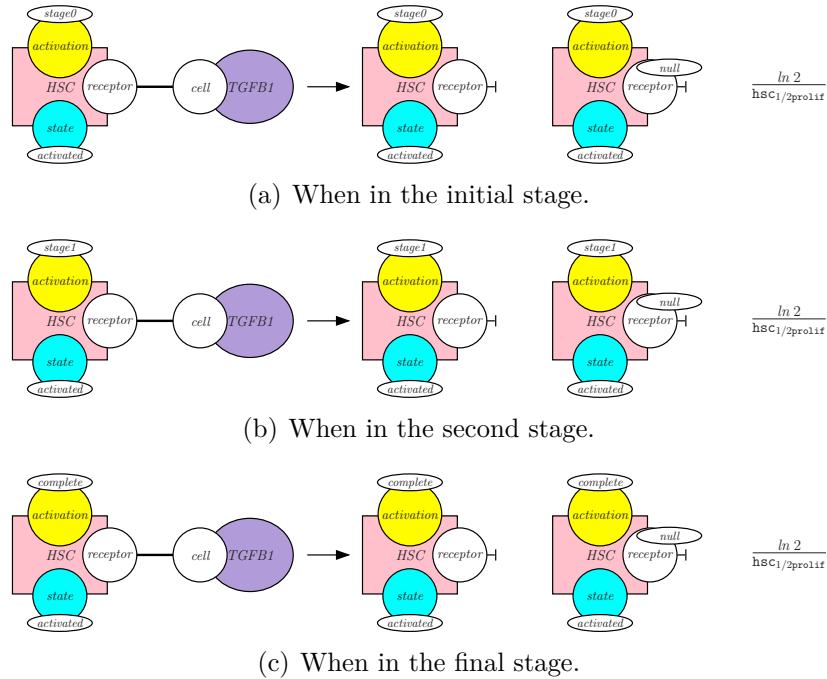


FIGURE 23 – Proliferation of activated hepatic stellate cells.

passage from the second stage into the last one (which is written *complete*).

The rates of activation and differentiation steps are defined by the means of a reaction half-time, following the guidelines that were given in Sec. 5.1.2. Each activation step shares the same rate, while each differentiation step shares another one.

5.1.3.4 Proliferation of activated hepatic stellate cells

The activation of hepatic stellate cells by some occurrences of the *TGFB1* protein may induce their proliferation (Fig. 23).

There are three different rules according to the activation stage of the cells. As a result of proliferation, the occurrences of the *TGFB1* agent are consumed, and the occurrences of the *HSC* agent are duplicated. It is worth noting that we have assumed that the new occurrences are in the same state and in the same activation stage as the occurrences which have given them birth. But, their receptors are not operational yet (which is written as *null*).

The rate of the rules are computed from the half-time of the proliferation reaction.

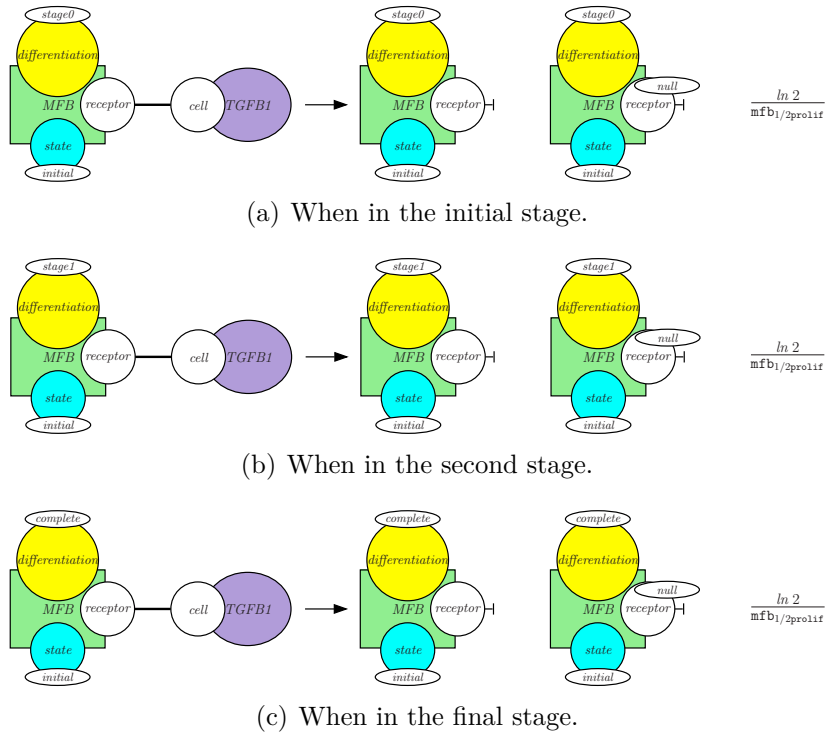


FIGURE 24 – Proliferation of myofibroblasts.

5.1.3.5 Proliferation of myofibroblasts

The proliferation of myofibroblasts works exactly as the proliferation of activated hepatic stellate cells. Myofibroblasts must be in the *initial* state and activated by some occurrences of the TGFβ1 protein. Occurrences of myofibroblasts may then be duplicated, conserving the state and the differentiation stage. The receptors of the newly created myofibroblasts are not operational yet (which is written as *null*). Proliferation speed is set by the means of the proliferation half-time of myofibroblasts.

5.1.3.6 Apoptosis and senescence of myofibroblasts

Upon the action of the TGFβ1 protein, hepatic stellate cells are activated and differentiated into myofibroblasts which are in charge of tissue repair. Moreover, when TGFβ1 is consumed, some of these MFB are eliminated (around 50%) through the apoptosis/senescence pathway which is modeled in Fig. 25, the rest of them becomes inactivated (for 50%) through an inactivation pathway which is depicted in Fig. 26.

Due to the lack of information, the processes of apoptosis and senescence have been

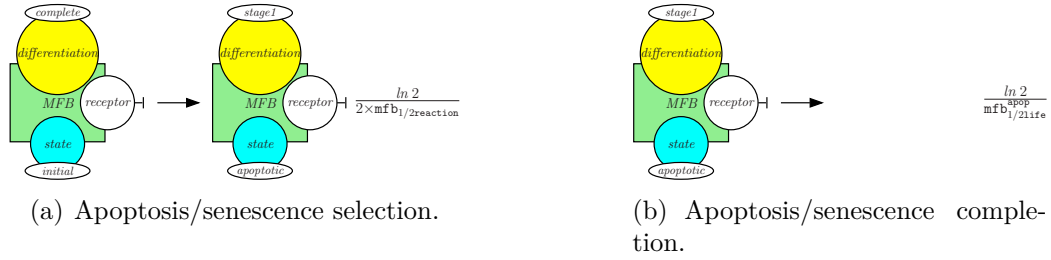


FIGURE 25 – Apoptosis/senescence pathway.

merged. The resulting pathway is made of two steps. The first step (Fig. 5.25(a)) consists in marking the myofibroblast for apoptosis/senescence. It requires the myofibroblast to be in its initial form (which is written *initial*), in the final stage of differentiation (which is written as *complete*), and its receptors to be free. As a result, the state is changed into *apoptosis*. We use a different scale for the differentiation level of the myofibroblasts that are marked for apoptosis, hence the differentiation level is set to *stage1*. The rate constant of the interaction accounts for the half-time reaction of myofibroblasts and the fact that only half of them follow this pathway.

The second step consists in the degradation of the myofibroblast (Fig. 5.25(b)). The degradation rate is computed from the half-life time of the myofibroblasts once they have entered the apoptosis/senescence pathway.

5.1.3.7 Inactivation of myofibroblasts

Inactivation of myofibroblasts is a two-step process which is depicted in Fig. 26. This process turns them back to hepatic stellate cells. Yet, since the cells that come from the inactivation of myofibroblasts have a different behavior, we call them inactivated cells.

The first step (Fig. 5.26(a)) consists in turning occurrences of myofibroblasts into hepatic stellate cells in inactivated form (which is written *inactivated*). This requires the occurrences of myofibroblasts to be in their initial state (hence they have not entered the apoptosis pathway, and they cannot come from the reactivation of an inactivated cell) and in the final stage of differentiation. Moreover, binding with occurrences of the TGFβ1 protein blocks this process, thus their receptors are assumed to be free. No further assumption is required on the state of the receptors of the occurrences of the myofibroblasts. As a result, the occurrences of myofibroblast are replaced with occurrences of hepatic stellate cell in the initial state and in the first stage of inactivation (which is written *stage1*). The state of the receptors is maintained. Since the agents *MFB* and *HSC* are different,

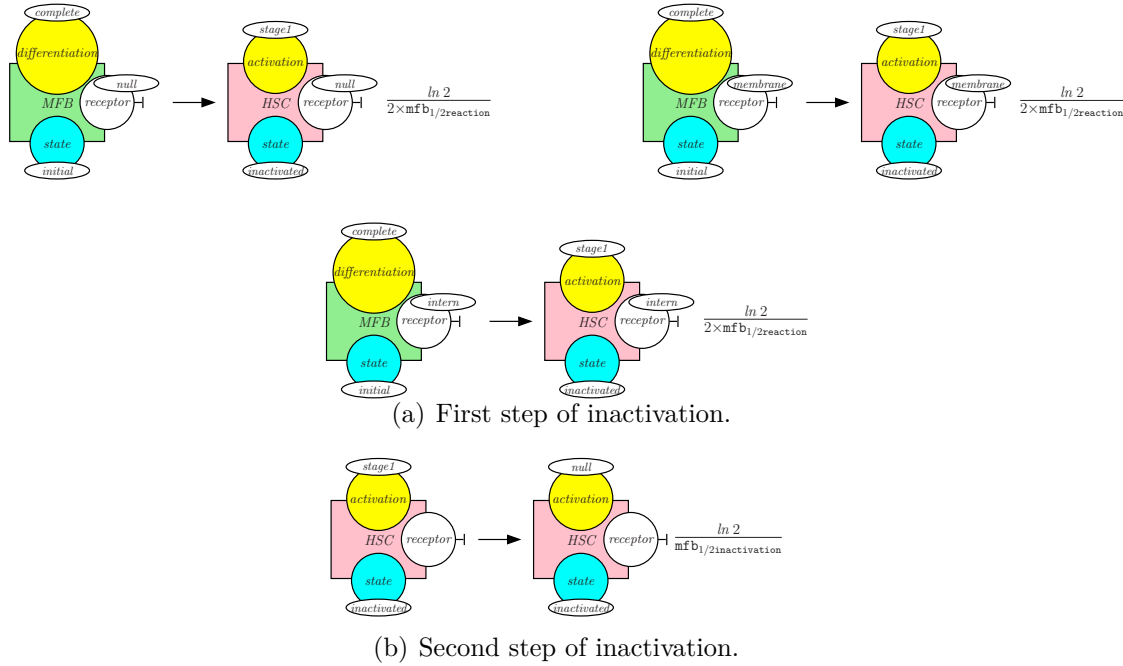


FIGURE 26 – Inactivation pathway.

we have to use three rules to model this (as we did already for describing the first step of differentiation of hepatic stellate cells (Fig. 5.22(d))). The rate of these rules are the same than the rules for apoptosis/senescence (Fig. 5.25(a)), since it stands for the behavior of the other half of the myofibroblasts.

The second step of the inactivation of hepatic stellate cells is described in Fig. 5.26(b). It still requires the cell not to be activated by the *TGFB1* proteins. As a result, the cell enters the final stage of inactivation (which is encoded by the activation state *null*). The rate of the rule for this second step of inactivation is defined by the half-time of inactivation of hepatic stellate cells.

5.1.3.8 Behavior of inactivated hepatic stellate cells

Inactivated hepatic stellate cells are either slowly eliminated through degradation process, or returned to the quiescent form, or reactivated by a new *TGFB1* stimulation into myofibroblasts.

Inactivated HSC can be eliminated through two processes, apoptosis/senescence or reversing. To undergo those pathways, *HSC* agents must be inactivated (which is written *inactivated*) and not under the process of activation (which is written *null*). Moreover,

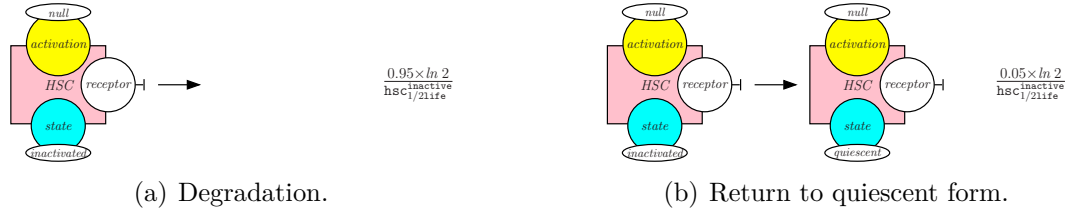


FIGURE 27 – Degradation and recycling of inactivated hepatic stellate cells.

these *HSC* agents must not be bound to *TGFB1* agents. We choose that 95% of iHSC will enter apoptosis/senescence (Fig. 5.27(a)), the remaining part will reverse into a quiescent state (Fig. 5.27(b)).

The 5% of iHSC undergoing reversion are similar to the one undergoing apoptosis/senescence and reverse into qHSC (Fig. 5.27(b)). The process of reversion forms similar qHSC than those previously described (Fig. 21). Nevertheless, the state of the receptor is conserved during reversing. The rates of degradation and return to quiescence rules depend on iHSC half-life time.

Upon a new *TGFB1* stimulation, inactivated hepatic stellate cells may be reactivated. This is a two-step process. In the first step, which is depicted in Fig. 5.28(a), the occurrences of the *TGFB1* agent are consumed and the conformation of the cell changes. They are now in the first stage of reactivation, which is formalized by the state *reactivated* and the activation level *stage0*. The second step, which is written in Fig. 5.28(b), puts the cell in the second stage of reactivation, its activation level is set to *complete*. The rate of the reactivation of iHSC is 4 times the rate of a *HSC* activation, that is why the parameter $\text{ihsc}_{1/2\text{reaction}}^{\text{stage}}$ is divided by the parameter $\text{react}_{\text{factor}}$ (that is set to the value 4).

Reactivated hepatic stellate cells follow the same process of differentiation than activated ones leading to the formation of reactivated myofibroblasts (which is written as *reactivated*) as described in Figs. 5.28(c) and 5.28(d). Similar to the differentiation of initial myofibroblasts (Fig. 22), the description of the first step of redifferentiation requires four rules (Fig. 5.28(c)). Those four rules aim at conserving the state of the receptors during the change from occurrences of the agent *HSC* into occurrences of the agent *MFB*. The rate of the rules for *MFB reactivated* stage depends on the half-time of the reaction needed to their formation.

After differentiation of *HSC* towards *MFB*, the latter undergoes another step of differentiation. The rule in Fig. 5.28(d) describes the passage into the last stage (written *complete*). The rate of the complete differentiation rule depends on the half-time of the

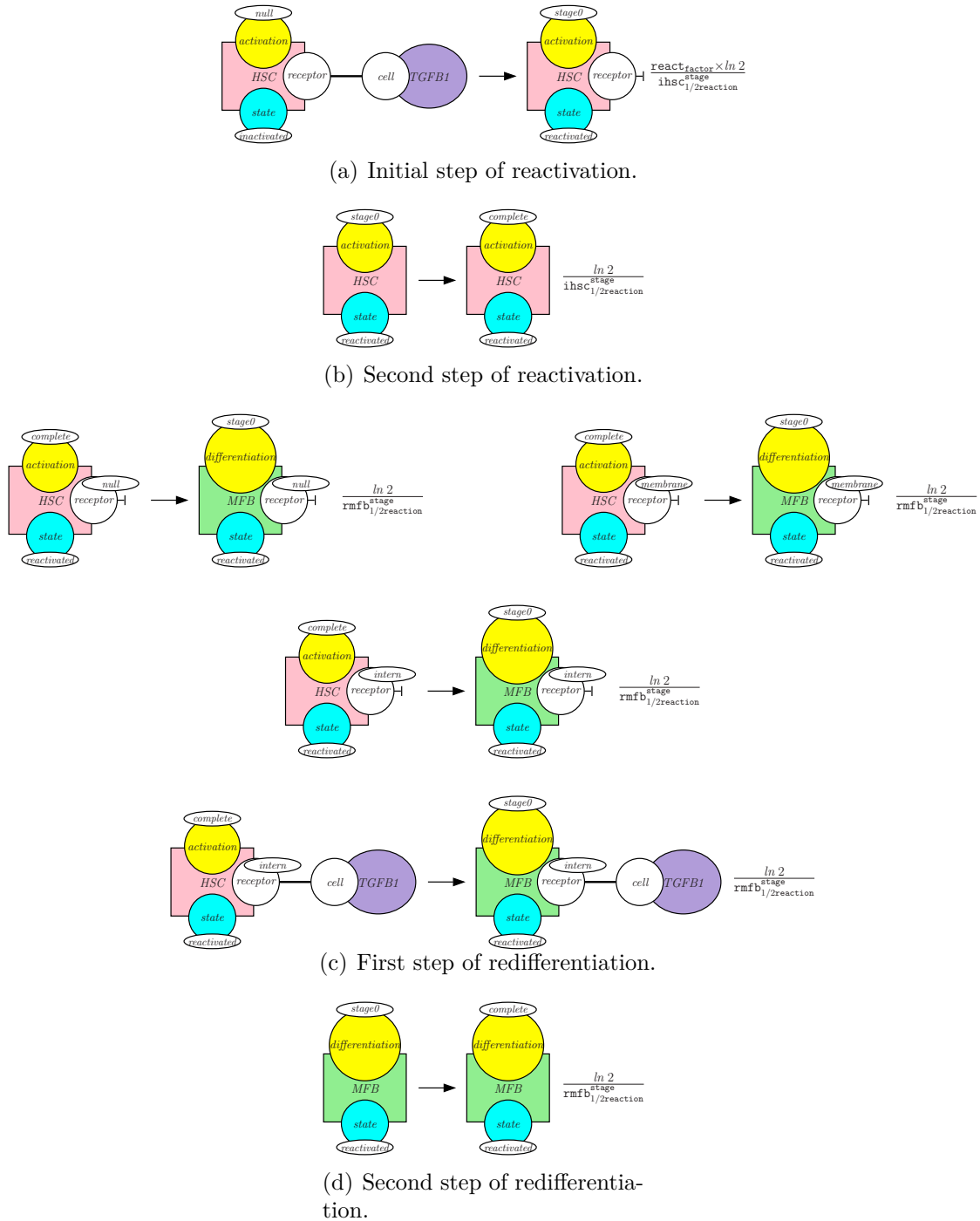


FIGURE 28 – Redifferentiation of inactivated hepatic stellate cells.

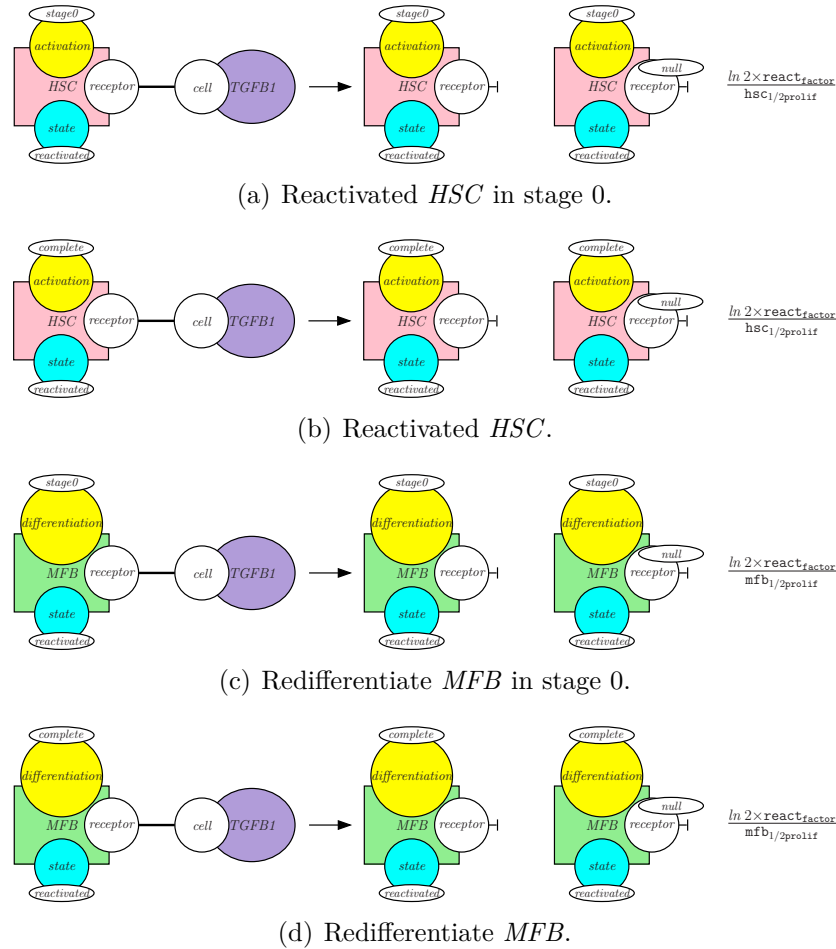


FIGURE 29 – Proliferation of reactivated cells.

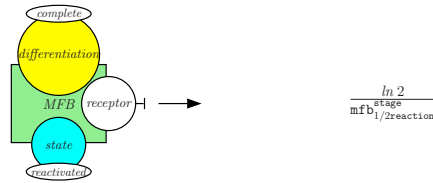
complete differentiation of the reactivated MFB.

5.1.3.9 Proliferation of reactivated cells

The proliferation of reactivated cells is formalized in Fig. 29. It works exactly as the proliferation of initial cells. Myofibroblasts and hepatic stellate cells must be in the state *reactivated* and bound to some occurrences of the protein TGFB1. Occurrences of *HSC* and *MFB* may then be duplicated, conserving the state of their sites *state* and *activation* for the formers and the state of their sites *state* and *differentiation* for the latter. The receptors of the newly created cells are not operational yet (which is written as *null*). Proliferation speed is set by the means of the proliferation half-time of each stage of activation and differentiation. In particular, the factor react_factor accounts for the fact

that reactivated cells proliferate more than firstly activated ones.

5.1.3.10 Degradation of reactivated *MFB*

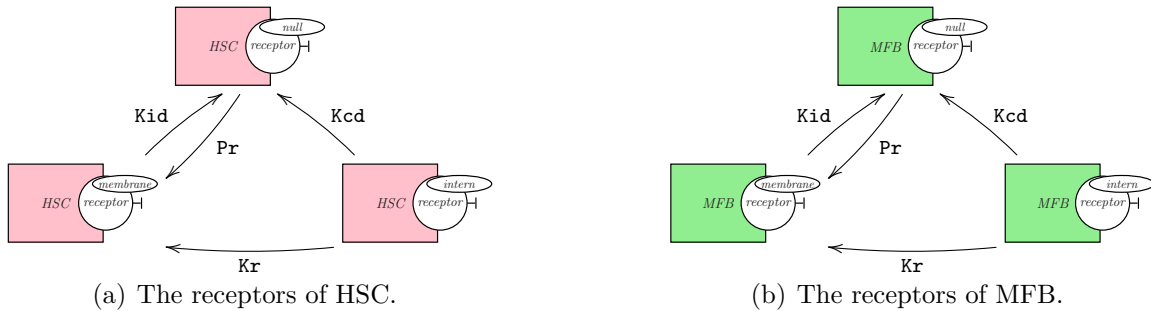


(a) Apoptosis/senescence.

FIGURE 30 – Apoptosis/senescence of reactivated myofibroblasts.

Unlike Myofibroblasts, no information shows a potential inactivation for reactivated MFB. Starting from that, the only way for *MFB* in a *reactivated* to disappear is through apoptosis/senescence (Fig. 30). It requires the myofibroblasts to be in their reactivated form (which is written *reactivated*), in the final stage of differentiation (which is written *complete*), and their receptors to be free. The reaction rate is computed from the half-life time of the myofibroblasts.

5.1.3.11 Behavior of receptors



(a) The receptors of HSC.

(b) The receptors of MFB.

FIGURE 31 – Receptors cycle.

The receptors of cells have also their own behavior which regulates the capability of cells to interact. This holds for the receptors of the protein TGF β 1 as well. The behaviors of these receptors in hepatic stellate cells and in myofibroblasts is sketched in Fig. 31 by the means of transition systems which show the possible changes from one state to

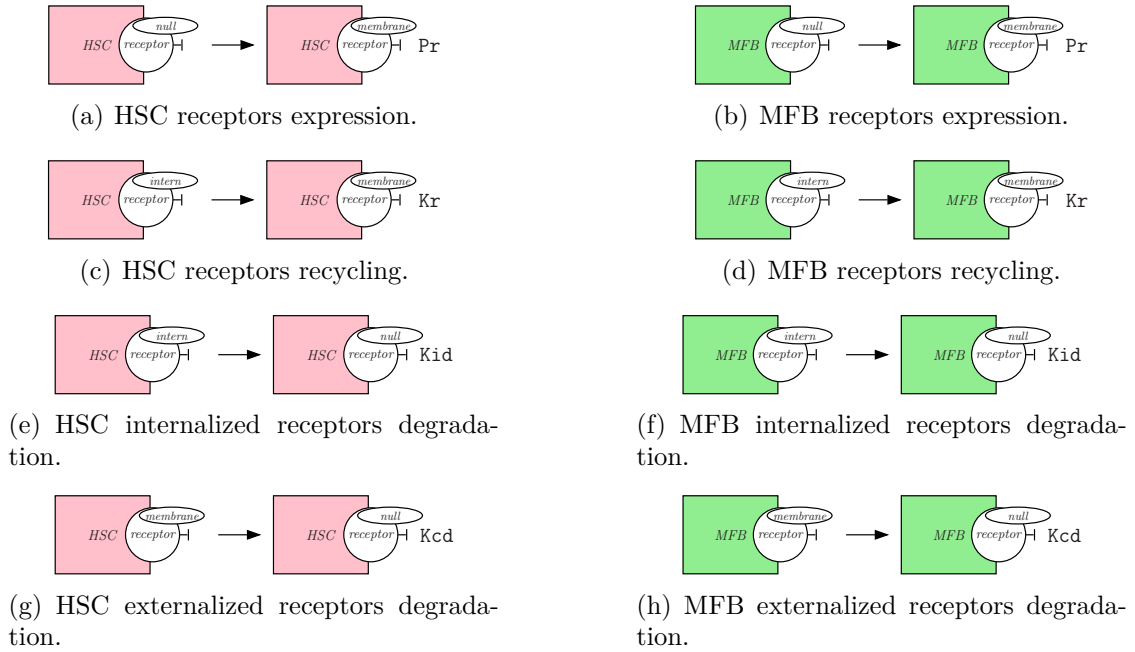


FIGURE 32 – Behavior of the receptors.

another, labelled with some rate constants. This behavior can be modeled by the means of four pairs of rules (Fig. 32) [261]. Those rules regulate the capability of the occurrences of the *HSC* and *MFB* agents to interact with occurrences of the *TGFB1* agent. The first pair of rules describes the production of the receptors for *HSC* and *MFB* agents (Figs. 5.32(a) and 5.32(b)). Hepatic stellate cells and myofibroblasts may produce their receptors when they are not present (which is written *null*). Newly produced receptors are expressed at the membrane of the cells (which is written *membrane*). The rate of production of receptors is given by the parameter *Pr*.

The second pair of rules describes the recycling of the receptors from the pool of internalized receptors to the membrane (Fig. 5.32(c) and 5.32(d)). Interacting with their ligands may induce the internalization of the receptors (which is written *intern*). This internalization may lead to the recovery of the receptor permitting its expression at the membrane (which is written *membrane*). The rate of this process is given by the parameter *Kr*.

The third pair of rules describes the degradation of the receptors from the pool of internalized receptors (Figs. 5.32(e) and 5.32(f)). This internalization may also lead to the degradation of the receptors (which is written *null*). The rate of this degradation process is given by the parameter *Kid*.

The fourth pair of rules describes the degradation of the receptors lying on the membrane (Figs. 5.32(g) and 5.32(h)). Receptors have a turnover leading to their degradation even without interaction with the protein *TGFB1*. In that case, they may change from the state *membrane* into the state *null*. The rate of this degradation process is given by the parameter *Kcd*.

5.1.4 Parameters

Definition	Symbol	Values	Acquisition	Reference
Environment				
Number of Disse's Space	Number_of_Disse_study	10	Calculated	[79]
Number of cells by Disse's Space	HSC_number_by_Disses	4	Calculated	[79]
Total number of cells	$hsc_{eq}^{quiescent}$	40	Calculated	
qHSC				
Half-life	$hsc_{1/2life}^{quiescent}$	90 days	Estimated	[145]
Half-activation time	$hsc_{1/2activation}$	0.17 hour	Calculated	[96]
Max of cell activated	<i>Vr</i>	0, 0.166 , 1	bibliography	[17]
aHSC				
Doubling time	$hsc_{1/2prolif}$	12.64 hours	Estimated	[224]
MFB				
Doubling time	$mf_{1/2prolif}$	8.43 hours	Estimated	[224]
Half-life	$mf_{1/2reaction}$	30 hours	Estimated	[145]
Apoptosis proportion		50 of total MFB	Bibliography	[81]
Inactivation proportion		50 of total MFB	Bibliography	[81]
MFB react				
Half-life	$mf_{1/2reaction}^{stage}$	25.29 hours	Estimated	
Doubling time	$\frac{mf_{1/2reaction}^{stage}}{reactfactor}$	1.58 hours	Estimated	[81]
iHSC				
Half-life	$hsc_{1/2life}^{inactive}$	90 days	Bibliography	[145]
Apoptosis proportion		95 of total iHSC	Estimated	[196], [163]
Quiescent return proportion		5 of total iHSC	Estimated	[196], [163]
Half-activation time	$\frac{ihsc_{1/2reaction}^{stage}}{reactfactor}$	0.0425 hours	Estimated	[81]
TGFB1				
Half-life	$tgfb_{1/2life}$	5 min	Bibliography	[268]
Inflammatory input	<i>TGFB_per_wave</i>	100 * number of qHSC	Calculated	
Fixing time	<i>fixing_cell</i>	3 min	Estimated	
Number of receptor by cell	<i>TGFB_factor</i>	7730	Bibliography	[177]
Receptor				
Recycling rate	<i>Kr</i>	0.5 hours	Bibliography	[261]
Production rate	<i>Pr</i>	0.066 hours	Bibliography	[261]
Degradation rate	<i>Kcd</i>	0.6 hours	Bibliography	[261]
Ligand induced degradation rate	<i>Kid</i>	0.066 hours	Bibliography	[261]
Time of transformation [110]				
aHSC transition stage	$hsc_{1/2reaction}^{stage}$	36.64 hours	Calculated	[17]
MFB differentiation stage	$mf_{1/2reaction}^{stage}$	46.22 hours	Calculated	[17]
MFB inactivation stage	$mf_{1/2inactivation}$	18.06 hours	Calculated	[81]
MFB apoptosis stage	$mf_{1/2life}^{pop}$	3.62 hours	Estimated	
IHSC reactivation stage	$ihsc_{1/2reaction}^{stage}$	6.31 hours	Calculated	[81]
MFB second differentiation stage	$rmf_{1/2reaction}^{stage}$	6.31 hours	Calculated	[81]
Factors				
IHSC reactivation stage	<i>reactfactor</i>	4	Bibliography	[81]
MFB TGFB1 resistance	$mf_{tgfb,resistance}$	0.416	Bibliography	[224]

TABLEAU 3 – Parameterization of the model.

In the previous section we described the rules present in our model, the table in Fig. 3 contains the parameters used to define the rate of these rules. Most of these parameters were found in the literature; for some of them, some calculations and estimations were done. Sometimes, information was lacking, that is why for few parameters we estimated their values using literature and our knowledge. As explained previously in Sec. 5.1.2, defining the reaction half-time of intermediary stages with respect to the overall completion time of the process is complicated. The computation of those reaction half-times has been made empirically.

5.2 Results

5.2.1 Static analysis

Before sampling the trajectories of the model, we use the static analyzer KaSa [32] so as to check the structural invariants. The objective is two-fold. Firstly the analyzer may detect some rules that will never be applied in the model. If so, this would mean that some parts are missing and that the model should be completed accordingly. It may also be due to some typos that should be corrected. Secondly, we want to check whether the intended relationships among the state of interactions sites in the different cell conformations hold effectively.

The analysis takes about 0.08 seconds on a 2,3 GHz Intel Core i9 8 cores MacBook Pro. The analysis detects no applicable rule and infers the structural invariants that are shown in Fig. 33. These invariants take the form of some refinement lemmas. They are written as logical implications. On the left hand side there is a pattern that specifies some conditions about the conformation of a cell. On the right hand side, this pattern is completed by providing the list of all the potential states some sites may take on every occurrence of this pattern in a reachable state.

In particular, the analysis detects and proves the following properties about the different stages of the different forms of cells.

- Quiescent hepatic stellate cells may be only in the stage *null* (Fig. 5.33(a));
- Activated hepatic stellate cells may be only in the stages *stage0*, *stage1*, and *complete* (Fig. 5.33(b));
- Inactivated hepatic stellate cells may be only in the stages *null* and *stage1* (Fig. 5.33(c));
- Reactivated hepatic stellate cells may be only in the stages *stage0* and *complete*

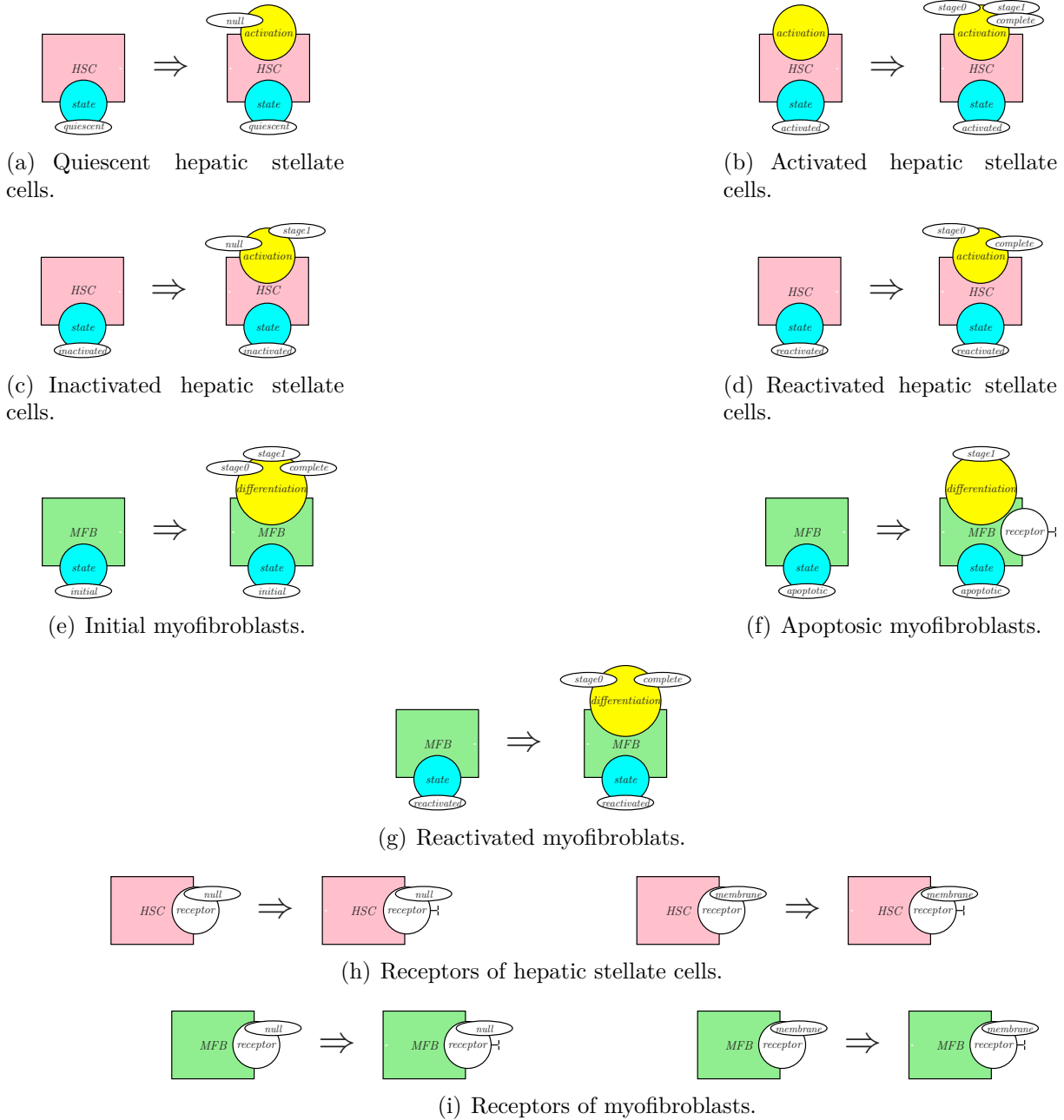


FIGURE 33 – The result of static analysis. For each implication, every occurrence of the pattern on the left hand side of an implication in a reachable state necessarily satisfies the conditions described in the right hand side.

- (Fig. 5.33(d));
- Myofibroblasts in their initial form may be only in the stages *stage0*, *stage1*, and *complete* (Fig. 5.33(e));
 - Myofibroblasts on the way to apoptosis may be only in the stage *stage1* and their receptors may not be bound (Fig. 5.33(f));
 - Reactivated myofibroblasts may be only in the stages *stage0* and *complete* (Fig. 5.33(g)).

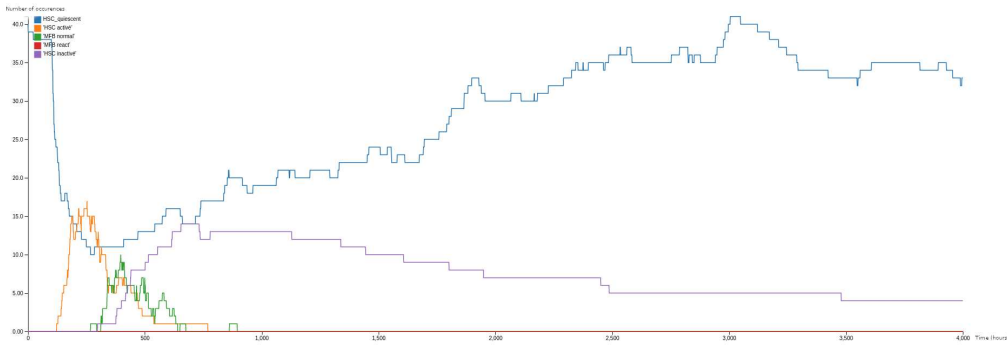
The analysis also discovers that the receptors that are either missing (*null*), or on the membrane (*membrane*) of the cells, are necessarily free both in the case of hepatic stellate cells (Fig. 5.33(h)) and myofibroblasts (Fig. 5.33(i)).

5.2.2 Underlying reaction network

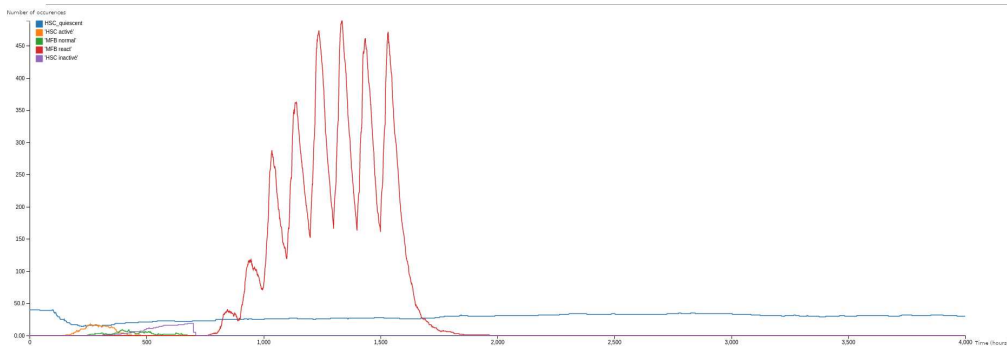
The set of rules may be compiled into a reaction network or equivalently into a system of ordinary differential equations thanks to the tool KaDe [40]. This computation takes about 0.04 seconds on a 2,3 GHz Intel Core i9 8 cores MacBook Pro. The resulting system involves 56 kinds of complexes.

5.2.3 Simulations

Simulating our model provides the results given in Fig. 34. In Fig. 5.34(a), we mimic an acute inflammatory aggression by one input of TGFB1. This is indeed the regime under which we have calibrated our model thanks to the information available in the literature. In Fig. 5.34(b), chronic inflammation was created by multiple inputs of TGFB1 (10 times here), and the expected behavior was successfully reproduced by our model. Each simulation takes around 14 seconds of CPU to simulate 4000 hours of biological time on a 2,3 GHz Intel Core i9 8 cores MacBook Pro. These curves show the evolution of the number of occurrences of each form of cells with respect to time (in hours). To parameterize the model, we choose to work in a volume corresponding to 10 Disse Spaces, containing among 40 quiescent HSC (calculated from [79]). Adding TGFB1 to the model at time 100 initializes the activation process, transforming quiescent HSC (qHSC, blue) towards activated HSC (aHSC, orange). After that, the newly activated HSC enter in differentiation process leading to the formation of MFB (green). When TGFB1 is completely consumed, MFB are either eliminated by apoptosis/senescence or inactivated (Fig. 5.34(a)). The inactivation of MFB leads to the formation of inactivated HSC (iHSC, purple). The iHSC are either eliminated by apoptosis/senescence or reverse to-



(a) Response to an acute inflammatory aggression.



(b) Response to chronic inflammation.

FIGURE 34 – Simulations of the model. Curves represent the time evolution of the number of occurrences of each form of cells in response to TGFB1 inputs. Time is expressed in hours. In Fig. 5.34(a), one input of TGFB1 at time 100. In Fig. 5.34(b), there are eleven inputs of TGFB1, first at time 100 then starting from time 700, each 100 hours there will be an input of TGFB till getting ten inputs.

wards quiescent HSC. Upon a new stimulation by TGFB1, iHSC can also reverse towards reactivated HSC first, then in reactivated MFB (red). Iterative inputs of TGFB1 in the model quickly favor the accumulation of reactivated myofibroblasts, the inactivation and apoptosis/senescence pathways being surpassed by MFB proliferation and reactivation. Moreover, TGFB1 iterative inputs induce a population switch with diminution of initial cells and augmentation of reactivated cells. After few inputs of TGFB1, some inactivated HSC remain but after more inputs, they completely disappear leading to the saturation of the environment by reactivated MFB 5.34(b). The timing for iterative inputs of TGFB1 is crucial. When too close, only MFP and HSC proliferation is observed. When too separate, much inactivated HSC are eliminated leading to decrease the inflammatory answer.

The primary results are encouraging, showing the dynamic of hepatic stellate cells in function of the TGFB1 inputs. Our model successfully describes the behavior of hepatic stellate cells shown in Sec 5.1.1 and respects the time link to this dynamic; activation process takes around 168 hours, differentiation around 336 hours, inactivation around 664 hours [17, 81]. However, some parts of the model need to be reviewed. Firstly the proliferation process is one of our main problems. Cells proliferation is lower than expected, for example the number of activated HSC should be 3 fold higher than the number of cells undergoing activation process. It could come from a part that is missing to explain the dynamics of the model (as discussed in conclusion Sect. 5.3). It could also be solved by considering more intermediary stages in Erlang time-distributions (which would lower the standard deviation of their overall process durations, hence favoring cell synchronization, and amplifying the amplitude of their abundance peaks). Also, reactivated cells elimination should be homogenized with initial MFB apoptosis/senescence. Last but not least, the model is quite rigid. Some encoding artifacts have been used. For instance controlling HSC activation by a factor should be removed and the rules firing should control it without any help. Globally, this model needs to be perfected but the primary results are promising.

5.3 Discussion

This chapter was devoted to the rule-based language Kappa. As a realistic case study, we developed a model for the activation of hepatic stellate cells by the protein TGFB1. We have used knowledge from the literature to calibrate the model according to its expected response to a single acute inflammatory aggression and we were able to reproduce

the expected behavior in case of chronic inflammation (up to minor differences in response amplitudes). Beyond the benefit of formalizing executable models, Kappa offers a convenient syntax, close to biochemistry, which eases the modeling process, the potential updates of the model, and the documentation of it. In Sec. 5.1, the description of the Kappa interactions provides a practical road map to navigate among the different elements of the model.

Yet modeling requires a constant search for the most adequate abstraction trade-off. Models may be made arbitrarily precise when detailed information about biochemical mechanisms is available in the literature and when this amount of details does not overcome the *in silico* computational resources. In this model, we favored simplicity while sticking to the experimental observations. We aim to investigate further different parts of the model and include more details gradually. During this, we will check that the overall behavior of the model is preserved and explore the impact of these updates in specific regimes. Rule-centric approaches make this empiric modeling approach easier.

More specifically, we plan to investigate further on two parts of the model. The first one is about the receptors of cells. With a view to simplifying, we have abstracted all the receptors of each occurrence of cells, by a single Kappa site, called *receptor*. Moreover each agent *TGFB1* indeed stands for a pack of occurrences of the protein TGFB1. The main motivation is to spare computation time. The impact on the model is that all the receptors of an occurrence of a cell are considered to be in the same state. This abstraction could be refined by modeling the receptors of each cell and each occurrence of the TGFB1 protein individually. Yet the cost may become prohibitive both with respect to memory (each occurrence of the TGFB1 protein would be described explicitly) and to computation time (binding between occurrences of the TGFB1 protein and their receptors would occur at a very fast scale and the simulation would have to account for very frequent instances of this event). The issue with memory could be solved easily by the means of counters. The total number of occurrences of free TGFB1 could be modeled as a numerical value (called 'token' in Kappa), and the number of receptors in each of the four different states (*null*, *membrane*, *intern* and free, and *intern* and bound) should be described by the means of some counters, as enabled by a recent extension of Kappa [28].

The second part concerns the memory of cells. As we have seen, activated and reactivated hepatic stellate cells exhibit a different behavior. This means that these cells have a memory and that the history of each cell impacts its further behavior. In the current version, this is modeled by introducing different forms of cells, and providing each form

with different capabilities of interaction and rates. This modeling directly operates at the level of the phenotypes of the cells. It would be interesting to better understand where these different behaviors come from by modeling an abstraction of the protein content of the cells. This way, the behavior of each cell would directly emerge from its protein content. Due to the difference of time-scales and the lack of details in the literature, the protein content cannot be modeled precisely. Instead it could be described by equipping each cell with some abstract counters [28]. The value of these counters could increase upon activation and decrease during cool-down phases, while influencing the capabilities of interaction of the cell.

In Kappa, basic elements are interactions. Their rates provide information about their time-distribution independently from the rest of the system. In the literature, information about reaction durations takes different forms. It can be specified as the time to complete a given ratio of a given process, or defined more phenotypically by the time period between peaks of concentrations. In Kappa, only exponential time-distributions can be assigned to an interaction. This means that the average time of a reaction and its standard deviation are fully entangled. More diverse time-distributions may emerge as a result of sequential composition of reactions. For instance, Erlang time-distributions [83] may be obtained by modeling a process as the sequential composition of k intermediary steps. Intermediary steps reduce the time-variability of the full process and eventually lead to a fixed duration. Yet, this comes with a computational overhead. From a parameterization point of view, rates of intermediary steps are more difficult to guess and must be data-fitted.

One critical point of modeling TGF β 1-dependent activation of HSC is the identification of the parameters because of the lack of quantitative values. To overcome this issue, we merged information about HSC activation from both *in vivo* and *in vitro* experiments, however the difference in dynamics of activation/regulation between these two approaches has been already widely documented such as for gene expression [65]. Indeed, seeding quiescent HSC on plastic dishes induces HSC activation through molecular mechanisms that differ in part from mechanisms within liver tissue because of the presence of other cell types and microenvironments. An evolution of the model might be to integrate other cellular components and the biomechanical constraints that play a critical role in activation of HSC and in regulating myofibroblasts phenotype [275, 197]. Importantly, the phenotype of activated HSC is more complex than initially reported and the recent development of single cells RNAseq analyses allows now to demonstrate the heterogeneity of activated HSC [154]. We need to get more information on phenotyping the different HSC

species including the MFB states in order to better characterize the different reversibility pathways and to understand what contributes to the disequilibrium towards the disease progression.

While the biomechanical aspect is generally discarded in modeling cell activation-differentiation and cell signaling, modeling biological processes needs to take into account the physical constraints occurring in situ. Obviously extracellular matrix is the paradigm of such constraints and is the major regulator of cell responses [186, 44]. Extracellular matrix is not an inert material supporting cells within tissue but a plastic and complex network associating insoluble molecules such as collagen mostly arranged as supramolecular assemblies, glycoproteins such as fibronectin, and proteoglycans that consist in polypeptide backbone decorated by glycoaminoglycans that confer viscoelasticity and hydrophilic properties. This core matrisome characterized by [127] comprises among 300 proteins and is associated with ECM-affiliated proteins, ECM-regulators and secreted factors [231]. ECM composition and mechanical properties are specific of tissue and change during physiopathological processes such as development, inflammation, wound-healing, fibrosis and cancer. In the liver, the matrisome analyses were recently reviewed in [11] and showed the incomplete and heterogeneous description of this network. Because of the impossibility to catch a spatial and evolutive view of this network, models based on differential equations searched for driving molecules of the network behavior such as the core matrisome proteins, fibrin and collagen [172], and the ECM regulators [15, 134] in order to reduce the complexity.

Using Kappa language allows for the creation of an agent ECM that could interact with cellular and molecular agents. Different states could be attributed to ECM agent such as low, intermediate and high stiffness that control in turn the activation of HSC. While the relationship between ECM stiffness and HSC activation is known since a long time mainly by using 2D cell culture [275, 197], the quantitative evaluation of each molecule implication in stiffness remains to be clarified. Of course, the role of several molecules in liver stiffness during fibrosis has been characterized and Lysyl Oxidases (LOX) that catalyze cross-linking of collagen and elastin are some of these major actors [218]. The development of 3D-multi cellular hepatic models and microfluidic organ-on-a-chip liver models might be useful to get quantitative data about the contribution of molecular agents in stiffness and activation of HSC [185, 56]. In line with this, our future challenge aims to integrate the Kappa model for extracellular matrix-dependent TGFB1 activation that we recently developed [32]. TGFB1 is synthesized as a latent form (LAP-TGFB1) associated with the

Latent TGFB1 binding protein (LTBP1) that sequesters it within the extracellular matrix networks. The release of active TGFB1 depends on enzymatic activities but above all on mechanical strengths involving matrix components and membrane receptors [119]. As a regulatory loop, the activated HSC synthesize extracellular matrix components including TGFB1 and are involved in regulation of matrix plasticity thereby affecting TGFB1 activation. Integrating matrix components implicated in TGFB1-dependent HSC activation might improve the present model and allow us to identify new regulators of the equilibrium between repair and fibrosis.

5.4 Conclusion

Dans ce chapitre, nous avons apporté une preuve de concept de l'applicabilité de Kappa dans la modélisation de systèmes multi-échelles. Ce premier modèle de la dynamique d'activation des cellules étoilées par le TGFB1 comportait 52 règles et 33 paramètres. Il a permis de décrire avec succès le comportement des cellules étoilées en respectant le lien temporel de cette dynamique.

En conclusion, ces premiers travaux nous ont permis d'élaborer un modèle de la dynamique des cellules étoilées à partir d'observations *in vivo* et *in vitro*, de le formaliser avec la syntaxe Kappa et de réaliser des études de simulations. D'un point de vue biologique, ce modèle nous a permis de proposer une description détaillée du processus d'activation des cellules étoilées pendant le développement de la fibrose et sa réversion, et de mettre en évidence l'importance de leur réactivation répétée dans le contexte des atteintes chroniques. Cette première exploration nous a permis d'identifier le coût de mémoire comme limites de l'écosystème Kappa, pour modéliser ce modèle multi-échelle et de rechercher des adaptations, notamment en introduisant l'utilisation de *tokens* (jetons) et *counters* (compteurs) comme décrit dans la seconde partie de nos contributions de thèse (Section 6).

MODÉLISATION DE LA DYNAMIQUE DES CELLULES ÉTOILÉES HÉPATIQUES, INTRODUCTION DES *tokens* ET DES *counters*

Dans le précédent chapitre, nous avons démontré que le langage Kappa permettait de modéliser la dynamique fonctionnelle des cellules étoilées du foie. Cependant, le coût computationnel et l'approche rigide (certains processus étant définis explicitement plutôt qu'observés par la dynamique) du premier modèle nous a conduit à entièrement repenser le modèle en introduisant des *tokens* et des *counters* ainsi qu'en affinant les règles et données utilisées. Par commodité et parce que les scripts des modèles sont en langue anglaise, nous utiliserons les termes anglais pour présenter les résultats de ce chapitre. Pour rappel :

- Cellules étoilées du foie (CEF) : hepatic stellate cells (HSC)
- Myofibroblaste (MFB) : Myofibroblast (MFB)
- CEF quiescente : quiescent HSC (qHSC)
- CEF activées : activated HSC (aHSC)
- CEF inactivées : inactivated HSC (iHSC)
- CEF réactivées : reactivated HSC (react_HSC)
- MFB réactivés : reactivated MFB (react_MFB)
- MFB apoptotique/senescent : apoptotic/senescent MFB (apop_sene_MFB)

6.1 Trois modèles basés sur des règles pour la dynamique des cellules étoilées du foie

Sur la base de la littérature, nous avons développé trois modèles décrivant la dynamique des cellules étoilées (HSC) pendant la fibrose du foie et sa réversion (voir Figure 35 A_1 , A_2 , A_3). Ces modèles sont accessibles sur Github¹. Dans le foie normal, les HSC sont maintenues dans un état quiescent non prolifératif (qHSC) et stockent la vitamine A. Dans le foie agressé (virus, toxique, etc), la réaction inflammatoire aboutit à la production de TGF β 1, les qHSC perdent leurs gouttelettes lipidiques, passent à un état activé (aHSC), prolifèrent et sécrètent la matrice extracellulaire représentée par le collagène de type I (COL1). Les HSC totalement activées se caractérisent par un phénotype myofibroblaste (MFB) et une sécrétion plus importante de COL1. Lorsque l’agression est interrompue, les MFB sont soit éliminés par les voies de l’apoptose et de la sénescence (apop_sene_MFB), soit inactivés (iHSC), c’est-à-dire qu’ils atteignent un état proche, mais différent de l’état quiescent. Les iHSC peuvent être réactivées (react_HSC) par une nouvelle agression qui se traduit par une nouvelle stimulation par le TGF β 1. De la même manière que les HSC se transforment en MFB, les HSC réactivées se transforment en MFB réactivés (react_MFB), mais nous manquons d’information sur le devenir de ces cellules. Nous avons donc conçu deux modèles dans lesquels ces react_MFB sont soit i) totalement éliminés par les voies de l’apoptose et de la sénescence et il n’y a pas d’inactivation, ce modèle nommé *reactMFB-wo-inactivation* (wo, without) est décrit dans la Figure 35 A_1), soit ii) ils sont éliminés de la même manière que les MFB, c’est-à-dire par la voie d’apoptose/senescence et la voie d’inactivation, ce modèle nommé *reactMFB-with-inactivation* est décrit dans la Figure 35 A_2). Une autre inconnue concerne le devenir des iHSC en absence de signal d’agression (i.e de TGF β 1 dans le modèle). En effet, la réactivation des iHSC par le TGF β 1 a été décrite *in vivo* et *in vitro* dans la littérature [143, 256] avec notamment une suractivation de ces cellules comparée à celle des HSC natives. Par contre, le devenir de ces cellules en l’absence de TGF β 1 n’a pas été analysé. Les données expérimentales de Kisseleva [143] montrent une diminution de la population de iHSC au cours de la réversion. Nous avons donc conçu des modèles dans lesquels en absence de TGF β 1, les iHSC sont soit éliminées lentement, soit elles retournent à un état quiescent, participant ainsi au repeuplement du foie lésé avec de nouvelles qHSC. Pour tenir compte de cette hypothèse, cette nouvelle source potentielle de qHSC devra être pondérée par rapport à

1. https://github.com/MBougueon/HSC_model_Kappa

l'auto-renouvellement des qHSC natives afin de restaurer l'homéostasie des HSC dans le foie réparé. Ce modèle dans lequel les iHSC reversent vers un état quiescent constitue le troisième modèle nommé *iHSC-reversion-to-qHSC* qui est illustré dans la Figure 35 A_3 .

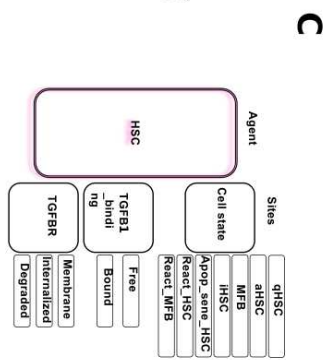
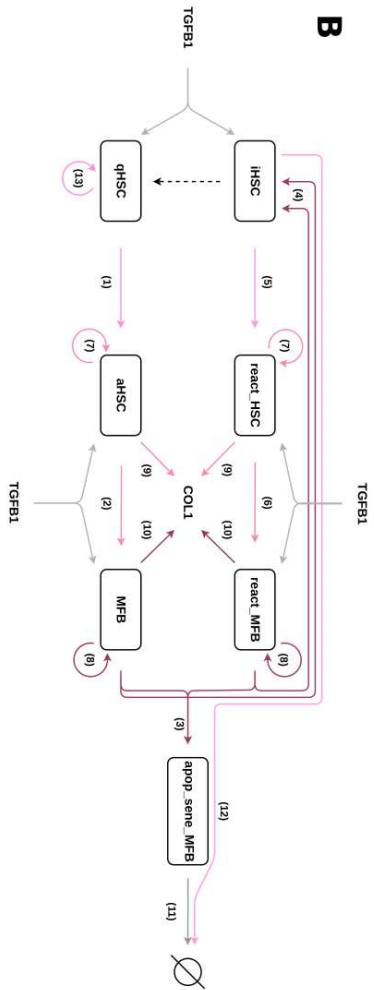
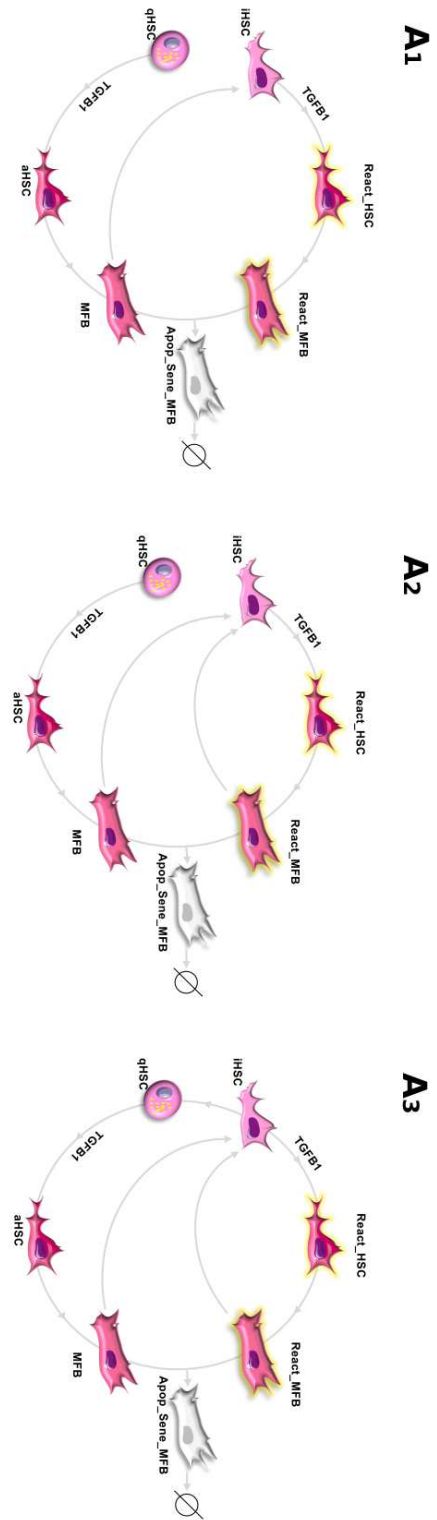
Finalement, nous avons défini trois modèles *reactMFB-wo-inactivation*, *reactMFB-with-inactivation*, *iHSC-reversion-to-qHSC*. Les processus biologiques de tous les modèles ont été intégrés dans la Figure 35 B et où sont aussi figurées les interactions entre le TGFB1 et les cellules aHSC, *react_HSC*, MFB et *react_MFB* qui conduisent à l'induction de leur prolifération et à la production de COL1.

Ces trois modèles ont été construits à l'aide du langage Kappa (version 4.2.12 de Kappa) et les simulations ont été effectuées à l'aide de l'outil KaSim (version 4.13.1 d'OCaml). Comme nous l'avons décrit précédemment, Kappa est un langage basé sur la réécriture de graphes à sites décrivant le comportement des agents et l'interaction entre ces derniers par des règles. Les agents ont des sites qui permettent de décrire des interactions physiques (ex : un site de liaison pour une protéine) ou des phénotypes (p.e. une modification post-traductionnelle pour une protéine ou des états cellulaires). Dans nos modèles, les cellules sont des agents et leurs états sont donc caractérisés par des sites. Pour notre agent *HSC*, ces sites sont au nombre de trois comme indiquée sur la Figure 35 C : 1) *cell_state* des HSC (*qHSC*, *aHSC*, *MFB*, *iHSC*, *react_HSC*, *react_MFB*, *apop_sene_MFB*), 2) le *TGFB1_binding* qui peut être libre ou lié et 3) l'état du récepteur au TGFB1 (*TGFBR*) qui peut être soit localisé à la membrane, soit internalisé, soit dégradé. Nous avons ensuite introduit des *counters* pour échelonner les étapes intermédiaires entre les états cellulaires et nous avons utilisé des *tokens* pour décrire les quantités de TGFB1 et de COL1 au lieu de détailler le comportement de chaque molécule, ce qui réduit considérablement le coût de calcul.

FIGURE 35 – Trois modèles pour décrire la dynamique des cellules étoilées (HSC).

A1) Représentation schématique du modèle *reactMFB-wo-inactivation*; **A2)** modèle *reactMFB-with-inactivation* et **A3)** modèle *iHSC-reversion-to-qHSC*. **B)** Représentation schématique des processus modélisés par Kappa dans tous les modèles. Les HSC sont des agents caractérisés par les sept états cellulaires (qHSC, aHSC, MFB, iHSC, *apop_sene_MFB*, *react_HSC* et *react_MFB*). Les règles décrivent les processus : (1) les qHSC sont activées en aHSC, (2) les aHSC sont transformées en MFB, (3) les MFB sont transformées en *apop_sene_MFB* pour être éliminées, (4) les MFB sont inactivées en iHSC, (5) les iHSC sont réactivées en *react_HSC* lors de la stimulation de TGFB1, (6) les *react_HSC* sont transformées en *react_MFB*. Comme les MFB, les *react_MFB* sont éliminés par apoptose et sénescence (3) ou par inactivation (4). (7) aHSC et *react_HSC* prolifèrent, (8) MFB et *react_MFB* prolifèrent, (9) aHSC et *react_HSC* produisent COL1, (10) MFB et *react_MFB* produisent COL1, (11) *apop_sene_MFB* sont dégradées, (12) iHSC sont dégradées. (13) auto-renouvellement des qHSC. L'effet de TGFB1 est représenté par des flèches grises, il induit l'activation des qHSC, la réactivation des iHSC, et la production de COL1 par aHSC, *react_HSC*, MFB et *react_MFB*. La réversion des iHSC en qHSC est représentée par une flèche en pointillé. **C)** Carte de contact Kappa. Les HSC sont des agents caractérisés par trois sites : *cell_state* (qHSC, aHSC, MFB, iHSC, *react_HSC*, *react_MFB*, *apop_sene_MFB*), *TGFB1_binding* (free ou bind) et *TGFBR* (membrane, internalized ou degraded)

6.1. Trois modèles basés sur des règles pour la dynamique des cellules étoilées du foie



6.2 Syntaxe de modélisation

Les trois modèles contiennent respectivement 75 règles et 41 paramètres pour le modèle *reactMFB-wo-inactivation*, 76 règles et 43 paramètres pour le modèle *reactMFB-with-inactivation* et 77 règles et 45 paramètres pour le modèle *iHSC-reversion-to-qHSC*. Ces règles peuvent se regrouper selon sept processus biologiques que sont 1) le renouvellement des qHSC, 2) la liaison du TGFB1 aux HSC, 3) la dynamique du TGFBR, 4) l'activation et la différenciation des HSC, 5) la prolifération des HSC, 6) la production et l'élimination du COL1 et 7) l'inactivation des MFB. Nous allons détailler la syntaxe utilisée pour modéliser ces sept processus. La syntaxe des règles utilisées est basée sur la syntaxe des *tokens* 6.2 et *counters* 6.1 qui pour rappel sont déclarés de la manière suivante :

$$A(x\{p\}, y[.], n\{\geq 1\}), B(y[.]) \rightarrow A(x\{p\}, y[1], n\{- = 1\}), B(y[1])@'k' \quad (6.1)$$

$$A(x\{p\}, y[.]), B(y[.]) \rightarrow A(x\{p\}, y[1]), B(y[1]) \mid - 10'ATP'@ |ATP| *' k' \quad (6.2)$$

Auto-renouvellement des qHSC. Le renouvellement des qHSC a été modélisé à l'aide de deux règles (voir Figure 36) Ces règles tiennent compte du fait que le site *TGFB1_binding* des qHSC doit être libre, puisque la liaison du TGFB1 induit l'activation des HSC.

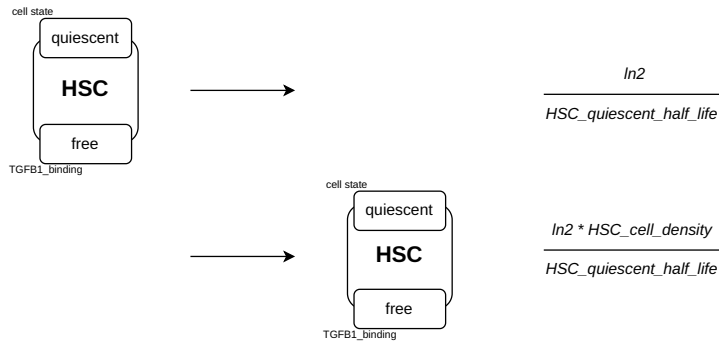


FIGURE 36 – Renouvellement des HSC quiescentes. Un agent *HSC* avec son site *cell_state* dans un état *quiescent* et son site *TGFB1_binding* dans un état *libre* disparaît (dégradation). À l'inverse, la seconde règle décrit la création de ce même agent. Le taux d'application de chaque règle est indiqué sur la droite

Fixation du TGFB1 aux cellules. Pour décrire l'interaction de TGFB1 avec les HSC, 7 règles différentes sont utilisées en fonction de l'état de la cellule. La Figure 37 montre la règle de fixation du TGFB1 aux qHSC. Seul l'état du site *cell_state* des agents diffère entre ces règles, à l'exception des règles pour MFB et *react_MFB*, où le taux de la règle est diminué. Le fait que chaque agent peut fixer le TGFB1 introduit une compétition entre les agents à chaque stimulation par le TGFB1.

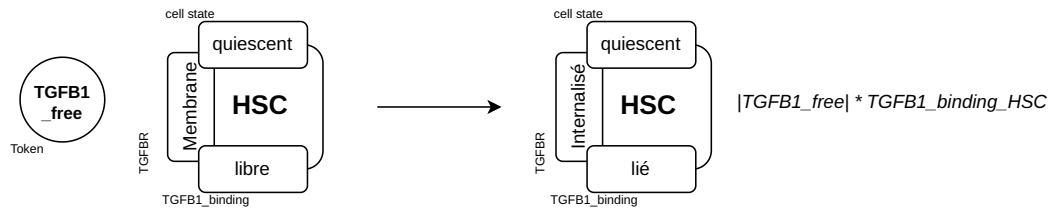


FIGURE 37 – Fixation du TGFB1 (sous forme de token *TGFB1_free*) sur un agent *HSC* avec son site *cell_state* dans un état *quiescent*, son site *TGFB1_binding* dans un état *libre* et son site *TGFB1_binding* dans un état *libre*. La valeur du token *TGFB1_free* est réduite. Pour l'agent *HSC* l'état du site *TGFB1_binding* passe à *lié* et l'état du site *TGFB1_binding* passe à *Internalisé*

Dynamique du TGFBR. La dynamique du TGFBR a été décrite précédemment par Di Guglielmo *et al.* [73] et modélisée par Zi *et al.* [293]. Dans ces modèles, la liaison du TGFB1 au TGFBR2 induit le recrutement du TGFBR1, conduisant à l'activation des voies de signalisation qui régulent les processus dépendant du TGFB1 tels que l'expression de COL1 et la prolifération cellulaire. Le trafic du complexe TGFBR1/TGFBR2 entre la membrane et le cytosol s'effectue en 30 minutes et n'est pas affecté par la liaison au TGFB1. Dans notre modèle, *TGFBR* est utilisé pour représenter le complexe TGFBR1-TGFBR2 et ne décrit pas les mécanismes de signalisation intracellulaire. Lors de la liaison avec TGFB1, le récepteur est internalisé pour transduire le signal, il peut être soit recyclé (Figure 38 haut), soit dégradé (Figure 38 bas). Pour réduire le modèle et le temps de calcul, la dynamique basale du récepteur n'a pas été incluse en l'absence de TGFB1. Pour modéliser la synthèse et la localisation du TGFBR sur la membrane, une règle similaire à celle de la dégradation a été utilisée, en changeant simplement le site *TGFBR* de l'état *dégradé* vers l'état *membrane*.

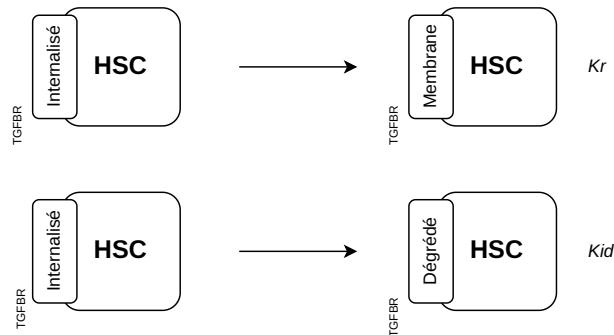


FIGURE 38 – Dégradation et recyclage du récepteur au TGFBR1

Un agent *HSC* avec son site *TGFBR* dans un état *Internalisé* aura ce site qui passera dans un état *Membrane* pour la règle de recyclage, et dans un état *dégradé* pour la règle de dégradation

Activation et différenciation des HSC. Pendant l’activation et la différenciation, les cellules ont le potentiel de proliférer et de produire du COL1 (en quantités variables selon leur état d’activation et de différenciation). Cependant, comme chaque agent ne peut entreprendre qu’une action à la fois, différentes règles ont été utilisées pour décrire l’activation, la différenciation, la prolifération et la production de COL1 des cellules.

Pour modéliser l’activation des HSC et leur différenciation en MFB, des *counters* ont été utilisés avec un intervalle de 1 à 14, correspondant aux 14 jours décrits pour l’activation *in vitro* (1 à 7 correspondant au passage de qHSC à aHSC ; 8 à 14 correspondant au passage de aHSC à MFB). D’une part, deux règles décrivent le changement d’état du site *cell_state* (qHSC vers aHSC et aHSC vers MFB) et d’autre part, deux règles décrivent la dynamique étape par étape en augmentant le *counter intermediate_step* et en diminuant le *counter control_counter*.

L’augmentation du *counter intermediate_state* permet de suivre le processus d’activation et de différenciation. La diminution du *counter control_counter* permet de s’assurer que la valeur du *counter intermediate_state* reste bien dans l’intervalle de 1 à 14 (limite du langage Kappa, détailler dans la Discussion III). La Figure 39 décrit la dynamique de l’activation, définie en 7 étapes, correspondant au sept jours du passage de qHSC à aHSC, cela avec l’implémentation d’une seule règle. Pour le processus de réactivation, des règles similaires à celles du processus d’activation ont été utilisées.

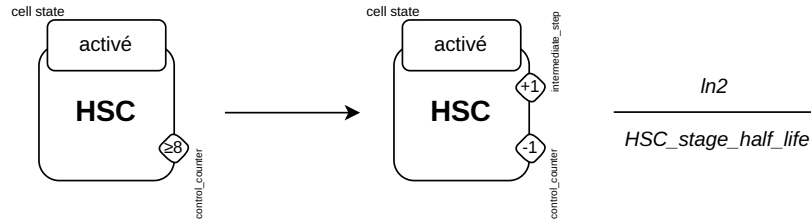


FIGURE 39 – Processus d’activation des HSC, la règle spécifie que tant qu’un agent *HSC*, dont le site *cell_state* est dans l’état *activé*, a son *counter control_counter* avec une valeur inférieur ou égal à huit alors, la valeur de son *counter control_counter* sera décrémenté de un et celle de son *counter intermediate_step* sera incrémenté de un. Cette simple règle permet de décrire les sept étapes d’activation

Prolifération des cellules. Pour chaque valeur de *counters*, une règle de prolifération a été définie. Si un agent *HSC* prolifère, tous les sites de l’agent nouvellement créés auront la même valeur que ceux de l’agent mère, à l’exception de TGFBR. La Figure 40 montre un exemple de règles de prolifération pour l’agent *HSC* dont le site *cell_state* est dans l’état *activé* et son *counter intermediate_state* égal à 7. Les étapes d’activation ou de différenciation n’affectent pas le taux de prolifération des agents et *aHSC* et *react_HSC* ont des règles supplémentaires pour la prolifération induite par le TGFBR1. Il existe 30 règles pour décrire le processus de prolifération. La valeur de *state*, *intermediate_step* et *control_counter* est différente pour chaque règle afin de permettre à toutes les cellules de proliférer. Parmi ces règles, 11 permettent aux agents dont le site *cell_state* est dans l’état *aHSC* ou *react_HSC* de proliférer en réponse au TGFBR1 en exigeant l’état *TGFBR1_binding{bound}*, puis en le transformant en *TGFBR1_binding{free}* après l’application de la règle.

Production et dégradation du Collagène 1. Le dépôt de COL1 durant la fibrose hépatique et la réversion résulte de l’équilibre entre sa production et sa dégradation. Lors de la stimulation par le TGFBR1, les HSC activés et les MFB sont la principale source cellulaire de COL1, tandis que les *apop_sene_MFB* contribuent à la dégradation du COL1 lors de la réversion de la fibrose [156, 223]. Pour surmonter les mécanismes moléculaires complexes impliqués dans la régulation de la dégradation de COL1, le collagène 1 a été décrit à l’aide de deux *tokens* différents : l’un pour le remodelage élevé (*COL1_remodelage_high*)

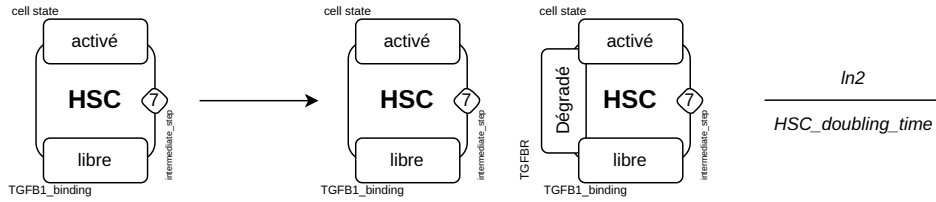


FIGURE 40 – Exemple de règles de prolifération pour un agent *HSC* dont le site *cell_state* est dans l'état *activé* et son *counter control_counter* égal à 7

et l'autre pour le remodelage faible (*COL1_remodelage_low*). Ces deux *tokens* permettent de décrire le faible taux de dégradation de COL1 pendant l'activation des HSC, tout en permettant un taux de dégradation élevé pendant le processus de réversion. La dynamique de COL1 est décrite à l'aide de 12 règles : 8 pour la production de *COL1_remodeling_high* par aHSC, MFB, react_HSC et react_MFB en présence et en absence de TGFB1 ; 2 pour la dégradation de chaque *tokens* (*COL1_remodeling_low* étant dégradé à un rythme plus lent que *COL1_remodeling_high*) et 2 pour permettre aux deux *tokens* d'être liés (voir Figure 41 haut). La transformation de *COL1_remodeling_high* en *COL1_remodeling_low* dépend du nombre de cellules productrices de COL1 (aHSC, react_HSC, MFB et react_MFB), toutes caractérisées par l'expression de la protéine α -actine musculaire lisse (α SMA).

Plus les agents cellulaires sont activés (ou réactivés) et différenciés en MFB (ou en react_MFB), plus la production de *COL1_remodeling_high* est importante. Pour modéliser cela, nous avons profité de la sémantique de Kappa en utilisant des *counters* dans le taux d'application des règles, au lieu de déclarer une règle pour chaque étape.

Au total, huit règles décrivent la production de COL1. L'état et la variable définissant la quantité de *COL1_remodeling_high* ont été modifiés en fonction des cellules (voir Figure 41 bas). Parmi ces huit règles, quatre correspondent aux conditions en absence de TGFB1 et quatre sont en présence de TGFB1. Pour ces quatre dernières, la cellule doit avoir *TGFB1_binding{bound}* et qui sera transformé en *TGFB1_binding{free}* après la production de *COL1_remodeling_high*.

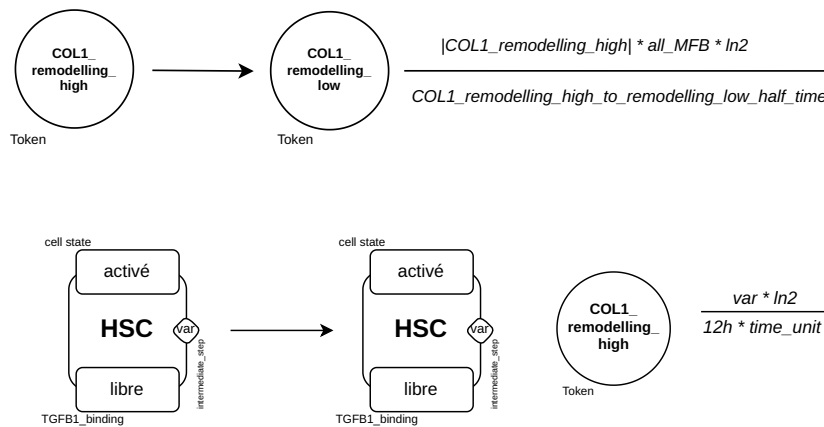
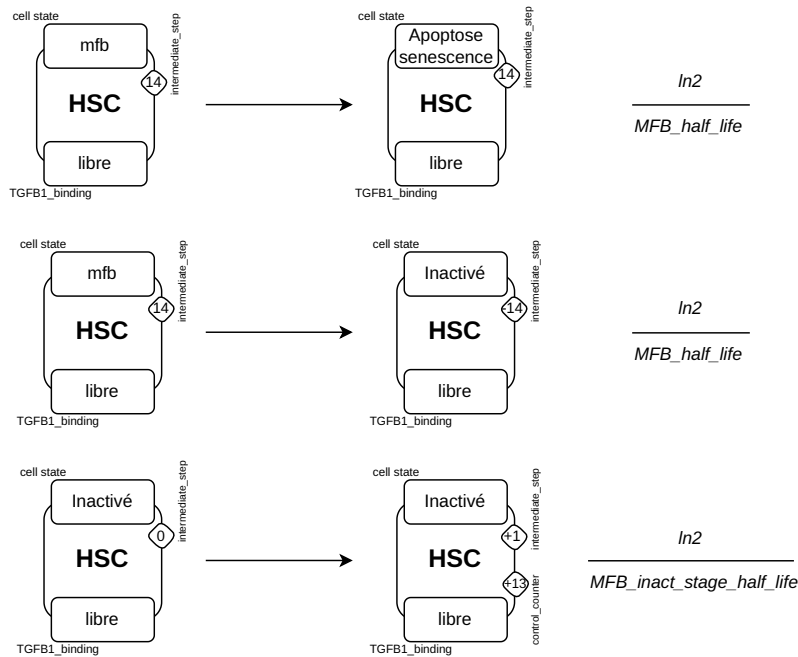


FIGURE 41 – Exemple des règles impliquées dans la dynamique du Collagène 1. Stabilisation du collagène, la valeur du *token* *COL1_remodeling_high* est diminuée et celle de *COL1_remodeling_high* est augmentée; Production du collagène par les HSC activées, le taux d'application de la règle est proportionnel à la valeur du *counter intermediate_step*. Cela permet aux cellules qui sont plus avancées dans le processus d'activation de produire plus de *COL1_remodeling_high* que celles qui viennent juste d'être activées.

Inactivation des cellules. Quand l'agression est interrompue, c'est-à-dire en absence de signal TGFB1, les MFB et *react_MFB* sont éliminés par les processus d'apoptose et de sénescence (voir Figure 42 haut). Une règle supplémentaire décrit le processus d'élimination des *apop_sene_MFB* (non présentée). La dynamique du processus d'inactivation est décrite en deux étapes à l'aide de deux règles (voir Figure 42 bas). La première règle fait passer les MFB à l'état de cellule inactivée (*iHSC*) et décrémente la valeur du *counters* nommé *intermediate_step* à 0. La seconde règle incrémente la valeur de ce même *counters* de 1 et incrémente la valeur du *counters* de *control_counters* de 13.

FIGURE 42 – Règles d'élimination des MFB par : apoptose/sénescence (haut) et inactivation (bas). La règle d'apoptose/sénescence consiste à faire passer le site *cellular_state* d'un agent *HSC* d'un état *mf* à *apoptose_senescence*. Pour l'inactivation, deux règles sont nécessaires, la première fait passer le site *cellular_state* d'un agent *MFB* d'un état *mf* à *inactivé* ainsi que réduire la valeur du *counter intermediate_step* de 14. La seconde règle augmente les valeurs de *control_counters* de 13 et de *intermediate_step* de 1.



6.3 Calibration et estimation des paramètres

La calibration des paramètres a été réalisée sous certaines contraintes, comme les données où l'utilisation de l'algorithme SSA induit par définition l'utilisation de données quantitatives discrètes. Le taux d'application des règles est tiré selon une loi exponentielle dont le paramètre est égal à la somme des propensités de tous les événements potentiels dans l'état du système. Les taux des règles doivent donc être définis comme la valeur $T_{1/2}$ du processus biologique qu'ils décrivent.

Pour tous les processus biologiques inclus dans notre modèle, un total de 43 paramètres a été utilisé, 8 paramètres ont été obtenus directement à partir de la littérature, 10 ont été calculés pour obtenir le $T_{1/2}$ de leur processus et 22 paramètres inconnus ont été estimés pour s'adapter aux observations biologiques en comparant les simulations du modèle avec les données expérimental par le lancement de centaines de simulations. Chaque simulation prenant ~ 1494.04 s et utilisant 1 033 MB de RAM (Dell Inc. Latitude 7410; 8 x 1.8 GHz; 16 GB RAM). Le comportement de certaines cellules étant inconnu (iHSC et MFB réactivés), des hypothèses ont été faites pour s'adapter aux données biologiques, ce qui nous a conduit à générer trois modèles basés sur les hypothèses suivantes : 1) les

MFB réactivés ne peuvent pas être inactivés (modèle *reactMFB-wo-inactivation*), 2) les MFB réactivés peuvent être inactivés (modèle *reactMFB-with-inactivation*), 3) les iHSC peuvent se transformer en qHSC (*iHSC-reversion-to-qHSC*, dans ce dernier modèle les MFB peuvent s'inactiver).

Principes généraux

Dans le foie sain, les HSC se trouvent dans les espaces de Disse entre l'endothélium sinusoidal et les microvillosités hépatocytaires et représentent 5 à 8 % des cellules totales. Sur la base des données de la littérature [267, 79], une densité de 4 HSC par espace de Disse a été estimée et des simulations utilisant 10 000 HSC quiescentes ont été effectuées. Les paramètres ont été extraits ou estimés à partir des données expérimentales *in vivo* et *in vitro* et sont résumées dans le Tableau 4. Nous détaillons ici brièvement le raisonnement qui a conduit à l'estimation de ces paramètres, communs aux trois modèles. Ces paramètres ont été estimés en procédant processus biologique par processus biologique.

Une attention particulière a été portée à la calibration des 17 paramètres inconnus, en raison des informations non documentées les concernant, mais aussi en raison de l'interdépendance entre les paramètres et l'approche stochastique. À certains moments de la calibration du modèle, certains paramètres ont été estimés empiriquement par groupe (par exemple, la production, la dégradation et la stabilisation de COL1), ce qui a entraîné un grand nombre de simulations avant de trouver l'ensemble correct de paramètres. Dans le Tableau 4, seuls les paramètres du modèle validé *reactMFB-with-inactivation* sont indiqués.

Paramètres des règles d'activation et de différenciation Le processus d'activation a été largement documenté à l'aide de cultures primaires des HSC. Les qHSC isolées s'activent spontanément lorsqu'elles sont cultivées sur du plastique et la transformation complète des qHSC en MFB s'étale sur 14 jours, avec un taux de prolifération plus rapide pour les aHSC que pour les MFB [16, 18]. Sur la base de cette dynamique, le modèle a été calibré avec 14 étapes, et le taux de prolifération des aHSC est deux fois plus élevé que celui des MFB. L'activation des HSC ce produit *in vivo* sous l'effet d'un signal inflammatoire qui implique principalement le TGFB1, celui-ci favorise la prolifération et la production de COL1. Alors que les qHSC sont très sensibles au TGFB1, les MFB le sont beaucoup moins [75, 18]. Nous avons donc choisi de ne pas mettre de règle pour la prolifération des MFB dépendante du TGFB1 et avons rendu la production de collagène dépendante du

TGFB1 cinq fois plus faible dans les MFB que dans les aHSC. Ces études de simulations nous ont conduites à l'estimation de huit paramètres (22-29) du Tableau 4.

Paramètre des règles d'impact du TGFB1. La cinétique de liaison du TGFB1 aux HSC a été décrite *In vitro* [96] et un taux de liaison plus lent pour les MFB a été utilisé en raison de la diminution du nombre de récepteurs [75]. Les paramètres de la dynamique des récepteurs sont ceux décrits par Vilar *et al.* [262] et le demi-temps nécessaire au TGFB1 pour induire un signal initiant le processus d'activation a été calculé à l'aide du modèle de Zi *et al.* [293]. Cette étape d'initialisation est nécessaire pour activer la machinerie cellulaire avant que le processus d'activation ne commence. Notons que ce temps d'initialisation a été divisé par 3,5 pour les iHSC, car ces cellules sont beaucoup plus réactives que les qHSC en réponse au TGFB1 [143, 256, 80]. De la même façon, les react_HSC peuvent proliférer immédiatement après avoir été réactivées, alors que la prolifération des aHSC commence à la troisième étape du processus d'activation. Ce choix de modélisation a été fait pour simuler le temps nécessaire aux qHSC pour initier la machinerie cellulaire de l'activation *in vivo* [183] alors que les react_HSC, en raison de leur historique d'activation, prolifèrent dès la première étape de leur réactivation. Ces observations nous ont conduites à l'estimation de huit paramètres (1-8) du Tableau 4.

Paramètre des règles d'auto-renouvellement des qHSC et d'élimination des iHSC. En l'absence d'information sur le taux d'auto-renouvellement des qHSC et la demi-vie des iHSC, ces paramètres ont été estimés à partir des données expérimentales de Kisseleva *et al.* [143]. Ces paramètres correspondent aux paramètres 14 et 15 du Tableau 4.

Paramètres des règles d'élimination des MFB et react_MFB. De même, la demi-vie des MFB et des react_MFB n'a jamais été documentée. Nous l'avons estimée sur la base des observations de l'accumulation de ces cellules pendant la stimulation par le TGFB1 et de leur disparition 1 à 2 mois après les derniers apports de TGFB1. Cette accumulation représente un changement de 10,6 à 14,3 % du nombre total de HSC après deux mois de traitement CCl4, comme l'indique Kisseleva *et al.*, où les qHSC et les cellules α -SMA représentent respectivement 10,6 % ($\pm 0,8$) et 14,3 % ($\pm 1,5$) de la population totale des cellules hépatiques. Ces observations nous ont conduites à l'estimation de deux paramètres (16 et 17) du Tableau 4.

Paramètre des règles dynamiques du collagène 1. Un point critique du modèle a été le paramétrage de la dynamique du collagène I (COL1). Les règles régissant la production, la dégradation, la stabilisation et la déstabilisation de *COL1_high_remodeling* et de *COL1_low_remodeling* ont été calibrées à l'aide des données de Kisseleva *et al.* [143]. Comme le COL1 est quantifié expérimentalement en tant que dépôts résultant d'un remodelage global, le COL1 total (COL1_tot) est défini comme la somme de *COL1_remodeling_high* et *COL1_remodeling_low*. Le COL1_tot a été multiplié par 12 et 14 après respectivement 1 et 2 mois de stimulation par le TGFβ1 (deux stimulations par semaine). Les aHSC et react_HSC ont produit 10 fois moins de *COL1_remodeling_high* que les MFB et react_MFB. Cependant, la production de COL1 dépendante du TGFβ1 a été multipliée par 10 pour les aHSC et react_HSC et seulement par 2 pour les MFB et react_MFB, puisque les aHSC et react_HSC sont plus sensibles au TGFβ1 [143, 256, 80]. En absence de TGFβ1, le modèle a été calibré pour observer une diminution rapide de COL1_tot au cours du premier mois, suivie d'un retour progressif à sa valeur initiale environ six mois plus tard [143]. Ces observations nous ont conduites à l'estimation de 10 paramètres (30-39) du Tableau 4.

Paramètre spécifique aux modèles *reactMFB-wo-inactivation* et *reactMFB-with-inactivation*

Lors de l'arrêt de l'agression, *i.e.* en l'absence de stimulation du TGFβ1 dans le modèle, 45 % des MFB sont inactivés et les 55 % restants sont éliminés par apoptose et par les voies de la sénescence [164, 256]. Troger *et al.* ont émis l'hypothèse que les react_MFB peuvent subir de nouveaux cycles d'inactivation/réactivation, mais que le nombre de cycles devrait être limité pour une cellule donnée [256]. Sur la base de cette hypothèse, deux familles de modèles ont été créées : le modèle *reactMFB-wo-inactivation* dans lequel les react_MFB ne peuvent pas être inactivés et le modèle *reactMFB-with-inactivation* dans lequel l'inactivation est possible. Sans connaître le pourcentage de MFB réactivés pouvant être inactivés, une série de modèles a été créée pour la dernière famille en faisant varier la valeur de 1 à 50 %. Ces simulations nous ont conduites à l'estimation de deux paramètres (18 et 19) du Tableau 4.

Paramètre spécifique au modèle *iHSC-reversion-to-qHSC*

En utilisant le modèle murin de fibrose hépatique induite par le CCl₄, Kisseleva *et al.* [143] ont évalué la quantité de iHSC par rapport aux qHSC après l'arrêt de l'injection de CCl₄. Ils ont rapporté que les qHSC et les iHSC sont en proportion égale après un mois de récupération, et que la somme des qHSC et des iHSC après six mois de récupération est égale au nombre initial de qHSC. Cependant, il n'y a pas d'information sur le comportement des iHSC en l'absence de CCl₄ (i.e. de TGFβ1 dans le modèle). Nous avons testé deux hypothèses : Leur élimination par des voies de dégradation ou leur retour à l'état de quiescence. Dans le modèle *iHSC-reversion-to-qHSC*, nous avons fait varier le pourcentage de iHSC qui retournent à l'état quiescent entre 10 à 100 %. Ces études de simulations nous ont conduites à l'estimation de deux paramètres (non montrés).

TABLEAU 4 – **Tableau des paramètres.** *U.A* : Unité arbitraire correspondant au pourcentage des surfaces de marquage histochimique permettant de quantifier le dépôt de COL1. *Calculé* : paramètre calculé à partir des valeurs trouvées dans la bibliographie. *Adapté aux données expérimentales* : paramètres pour lesquels les valeurs sont estimées empiriquement afin d'adapter le modèle aux observations. Tous les paramètres décrivant une durée sont calculés comme le demi-temps de leurs réactions. L'utilisation de la loi exponentielle permet de déterminer la dynamique moyenne des réactions correspondant aux observations expérimentales [70] ;

6.3. Calibration et estimation des paramètres

Paramètre	Symbole/Name	Value	Référence
1 - Taux de recyclage	Kr	4.5 min-1	[262]
2 - Taux de production	Pr	1.2 min-1	[262]
3 - Taux de dégradation induite par le ligand	Kid	0.6 min-1	[262]
4 - Temps de demi-vie	TGFB1_half_life	3 min-1	Calculé à partir de [268, 118]
5 - Demi-temps de fixation au récepteur des HSC	TGFB1_binding_HSC	0.001 min-1	Calculé à partir de [96]
6 - Demit-temps de fixation au récepteur des MFB	TGFB1_binding_MFB	0.004 min-1	Calculé à partir de [96]
7 - Demi-temps requis par TGFB1 pour induire le signal initiant le processus d'activation	HSC_half_activation	0.15 heure-1	Calculé à partir de [293]
8 - Demi-temps requis par TGFB1 pour induire le signal initiant le processus de ré-activation	react_HSC_half_activation	0.043 heure-1	Calculé à partir de [293, 143, 256]
9 - Nombre de TGFBR par cellules	TGFB1_factor	10 000	[177, 138, 169]
10 - Quantité saturante de TGFB1	TGFB1_input	60 000/cellules	[293]
11 - Nombre d'inputs de TGFB1	nb_iteration	16	[143]
12 - Temps entre chaque input de TGFB1	interval	84 heures	[143]
13 - Nombre de qHSC à T0	HSC_cell_density	10 000	
14 - Temps de demi-vie des qHSC	HSC_quiescent_half_life	75 jours-1	Adapté aux données expérimentales [143]
15 - Temps de demi-vie des iHSC	iHSC_half_life	2 années-1	Adapté aux données expérimentales [143]
16 - Temps de demi-vie des MFB	MFB_half_life	3.54 jours-1	Adapté aux données expérimentales [143]
17 - Temps de demi-vie des react_MFB	react_MFB_half_life	3.54 jours-1	Adapté aux données expérimentales [143]
18 - Pourcentage de MFB inactivés	inactivation_percentage	45 %	[143]
19 - Pourcentage de react_MFB inactivés	inactivation_percentage_react	5 %	Adapté aux données expérimentales [143]
20 - Demi temps de l'étape d'inactivation	MFB_inact_stage_half_life	0.75 jour-1	Calculé à partir de [256, 80]
21 - Temps de demi-vie des apop_sene_MFB	apop_sene_MFB_half_life	3.6 heure-1	
22 - Temps prolifération des aHSC	HSC_doubling_time	15.5 jours-1	Adapté aux données expérimentales [143]
23 - Temps prolifération des react_HSC	react_HSC_doubling_time	4.43 jours-1	Adapté aux données expérimentales [143, 256, 80]
24 - Temps prolifération des MFB	MFB_doubling_time	31 jours-1	Adapté aux données expérimentales [143, 75, 18]
25 - Temps prolifération des react_MFB	react_MFB_doubling_time	8.86 jours-1	Adapté aux données expérimentales [143, 256, 80]
26 - Demi temps de l'étape d'activation des aHSC	HSC_stage_half_life	17 heure-1	Calculé à partir de [16, 18]
27 - Demi temps de l'étape de différenciation des MFB	MFB_stage_half_life	17 heure-1	Calculé à partir de [16, 18]
28 - Demi temps de l'étape d'activation des react_HSC réactivation	react_HSC_stage_half_life	4.86 heure-1	Calculé à partir de [143, 256, 80]
29 - Demi temps de l'étape de différenciation react_MFB	react_MFB_stage_half_life	4.86 heure-1	Calculé à partir de [143, 256, 80]
30 - % Aire dans le foie sain	col1_healthy_concentration	0.5 A.U	[143]
31 - Production de high remodeling COL1 par les aHSC	COL1_synthesis_aHSC	0.019 A.U / 12h	Adapté aux données expérimentales [143]
32 - Production de high remodeling COL1 par les MFB	COL1_synthesis_MFB	0.19 A.U / 12h	Adapté aux données expérimentales [143]
33 - Production de high remodeling COL1 par les react_HSC	COL1_synthesis_react_HSC	0.0665 A.U / 12h	Adapté aux données expérimentales [143]
34 - Production de high remodeling COL1 par les react_MFB	COL1_synthesis_react_MFB	0.665 A.U / 12h	Adapté aux données expérimentales [143]
35 - Production de high remodeling COL1 par les aHSC induite par le TGFB1	COL1_synthesis_aHSC_TGFB1	0.19 A.U / 12h	Adapté aux données expérimentales [143]
36 - Production de high remodeling COL1 par les MFB induite par le TGFB1	COL1_synthesis_MFB_TGFB1	0.38 A.U / 12h	Adapté aux données expérimentales [143]
37 - Production de high remodeling COL1 par les react_HSC induite par le TGFB1	COL1_synthesis_react_HSC_TGFB1	0.665 A.U / 12h	Adapté aux données expérimentales [143]
38 - Production de high remodeling COL1 par les react_MFB induite par le TGFB1	COL1_synthesis_react_MFB_TGFB1	1.33 A.U / 12h	Adapté aux données expérimentales [143]
39 - Quantité de high remodeling COL1 and low remodeling COL1 dégradée	COL1_deg	0.0002 A.U	Adapté aux données expérimentales [143]
40 - Temps de demi-vie high remodeling COL1	COL1_remodeling_high_half_life	17.28 s-1	Adapté aux données expérimentales [143]
41 - Temps de demi-vie low remodeling COL1	COL1_remodeling_low_half_life	15.84 min-1	Adapté aux données expérimentales [143]
42 - Temps de transition du High remodeling COL1 vers low remodeling COL1	COL1_remodeling_high_to_remodeling_low_half_time	20 jour-1	Adapté aux données expérimentales [143]
43 - Temps de transition du Low remodeling COL1 vers COL1 High remodeling COL1	COL1_remodeling_low_to_remodeling_high_half_time	2.16 min-1	Adapté aux données expérimentales [143]

6.4 L'inactivation des myofibroblastes réactivés est essentielle pour maintenir l'accumulation de collagène 1 au cours d'une agression chronique du foie

Pour identifier le ou les modèles qui correspondent aux données expérimentales de Kisseleva *et al.* [143], nous avons effectué des études de simulations en utilisant le protocole de stimulation décrit par les auteurs, c'est-à-dire des injections de CCl4 (TGFB1 dans le modèle) deux fois par semaine pendant huit semaines et une période de réversion globale de six mois. Les observations de Kisseleva *et al.*[143] ont montré une augmentation de $\sim 1,43$ -fois du nombre de cellules exprimant la protéine α -SMA et de ~ 14 -fois du pourcentage de l'aire de COL1, à 8 semaines, avec un retour aux valeurs initiales à six mois après l'arrêt de l'agression. De plus, les auteurs ont évalué les rapports iHSC/qHSC et ont identifié une distribution de 50/50 à un mois de réversion et un nombre total d'HSC similaire aux conditions initiales à six mois de réversion. Comme le montre la Figure 43 quels que soient les modèles, la majorité des qHSC (99 %) ont été immédiatement activées lors de la première entrée de TGFB1, un petit nombre (2 ± 1 %) de qHSC a persisté pendant la période de stimulation. L'activation des qHSC a conduit à la production de aHSC qui se sont différenciées en MFB. Ces cellules expriment la protéine α -SMA et nous avons suivi les cellules positives pour α -SMA en additionnant les occurrences de aHSC, de MFB, de `react_HSC` et de `react_MFB`. Il est intéressant de noter que la proportion entre les cellules initialement activées et les cellules réactivées change au cours des stimulations successives. Au cours de l'apport initial de TGFB1, les qHSC sont la principale source de cellules α -SMA, comme le montrent les courbes pour les MFB et aHSC. Après environ un mois, les iHSC deviennent la source principale, générant plus de cellules réactivées (MFB et aHSC réactivées) qui deviennent les principales cellules productrices de COL1. Cela peut s'expliquer par la plus grande affinité des cellules inactivées pour le TGFB1 par rapport aux qHSC [80]. Notons que même si le nombre de qHSC et iHSC est faible, elles peuvent être activées ou réactivées par le TGFB1 au cours des stimulations répétées durant les deux mois. Conformément à cette observation, Rosenthal *et al.* [214] a récemment suggéré que les iHSC persistent à un faible niveau même pendant la phase d'activation.

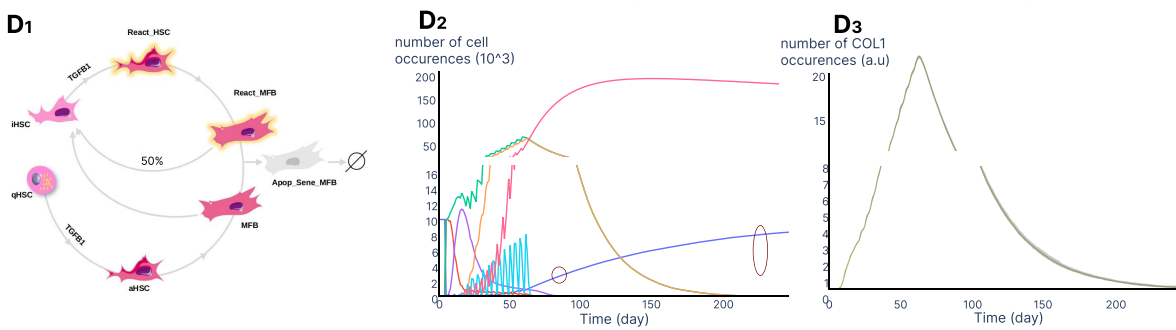
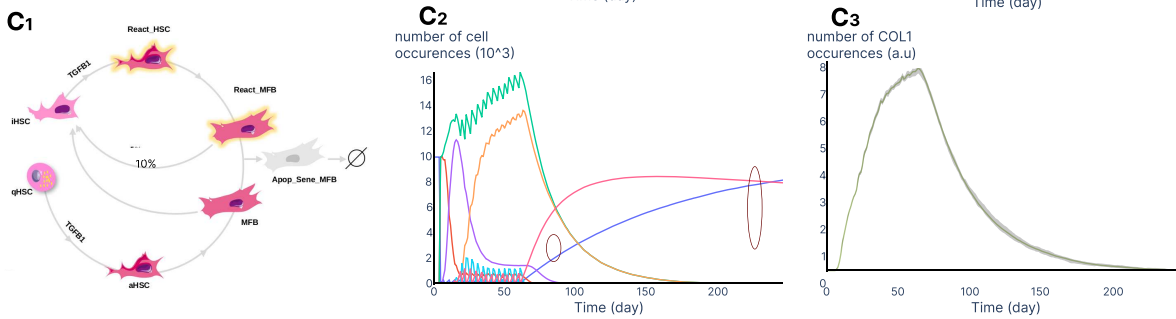
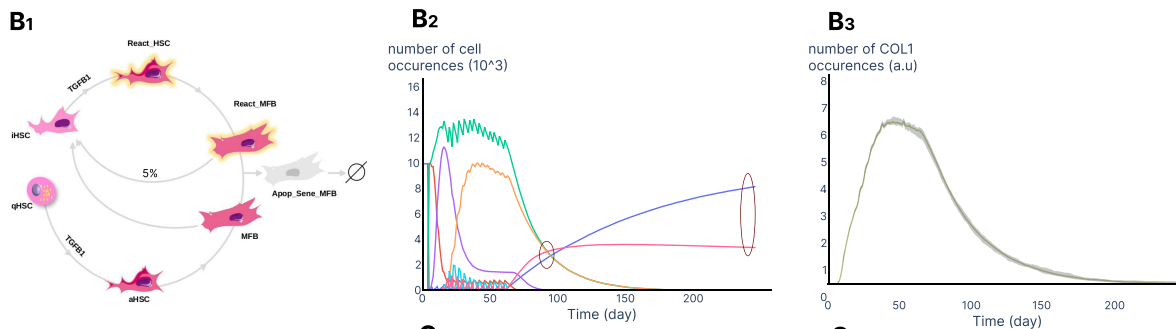
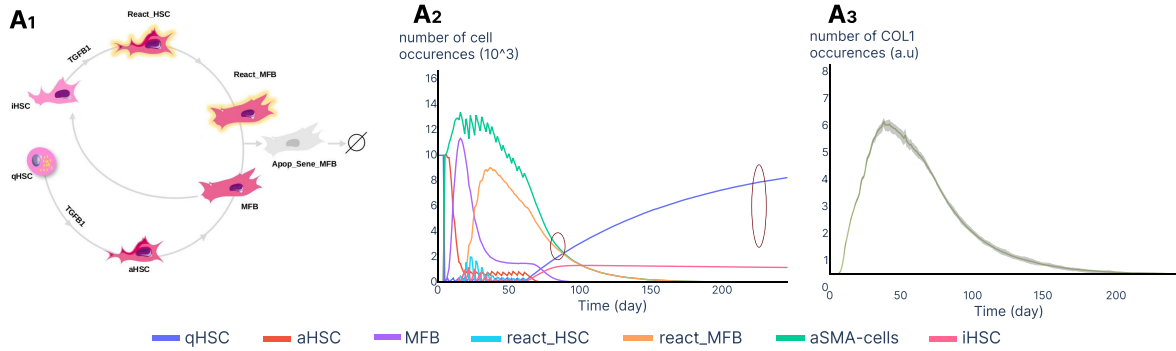
La comparaison des modèles a montré que l'existence d'une boucle d'inactivation des `react_MFB` affecte grandement le comportement du modèle. Lorsque l'inactivation des

6.4. *L'inactivation des myofibroblastes réactivés est essentielle pour maintenir l'accumulation de collagène 1 au cours d'une agression chronique du foie*

react_MFB n'est pas permise (modèle *reactMFB-wo-inactivation*) (voir Figure 43 (A)), nous avons observé une quantité de cellules α -SMA et de COL1 inférieure à celle observée expérimentalement à huit semaines, les quantités diminuant même avant l'arrêt de la stimulation. Inversement, lorsque l'inactivation de react_MFB est autorisée (modèle *reactMFB-with-inactivation*) Figure 43 (B, C, D) nous avons observé une accumulation de cellules α -SMA et de COL1 qui variait en fonction du pourcentage d'inactivation de react_MFB. Alors que cinq pour cent rendent le modèle conforme aux observations expérimentales (cercle rouge dans la Figure 43), l'augmentation du pourcentage à 10 et 50 % a conduit à des quantités plus élevées de cellules α -SMA (1,6 fois et 5 fois, respectivement) et de COL1 (16 et 40, respectivement) à huit semaines. Seul le modèle avec une inactivation de 5 % des react_MFB a conduit à une distribution 50/50 du ratio iHSC/qHSC à un mois, et à un retour au nombre total initial de cellules après six mois de réversion, comme le décrivent les données expérimentales.

FIGURE 43 – Analyses temporelles des cellules et de COL1 dans les modèles sans ou avec inactivation des react_MFB Une série de simulations a été réalisée en utilisant les conditions de stimulation de Kisseleva *et al.* [143]. Les paramètres du TGFB1 étaient de 10 000 molécules par cellule, 16 stimulations par TGFB1 (deux fois par semaine) pendant deux mois. **A)** modèle *reactMFB-wo-inactivation*. **B)** modèle *reactMFB-with-inactivation* avec 5% de react_MFB inactivés. **C)** modèle *reactMFB-with-inactivation* avec 10% de react_MFB inactivés. **D)** modèle *reactMFB-with-inactivation* avec 50% de react_MFB inactivés. La stimulation des cellules avec le TGFB1 commence le jour 4 pour permettre au modèle de s'équilibrer. Pour **A2**, **B2**, **C2** ou **D2** il s'agit de la variation du nombre de qHSC, aHSC, MFB, iHSC, react_HSC, react_MFB et apop_sene_MFB. Les cellules à α -SMA sont la somme des aHSC, MFB, react_HSC et react_MFB. Pour **A3**, **B3**, **C3** ou **D3** il s'agit de la variation du nombre d'occurrences de COL1 (unité arbitraire).

Partie II, Chapitre 6 – Modélisation de la dynamique des cellules étoilées hépatiques, introduction des tokens et des counters



6.4. L'inactivation des myofibroblastes réactivés est essentielle pour maintenir l'accumulation de collagène 1 au cours d'une agression chronique du foie

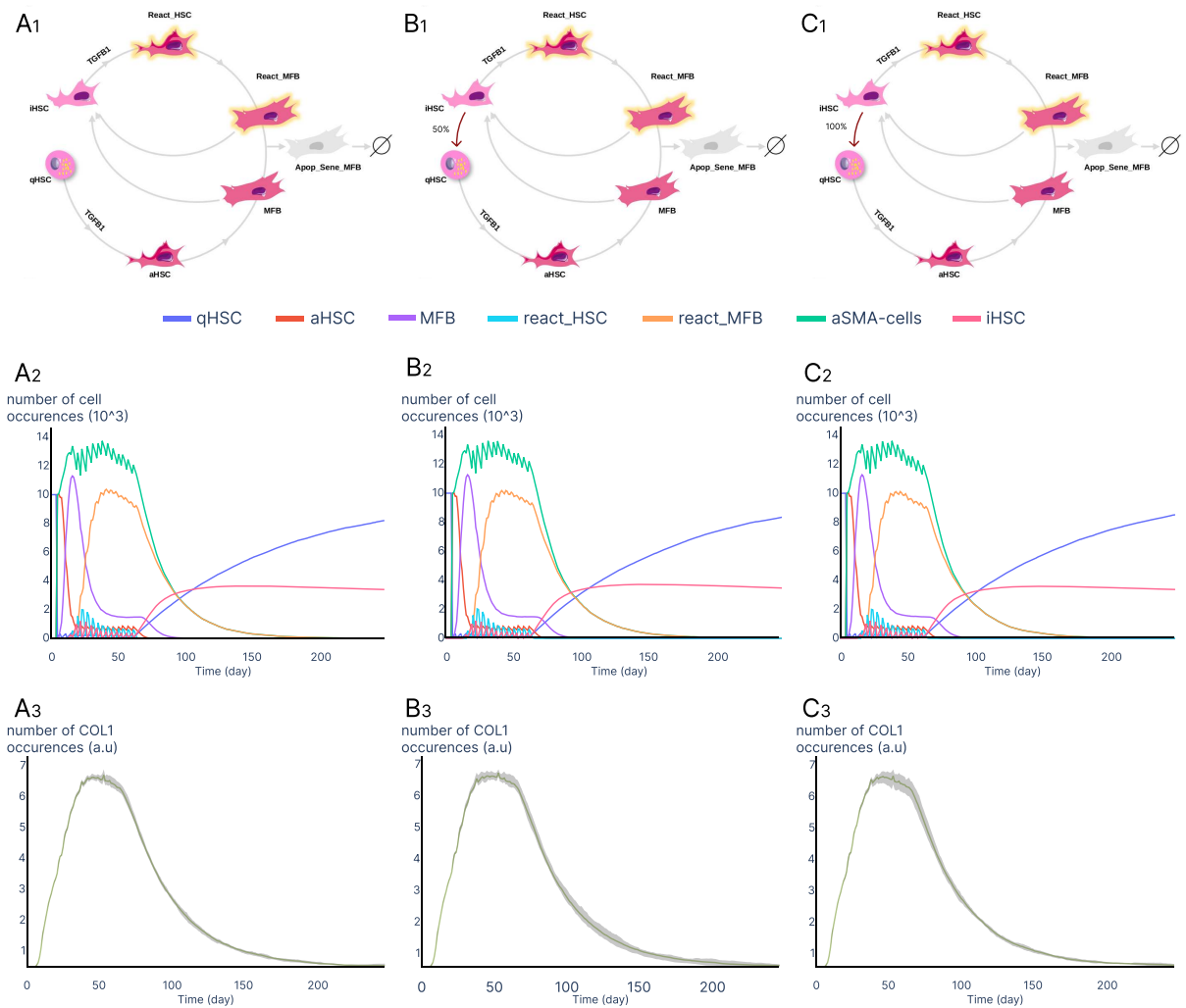


FIGURE 44 – Analyses temporelles des cellules et de COL1 dans les modèles sans ou avec réversion des iHSC vers un état quiescent Une série de simulations a été réalisée en utilisant les conditions de stimulation de [143]. Les paramètres du TGFβ1 étaient de 10 000 molécules par cellule, 16 entrées (deux fois par semaine) pendant deux mois. **A)** modèle *reactMFB-with-inactivation*. **B)** modèle *iHSC-reversion-to-qHSC* avec 50 % de iHSC qui retournent vers un état quiescent. **C)** modèle *iHSC-reversion-to-qHSC* avec 100 % de iHSC qui retournent vers un état quiescent. La stimulation des cellules avec le TGFβ1 commence le jour 4 pour permettre au modèle de s'équilibrer. Pour **A,B,C (2)** est la variation du nombre de qHSC, aHSC, MFB, iHSC, react_HSC, react_MFB et apop_sene_MFB. Les cellules à α-SMA sont la somme des aHSC, MFB, react_HSC et react_MFB. **(3)** Variation du nombre d'occurrences de COL1 (unité arbitraire).

Sur la base du modèle *reactMFB-with-inactivation* paramétré avec 5 % d'inactivation de `react_MFB`, nous avons ensuite évalué l'impact d'une réversion potentielle des iHSC vers un état de quiescence. Comme décrit dans ces nouveaux modèles *iHSC-reversion-to-qHSC*, nous avons observé que quel que soit le pourcentage de cellules qui reversent, cela n'affectait pas le comportement global des autres composants du modèle (voir Figure 44).

Dans l'ensemble, nos données ont démontré que l'inactivation de `react_MFB` semble importante pour soutenir l'accumulation de COL1 en maintenant le niveau des cellules α -SMA, validant ainsi le modèle *reactMFB-with-inactivation*. Les analyses de simulation ont identifié un rapport de 5 % entre l'inactivation des `react_MFB` et les voies de l'apoptose et de la sénescence. Notons que la probabilité qu'un nombre de processus d'inactivation et de réactivation se produisent est égale à

$$\left(\frac{1}{20}\right)^n$$

où n est le nombre de processus d'inactivation successifs. En ce qui concerne le comportement des iHSC, nos observations suggèrent que la réversion complète des iHSC vers l'état de quiescence est une voie possible d'élimination des iHSC en l'absence de TGFB1. Le modèle *iHSC-reversion-to-qHSC* étant insensible à la variation de la réversion des iHSC, nous avons choisi le modèle *reactMFB-with-inactivation* pour tester la sensibilité du modèle au TGFB1.

6.5 L'accumulation de collagène I est sensible à la dynamique de la stimulation par le TGFB1

Nos modèles sont calibrés à l'aide de données expérimentales provenant d'un modèle murin de fibrose induite par le CCl4, un modèle expérimental largement utilisé pour étudier la fibrose et la réparation [225]. Bien que les protocoles varient quelque peu en termes de nombre, de fréquence et de modalité des injections de CCl4, ces modèles expérimentaux induisent une fibrose hépatique similaire. D'autres modèles d'induction de la fibrose hépatique chez le rongeur ont été développés, notamment : modèle par injection d'autres agents chimiques (comme le thioacétamide (TAA) et la diméthylnitrosamine (DMN)), modèles utilisant des régimes riches en graisses et en cholestérol, modèle de ligature des voies biliaires [281, 284]. Quelle que soit l'étiologie, les dommages induits conduisent à une activation des HSC, cependant la dynamique du signal d'agression est variable sui-

vant le modèle. Il peut être très rapide dans un modèle induit par le CCl4 et plus lent dans un modèle induit par un régime gras. Pour explorer la sensibilité de notre modèle *reactMFB-with-inactivation*, nous avons fait varier la concentration de TGFB1 ainsi que le nombre et la périodicité des stimulations. Comme le montre la Figure 45, le modèle est très sensible à la quantité de TGFB1, mais cette sensibilité varie en fonction du nombre d'entrées et de la périodicité, en particulier pour les valeurs de TGFB1 inférieures à 10 000 molécules par cellule. Au-delà, le nombre de molécules dépasse le nombre de récepteurs par cellule et le modèle devient saturé, comme le montre la superposition des courbes de cinétiques à 10 000 (vert) et 100 000 (jaune) molécules par cellule.

À des niveaux inférieurs de TGFB1, l'accumulation de collagène est beaucoup plus subtile. Lorsque les entrées sont suffisamment espacées, les niveaux de collagène reviennent à la ligne de base et de façon remarquable, nous avons observé une induction beaucoup plus forte lors de la deuxième stimulation, en particulier à des doses de 10 000 et 100 000 molécules par cellule (Figure 45 A, périodicité de 60 et 120 jours). Cette induction du collagène 1 est moins importante aux faibles doses, mais l'augmentation du nombre de stimulations par le TGFB1 (Figure 45B, périodicité 15 et 30 jours) montre que le collagène continue à s'accumuler à chaque nouvelle stimulation pour la dose 1000, alors qu'il diminue pour les doses 10 000 et 100 000. Cet effet reste visible pour un plus grand nombre de stimulations (Figure 45 C, périodicité 180 h et 15 jours et Figure 45 D, périodicité 90 h et 180 h).

Finalement, nous avons donc développé un modèle permettant de décrire la dynamique fonctionnelle des cellules étoilées au cours du développement de la fibrose et pendant sa réversion. Pour ce faire, nous avons utilisé les *tokens* et *counters* proposés par le langage Kappa pour nous affranchir du coût de mémoire observé lors de la création de notre modèle initial (Résultats, section 5). Nous avons mis en évidence les effets de deux nouveaux phénomènes biologiques, à savoir d'une part l'inactivation des cellules MFB réactivées qui est nécessaire pour expliquer l'accumulation de collagène I dans les modèles animaux de fibrose induite par le CCl4 et d'autre part le retour possible des cellules inactivées vers un état de quiescence. Enfin, les analyses de sensibilité de notre modèle montrent que la dynamique de dépôt de collagène I varie en fonction des paramètres de stimulation comme le suggèrent les observations dans les différents modèles biologiques.

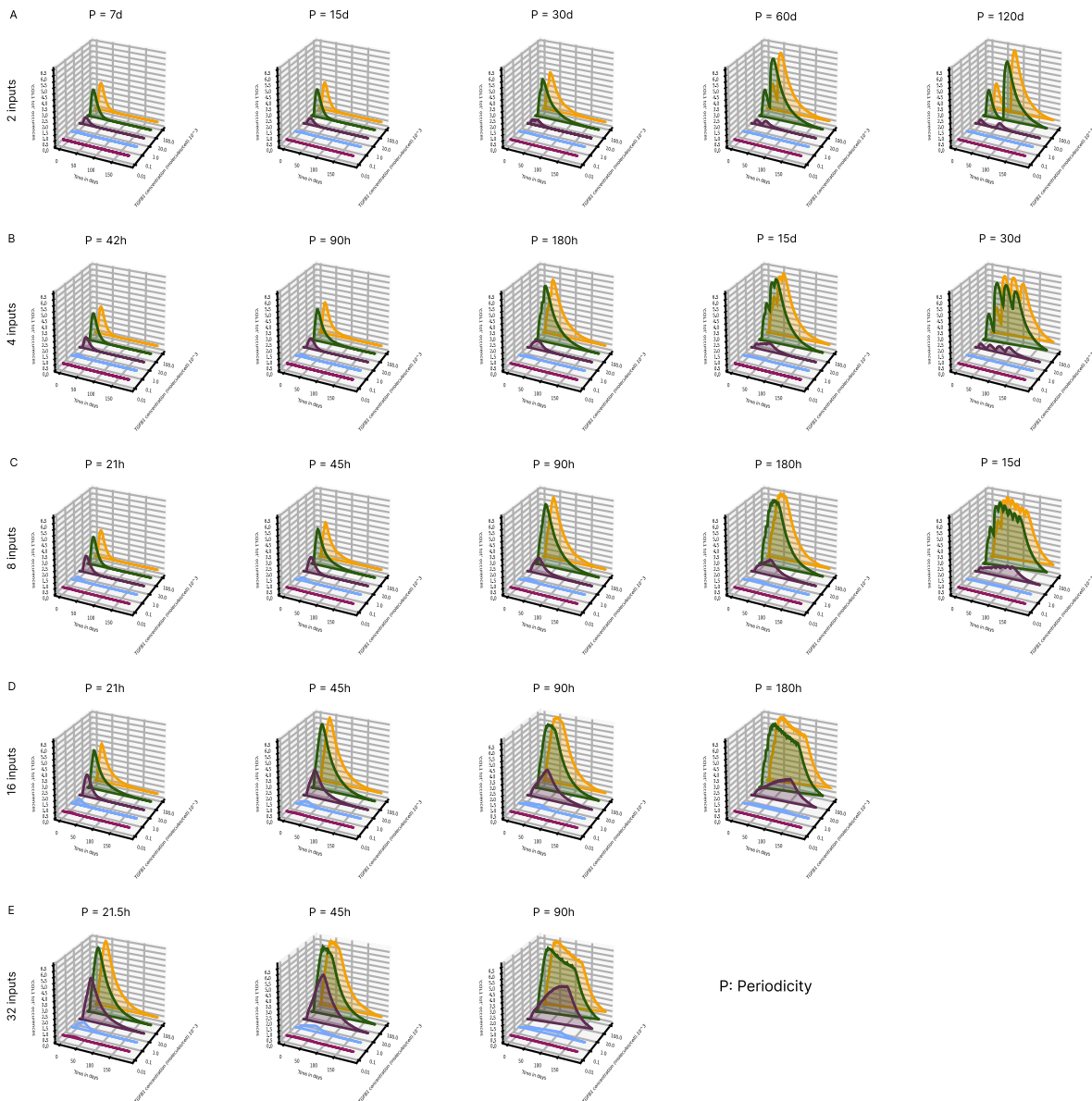


FIGURE 45 – Analyse de l'évolution temporelle de COL1 en fonction de la variation de la concentration, du nombre et de la périodicité des stimulations par le de TGFβ1. Les résultats sont exprimés en nombre d'occurrences de COL1 en fonction du nombre de stimulations par le TGFβ1 : (A) 2, (B) 4, (C) 8, (D) 16, (E) 32, la périodicité sur 1, 2, 4, 8 et 16 semaines et la concentration de TGFβ1 : 10 (rouge), 100 (bleu), 1 000 (violet), 10 000 (vert) et 100 000 (jaune) molécules par cellule. La première stimulation a lieu au jour 4. Les données sont exprimées comme la moyenne de cinq itérations. P, Périodicité

VALIDATION ET PRÉDICTIONS BIOLOGIQUES DU MODÈLE :

La troisième contribution de nos travaux de thèse portent sur la validation des prédictions du modèle finalement sélectionné (*reactMFB-with-inactivation*). Ce chapitre se décompose en deux parties. La première décrit la validation de la dynamique prédite par les modèles en utilisant de nouvelles données expérimentales provenant d'un modèle murin de fibrose hépatique induite par le CCl₄. La seconde partie décrit la validation de la prédiction d'accumulation des cellules étoilées inactivées au cours de la progression de la fibrose en utilisant les données de séquençage de patients atteints de maladie du foie gras non alcoolique (NAFLD).

7.1 Validation du modèle par de nouvelles données expérimentales chez la souris

Pour valider les prédictions du modèle, nous avons utilisé des données expérimentales produites par les biologistes de l'équipe dans le cadre de la thèse de Jérémy Bomo. Brièvement, les conditions expérimentales étaient les suivantes : des souris femelles C57BL/6 âgées de sept semaines ont été traitées par administration orale de CCl₄ (Sigma-Aldrich, St. Louis, MO, USA) dilué dans de l'huile d'olive. Une première dose de 2,4 g/kg a été administrée aux souris trois jours avant de commencer le traitement hebdomadaire avec une dose de 1,6 g/kg pendant dix semaines. Les souris témoins ont été traitées avec l'huile d'olive seule. Les souris ont été sacrifiées 24 heures après la dernière dose de CCl₄. Pour les expériences de réversion de la fibrose, les souris ont reçu des injections de CCl₄ pendant 4 semaines et ont été sacrifiées 8 jours (une semaine de réversion) ou 15 jours (2 semaines de réversion) après la dernière injection. Chaque groupe (témoin, fibrose et réversion) à chaque moment du sacrifice est constitué de cinq souris. La quantification du collagène a

été réalisée par microscopie par génération de seconde harmonique (SHG) sur des coupes de tissus congelés de 20μ . L'équipe a déjà démontré l'efficacité de la microscopie SHG pour quantifier le collagène fibrillaire dans le foie humain [97] et le foie de souris [27]. Il convient de noter que l'évaluation globale du dépôt de collagène à l'aide du rouge Sirius, mais aussi de la microscopie SHG peut surestimer la contribution des cellules étoilées puisque d'autres cellules contribuent au dépôt de COL1, comme les fibroblastes portaux et les cellules dérivées de la moelle osseuse [130]. Cependant, il a été récemment démontré que les cellules étoilées sont la source exclusive de myofibroblastes dans le modèle murin de fibrose induite par le CCl4 [283] et les données de quantification du COL1 qui ont servi à calibrer notre modèle Kappa proviennent justement de ce type de modèle [143].

Nous avons donc comparé les données de quantification du signal obtenu par microscopie SHG dans notre modèle expérimental avec les données obtenues par simulation. Comme le montre les Figures 46 A_1 , le signal SHG montre une accumulation maximale de COL1 à 6 semaines de traitement et une stabilisation à un niveau inférieur après 8 et 10 semaines de traitement. Dans les échantillons de tissu de souris ayant eu le protocole de réversion, le signal SHG montre une diminution du collagène à 4 et 8 semaines après le retrait du CCl4 (Figure 46 A_2).

Pour les études de simulation, les paramètres du TGFB1 ont été adaptés au protocole expérimental, c'est-à-dire trois stimulations de TGFB1 au cours de la première semaine et une stimulation par semaine pendant dix semaines. Le CCl4 induit des lésions hépatocytaires rapides entraînant une inflammation et une production de TGFB1 que nous avons modélisée en utilisant des quantités de TGFB1 saturante (10 000 à 100 000 molécules/cellule). Comme le montre la Figure 46 B_1 , les courbes prédites pour l'accumulation de COL1 ont atteint une valeur maximale similaire aux données expérimentales, puis à 8 et 10 semaines, le plateau de stabilisation est sensiblement plus élevé dans notre simulation. Concernant le protocole de réversion, les courbes prédites pour le COL1 sont pleinement en accord avec les données expérimentales (Figure 46 B_2).

Nous avons également validé la prédiction de la dynamique de COL1 par rapport à des données expérimentales indépendantes, notamment des modèles de fibrose hépatique induite par la diméthyl_nitrosamine [220] et le thioacétamide [1] comme le montre les Figures 47 A et B. Contrairement aux modèles chimiques, les modèles de fibrose induite par un régime riche en graisse, servant à imiter la progression des maladies du foie gras (NAFLD) vers la stéatohépatite non alcoolique (NASH), sont caractérisés par un niveau de fibrose plus faible. En outre, la dynamique de la fibrose est très variable selon les souches

de rongeurs ou la composition du régime alimentaire et le modèle optimal de stéatohépatite conduisant à la fibrose (régime déficient en choline et défini en acides aminés) est peu réversible [274]. Pour surmonter cette hétérogénéité, nous avons choisi un modèle de régime réversible [87] que nous comparons aux prédictions du modèle. Nous avons défini les paramètres du TGFB1 comme des stimulations faibles (4 500 molécules/cellule, 6 stimulations et une périodicité de 14 jours). Comme le montre la Figure 47 C, le modèle a capturé l'augmentation de COL1, mais la dynamique de réversion a été surestimée, ce qui suggère que des régulateurs manquent dans le modèle pour cette condition.

7.2 Association entre l'accumulation des cellules étoilées inactivées et le score fibrose

Une observation majeure des travaux de Kisseleva *et al.* [143] consiste en l'exacerbation de la réponse fibrotique à l'agression chez des souris ayant subi une première agression chronique suivi d'une période de réversion permettant de retrouver un phénotype de tissu sain. Cette observation a été associée à l'accumulation des cellules étoilées inactivées (iHSC) qui ont la capacité à répondre davantage à la stimulation par le TGFB1 entraînant une production plus rapide et plus importante de COL1 [143, 256, 80]. Ces cellules seraient donc encore présentes dans le foie "réparé" après six mois de réversion. Pour tenir compte de ces observations, nous avons calibré notre modèle *reactMFB-with-inactivation* sur les observations de Kisseleva *et al.* [143] après la deuxième stimulation en introduisant l'hypothèse d'une élimination lente des iHSC. Cependant, il n'existe aucune donnée expérimentale décrivant le devenir de ces cellules lors de cycles agression/réparation. Nous avons donc étudié ce comportement des iHSC au cours de répétitions du cycle agression/réversion en suivant le protocole décrit par Kisseleva *et al.* [143] qui comprend une agression chronique induite par huit injections de CCl4 sur un mois, suivie d'une période de réversion de six mois, puis d'une nouvelle agression chronique similaire à la première. Nous avons répété ce cycle trois fois (Figure 48).

Nos simulations montrent d'abord que la dynamique des cellules exprimant α -SMA et des quantités de COL1 étaient conformes aux observations expérimentales de Kisseleva *et al.* [143] pendant le premier cycle, à savoir des taux plus élevés de ces deux marqueurs après la deuxième stimulation par rapport à la première (Figure 48 A, B, C cycles 1, 2). Sur cette même période, nous avons observé une accumulation des iHSC. Nous avons ensuite répété ce cycle agression/réversion (Figure 48 C, cycles 3, 4) et nos résultats montrent une

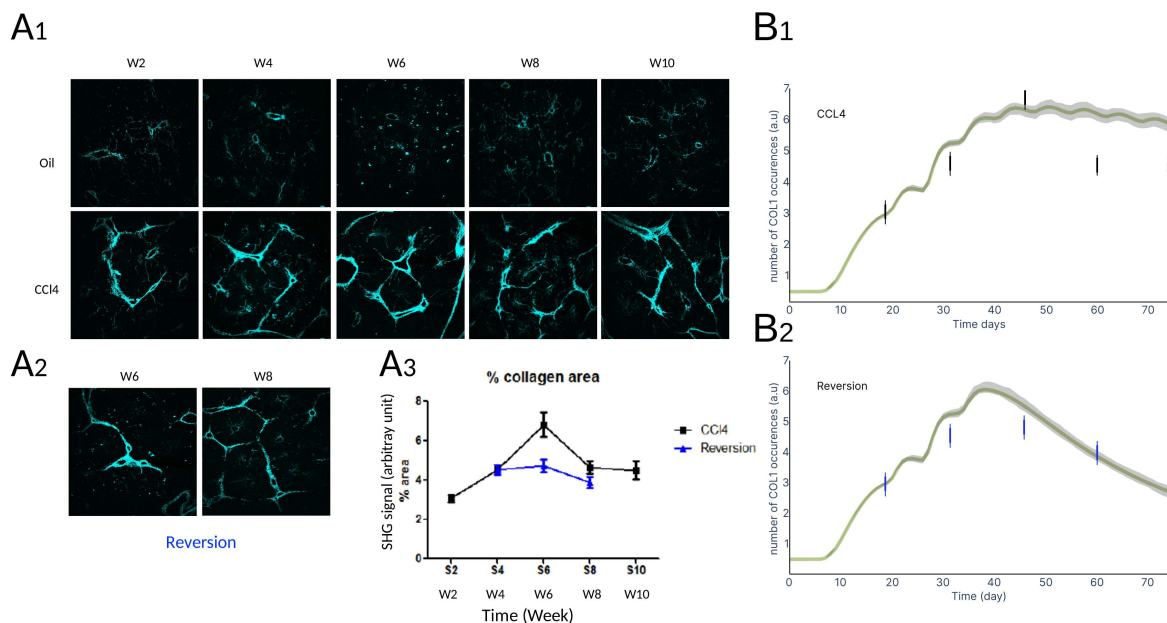


FIGURE 46 – Comparaison des prévisions du modèle et des nouvelles données expérimentales acquises dans un modèle de fibrose induite par le CCl4. **A1**) Images de microscopie SHG représentatives du dépôt de collagène dans les foies de souris après 2, 4, 8 et 10 semaines de traitement au CCl4. **A2**) Pour les expériences de réversion, le traitement a été arrêté à la semaine 4. **A3**) La quantification du collagène a été réalisée avec ImageJ et les résultats sont exprimés en pourcentage du signal SHG (moyenne \pm SD). **B1**), Les simulations du modèle ont été réalisées en utilisant des paramètres correspondant au protocole d’injection de CCl4 (trois stimulations de TGFB1 la première semaine et une stimulation par semaine pendant dix semaines). L’accumulation de collagène a été représentée graphiquement et les données expérimentales (de A1) sont indiquées par des barres noires. **B2**) Les simulations du modèle ont été réalisées en utilisant des paramètres correspondant au protocole de réversion (trois stimulation de TGFB1 au cours de la première semaine et une stimulation par semaine pendant quatre semaines). L’accumulation de collagène a été représentée graphiquement et les données expérimentales (de A2) sont indiquées par des barres bleues.

accumulation des iHSC. Cette augmentation peut résulter de l’équilibre entre le nombre plus élevé de cellules α -SMA conduisant à l’inactivation d’un plus grand nombre de iHSC et le temps de demi-vie des iHSC.

Sur la base de ces observations, le modèle prédit une accumulation progressive de iHSC au cours des cycles d’agression/réversion, suggérant un rôle critique des iHSC dans

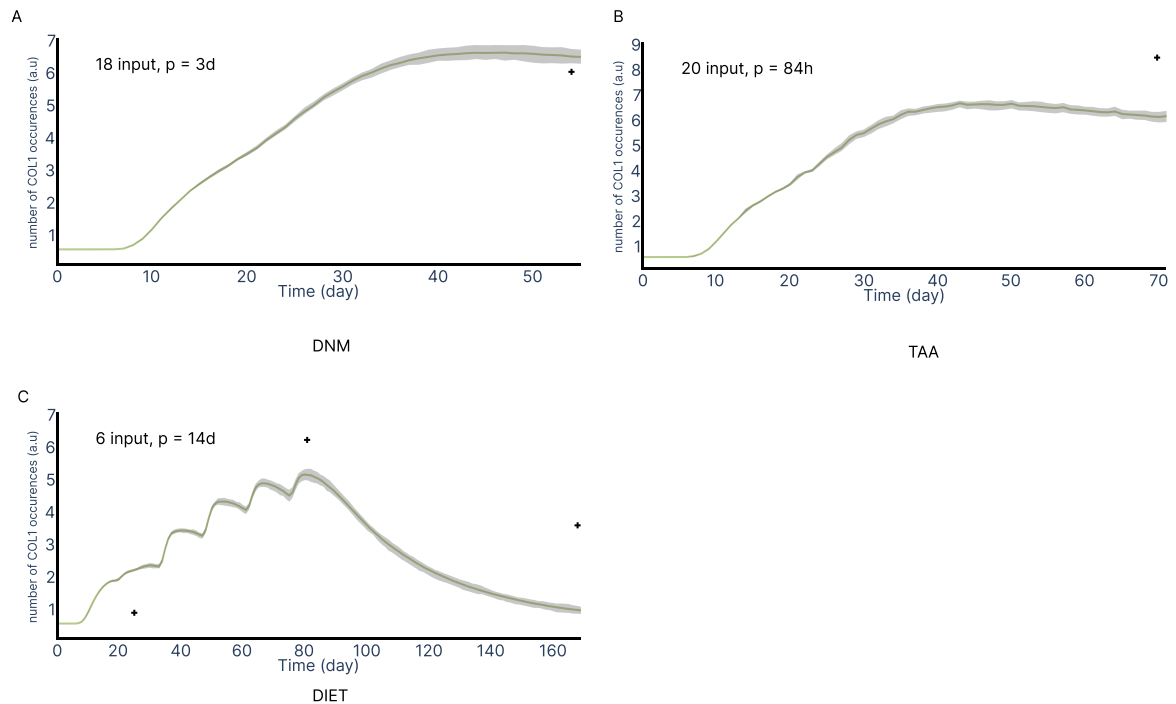


FIGURE 47 – Comparaison des prévisions du modèle et des données expérimentales acquises dans des modèles de fibrose induite par la diméthyl_nitrosamine, le thioacétamide et un régime riche en graisse. Les simulations ont été réalisées en utilisant les protocoles de stimulation décrits dans les différents modèles. A) modèle diméthyl_nitrosamine : 3 stimulations par le TGFB1 par semaine pendant 6 semaines [220], B) modèle thioacétamide : 2 stimulations par le TGFB1 par semaine pendant 10 semaines [1], C) modèle riche en graisse : 6 stimulations par le TGFB1 tous les 14 jours pendant 84 jours [87].

la dynamique de la fibrose, les iHSC accumulées constituant une source de cellules sur-productrices de COL1 lors de la réactivation.

Pour valider cette prédiction, nous avons cherché à caractériser l'accumulation de cellules inactivées dans des échantillons de patients atteints de fibrose hépatique. Pour cela, nous avons exploité les données RNAseq d'une étude multi-centrique comprenant 206 échantillons de foie histologiquement caractérisés de patients atteints de stéatose hépatique non alcoolique (NAFLD) dont 102 patients atteints de stéato-hépatite non alcoolique (NASH) associée à une fibrose hépatique [107]. Les données RNAseq ont été filtrées en utilisant la base de données BIOMART [176] pour sélectionner les gènes codant pour des protéines et les données ont été ensuite normalisées en utilisant l'outil DESeq2 [167] avec un seuil de 100 reads par gène. Cette étape de filtration a permis de réduire le

nombre de gènes de 21 595 à 16 021.

Pour identifier les iHSC dans les échantillons de foie humain, nous avons recherché l'expression de gènes spécifiquement exprimés par les iHSC. Pour cela, nous avons utilisé d'une part la signature d'expression génique des iHSC (39 gènes) identifiée par Rosenthal *et al.* [214] dans un modèle murin de stéatohépatite non alcoolique (NASH). D'autre part, nous avons utilisé 10 marqueurs géniques d'iHSC identifiés dans un modèle murin de fibrose hépatique induite par le CCl4 [143] et retrouvés dans un modèle de réversion *in vitro* de cellules activées humaines [80]. La liste de ces gènes est détaillée dans le Tableau 5. À noter que trois gènes (KRT20, GABRA3 et GSTT1) de ces signatures présentaient une expression trop faible dans les données de RNAseq des patients NASH et ont été éliminés après l'étape de filtration.

L'enrichissement de la signature de l'expression des gènes associés aux iHSC dans les échantillons fibrotiques de patients NASH a été évalué par la méthode GSEA, pour analyse d'enrichissement des ensembles de gènes [243, 184]. L'outil GSEA nous a permis de déterminer si l'expression des gènes associés aux iHSC présentait une différence statistiquement significative entre deux groupes de patients présentant respectivement un grade de fibrose faible (F0-F1, n=34) et un grade de fibrose élevé (F3-F4, n=68).

Comme le montre la Figure 49 A, les analyses GSEA réalisées avec la signature identifiée par Rosenthal *et al.* [214] montrent que les gènes iHSC sont plus exprimés dans le groupe de patient à fibrose élevée que dans le groupe à fibrose faible (score enrichi normalisé=1,38, valeur p = 0,097). Nous avons effectué une nouvelle analyse en ajoutant les 10 gènes marqueurs identifiés par Kisseleva *et al.* et El Taghdouini *et al.* [143, 80] à la liste de gènes identifiés par Rosenthal *et al.*. Comme le montre la Figure 49 B, cet ajout a augmenté la puissance statistique de l'analyse GSEA (score enrichi normalisé = 1,57, valeur p = 0,048), confirmant que l'expression de cette signature génique est augmentée chez les patients présentant une fibrose élevée, par rapport à ceux présentant une fibrose faible. Ces observations suggèrent donc une augmentation du nombre de cellules iHSC au cours de la progression de la fibrose hépatique, validant ainsi la prédiction de notre modèle.

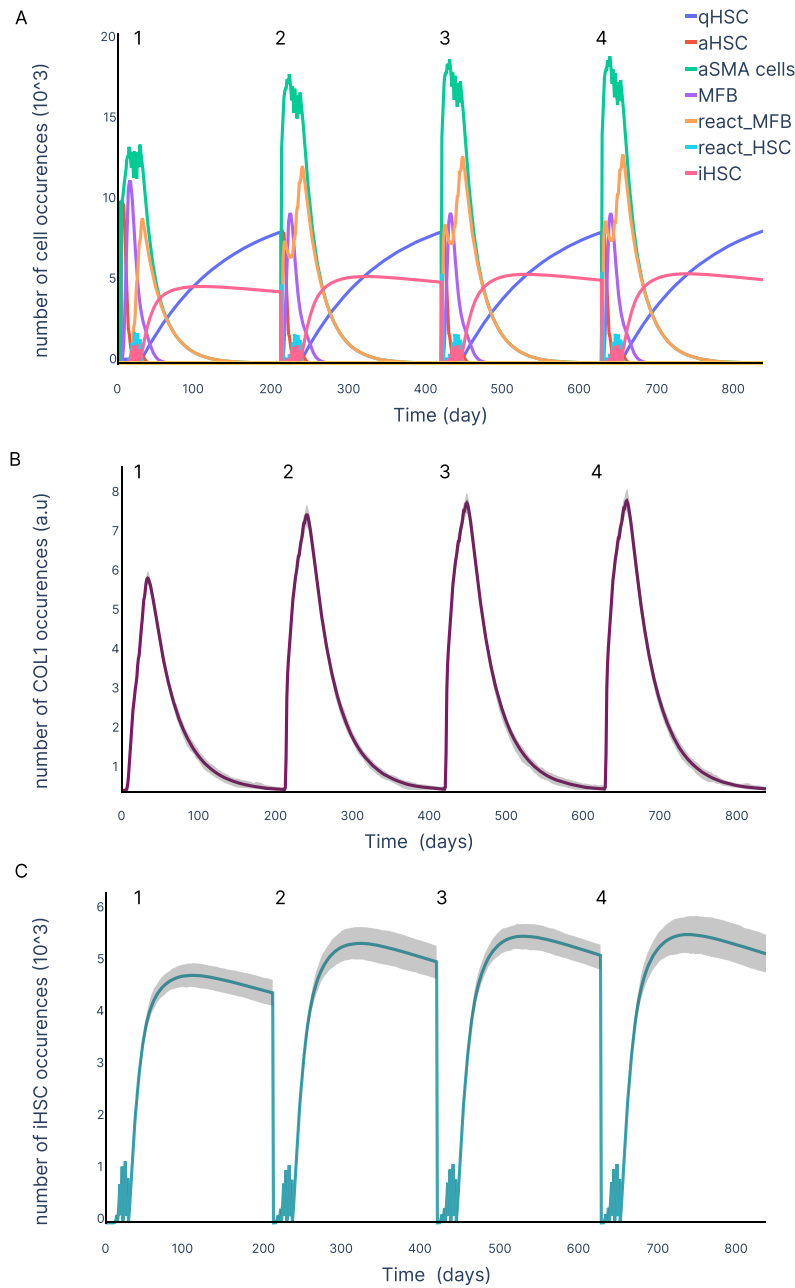


FIGURE 48 – **Accumulation des iHSC au cours des répétitions du cycle agression/réversion.** Les simulations ont été effectuées à l'aide d'un cycle de trois répétitions du protocole comprenant une agression chronique (huit injections deux fois par semaine pendant un mois), suivi d'une période de réversion de six mois et d'une deuxième agression, comme décrit précédemment [143]. Les données sont exprimées comme la moyenne de 50 simulations pour les cellules (A), et les 50 simulations ont été représentées pour le COL1 (B) et les iHSC (C).

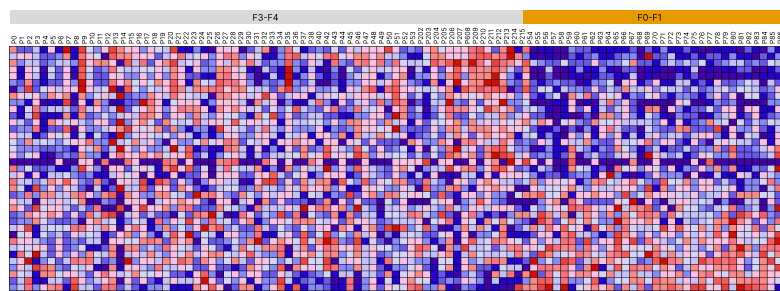
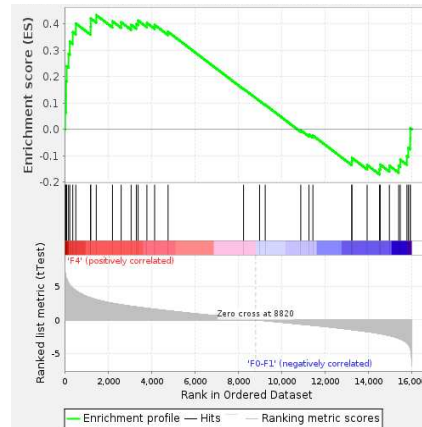
TABLEAU 5 – Liste des gènes utilisés pour l’analyse GSEA, la première colonne contient les gènes des travaux de Rosenthal *et al.*. Elle regroupe 39 gènes qui sont significativement exprimés de manière différentielle dans les cellules étoilées inactivées par rapport aux cellules étoilées activées et aux cellules étoilées quiescentes [214], la seconde liste contient les huit gènes marqueurs des cellules étoilées inactivées identifiés à la fois dans un modèle murin de fibrose hépatique induite par le CCl4 [143] et dans un modèle de réversion *in vitro* de cellules activées humaines [80]

Signature de Rosenthal		Signature combinée	
Ensembl code	Gene name	Ensembl code	Gene name
ENSG00000163739	CXCL1	ENSG00000077274	CAPN6
ENSG00000115380	EFEMP1	ENSG00000169245	CXCL10
ENSG00000148180	GSN	ENSG00000147257	GPC3
ENSG00000152583	SPARCL1	ENSG00000081041	CXCL2
ENSG00000112562	SMOC2	ENSG00000112799	LY86
ENSG00000164949	GEM	ENSG00000163131	CTSS
ENSG00000011465	DCN	ENSG00000168329	CX3CR1
ENSG00000177606	JUN	ENSG00000172602	RND1
ENSG00000171223	JUNB		
ENSG00000120738	EGR1		
ENSG00000170345	FOS		
ENSG00000137193	PIM1		
ENSG00000142871	CYR61		
ENSG00000130522	JUND		
ENSG00000120708	TGFBI		
ENSG00000168477	TNXB		
ENSG00000161544	CYGB		
ENSG00000272573	MUSTN1		
ENSG00000170801	HTRA3		
ENSG00000125740	FOSB		
ENSG00000162545	CAMK2N1		
ENSG00000130208	APOC1		
ENSG00000171867	PRNP		
ENSG00000162496	DHRS3		
ENSG00000163737	PF4		
ENSG00000000971	CFH		
ENSG00000224389	C4B		
ENSG00000121207	LRAT		
ENSG00000144152	FBLN7		
ENSG00000143369	ECM1		
ENSG00000131620	ANO1		
ENSG00000137491	SLCO2B1		
ENSG00000146374	RSPO3		
ENSG00000165795	NDRG2		
ENSG00000147481	SNTG1		
ENSG00000133110	POSTN		
ENSG00000069431	ABCC9		

7.2. Association entre l'accumulation des cellules étoilées inactivées et le score fibrose

A

Enrichment Score (ES): 0.4256573
 Normalized Enrichment Score (NES) 1.3889463
 Nominal p-value: 0.097247705
 FWER p-Value: 0.053



B

Enrichment Score (ES): 0.46911344
 Normalized Enrichment Score (NES) 1.5760338
 Nominal p-value: 0.04789272
 FWER p-Value: 0.025

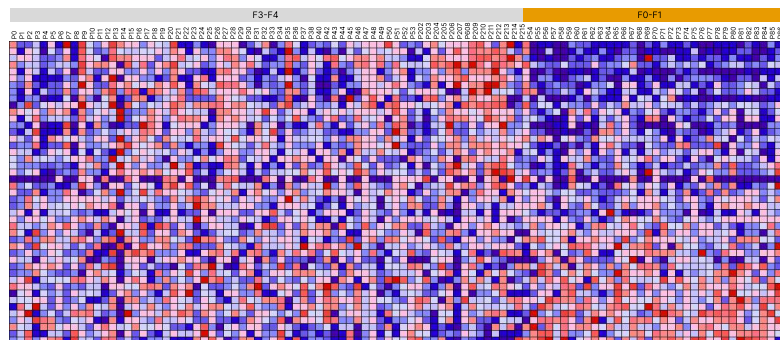
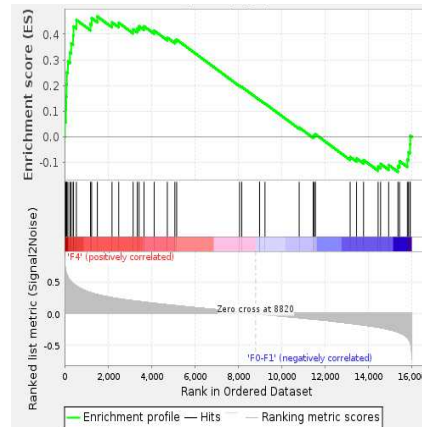


FIGURE 49 – L’accumulation des iHSC est associée à la fibrose hépatique chez les patients atteints de NASH. Une analyse d’enrichissement de l’expression des gènes des iHSC a été réalisée à partir d’échantillons de patients atteints de NASH à faible fibrose (F0-F1, n=34) et à forte fibrose (F3-F4, n=68) ([107]). **A** Analyse réalisée avec la signature génique des iHSC identifiée par Rosenthal *et al.* [214]. **B** Analyse réalisée avec la signature génique des iHSC identifiée par Rosenthal à laquelle sont ajoutés les dix marqueurs des iHSC identifiés dans les deux études [143, 80]. Les résultats sont présentés sous la forme de profils de corrélation entre les listes de gènes, de scores d’enrichissement (ES) qui indiquent à quel phénotype (faible ou forte fibrose) l’ensemble de gènes est le plus positivement corrélé et de cartes thermiques visualisant les clusters et la distribution des gènes.

Pour conclure, nos études de simulations montrent que notre modèle est généralisable à différents protocoles de fibrose chronique induite chez la souris. De nouvelles données expérimentales ont permis de valider la dynamique des HSC prédite par notre modèle. Enfin, le modèle identifie une nouvelle propriété de la dynamique des HSC à savoir une accumulation des iHSC au cours des cycles d’agression/réversion. Cette prédiction a été validée avec des données de RNAseq chez les patients NASH présentant une fibrose.

La persistance des iHSC au sein des tissus ayant été agressés constitue une trace cicatricielle susceptible d’être réactivée par une nouvelle agression, cette réactivation des iHSC conduira à une réponse fibrotique plus importante que sur un foie sain. Des données de suivi longitudinal de patients permettraient de mieux caractériser la dynamique de cette trace cicatricielle qui pourrait constituer un nouveau marqueur d’évolution/régression de la fibrose.

TROISIÈME PARTIE

Discussion et Perspectives

DISCUSSION ET PERSPECTIVES

Quelle que soit l'étiologie sous-jacente, toutes les maladies chroniques du foie sont associées au développement d'une fibrose hépatique, qui constitue un problème majeur de santé publique. La compréhension de la dynamique de la fibrose hépatique et de sa réversion est essentielle pour la prise en charge clinique des patients et l'adaptation des thérapies. La force motrice de la fibrose hépatique est l'activation dépendante du TGFB1 des cellules étoilées du foie (CEF) dont la plasticité est maintenant largement reconnue [255, 151]. Dans le cadre de cette thèse, nous avons développé un modèle de la dynamique des cellules étoilées et identifié de nouvelles propriétés liées à leur plasticité. Les contributions de nos travaux sont triples, tout d'abord, en termes d'approche de modélisation, nous avons démontré l'applicabilité du langage Kappa dans les approches multi-échelles puis nous avons développé des familles de modèles en introduisant : des *tokens*, pour représenter de manière compacte la quantité présente de certaines protéines, et des *counters* pour abstraite l'état interne des cellules. En termes de modélisation de la fibrose, notre modèle final est le premier à intégrer différents phénotypes des cellules étoilées. Il est aussi le premier à s'intéresser à la dynamique de ces cellules durant la phase de réversion de la fibrose. Enfin, notre modèle fournit des prédictions biologiques sur la boucle de réactivation des cellules, sur la dynamique des cellules inactivées et sur l'accumulation de celles-ci au cours de la progression de la fibrose. Ces prédictions ont été validées soit expérimentalement dans un modèle murin de fibrose induite par le CCl4, soit *in silico* en utilisant des données de patients NASH présentant une fibrose.

Modéliser la fibrose en langage Kappa

Kappa, un langage adapté à la modélisation multi-échelles

Alors que les modèles basés sur les équations différentielles ordinaires (ODE) sont utilisés depuis longtemps pour modéliser la dynamique biologique, les modèles basés sur des règles (RBM) sont beaucoup plus appropriés pour traiter les réseaux avec des interactions combinatoires multi-états [49]. En effet, les approches ODEs nécessitent l'implémentation d'une équation pour chacune des espèces. Les modèles ABMs, bien que plus appropriés aux approches bottom-up que les ODEs, demandent une description de tous les comportements possibles des agents, qu'il s'agisse d'un objet [69] ou d'un processus [50]. Dans le cas de notre modèle, notre unique agent possède 4 sites dont certains comportent 8 et 14 états différents. L'utilisation d'une approche ABM aurait fait exploser le nombre de règles du modèle. Les approches RBMs étaient donc les plus adéquates pour représenter les interactions connues qui opèrent sur les cellules étoilées durant le développement et la réversion de la fibrose.

Basés sur la réécriture de graphes de sites, BioNetGen [115] et Kappa [59] sont les deux principaux langages utilisés dans la littérature. Pour ces deux approches, les outils de simulation qui ont été développés offrent différentes options ; par exemple, avec BioNetGen, il est possible de choisir une simulation déterministe avec des ODEs ou stochastique avec SSA. Il en est de même pour Kappa avec son simulateur stochastique KaSim et son générateur d'ODEs : KaDE [41]. Une différence importante entre les deux langages réside dans la description des sites d'interaction des agents ; BNGL (BioNetGen language) autorise plusieurs sites de noms identiques alors que langage Kappa propose des *counters* qui peuvent remplacer des sites portant le même nom dans un certain limite. Si on a un site X possédant deux états u pour non-phosphorylé et p pour phosphorylé qui est présent 3 fois sur un Agent A en BNGL, en Kappa, on crée un *counter* X qui a un intervalle de 0-3. Un counter permet ainsi de compter le nombre de sites X dans l'état p. Cette simple différence explique en partie pourquoi le langage Kappa est plus rapide dans l'exécution de ces simulations, car ne pas autoriser deux sites de nom identiques permet de réduire grandement le coût de recherche des occurrences possédant le motif décrit à gauche des règles. Du fait de cette différence, la structure des données est plus efficace dans le langage Kappa.

Concernant les outils d'analyse du modèle, Kappa propose deux outils KaSa [33] et KaStor [60] permettant respectivement de réaliser des analyses statiques et causales du

modèle. De son côté, si BNGL propose des outils qui permettent son implémentation en Python avec PyBioNetGen ou de traduire des modèles SBML en BNGL via l’outil Atomizer [248], il ne comporte aucun outil d’analyse statique ou causale. Ces deux formalistes sont, par ailleurs, intégrés en python au travers de PySB qui est un module python permettant l’utilisation de différent formalisme RBM incluant Kappa et BNGL [165].

L’apport des *tokens* et des *counters* dans la modélisation Kappa

Le formalisme Kappa a été largement utilisé pour modéliser des réseaux de réactions moléculaires impliqués dans la signalisation cellulaire [62], la régulation des gènes [202], la régulation épigénétique [155], les circuits génétiques de type *repressilator* [278], la réparation de l’ADN [150] et la matrice extracellulaire [251]. Le passage des modèles d’interactions moléculaires aux modèles d’interactions cellule-molécule a constitué un nouveau défi, et nous avons au cours de cette thèse démontré l’applicabilité du langage Kappa à la modélisation multi-échelles [26]. Toutefois, pour surmonter l’explosion de la demande en mémoire due à la déclaration des molécules et des cellules en tant qu’agents, nous avons utilisé des *tokens* et des *counters*. La façon dont les *tokens* sont stockés en mémoire réduit considérablement le coût de calcul associé à l’existence de différentes échelles dans le modèle. Cela implique l’utilisation de règles hybrides entre agents et *tokens*, dont l’applicabilité a été démontrée dans nos modèles.

Les *counters* nous ont permis une description plus détaillée des étapes d’activation et de différenciation des CEF en créant des étapes intermédiaires qui ont conduit à un processus moyen proche de celui observé expérimentalement. L’utilisation des *counters* a également permis de décrire les différents comportements des cellules en fonction de leur stade d’activation/différenciation. Pour résumer, les *counters* et les *tokens* sont la clé de la modélisation multi-échelles dans Kappa. Les *tokens* permettent de représenter de manière plus compacte des populations d’agents, les *counters* permettent de représenter de manière plus compacte des populations de sites, tout en facilitant la recherche et l’abstraction des applications potentielles de règles dans les deux cas. À notre connaissance, c’est la première fois que les *counters* et les *tokens* sont utilisés dans un modèle Kappa. Le modèle non publié de William Waites sur les modèles d’infections virales pulmonaires¹ utilise lui aussi des *counters* dans son modèle.

1. <https://git.sr.ht/~wwaites/immune-transmission>

Limitations du langage Kappa

Au cours de l'élaboration de ce modèle multi-échelles constituant une nouvelle application pour le langage Kappa, nous avons identifié des limitations qui constituent autant de nouvelles pistes d'amélioration pour l'utilisation du langage Kappa. Les règles hybrides permettent de faire interagir des agents classiques avec des *tokens*, cependant ces deux types de composants n'ont pas du tout les mêmes propriétés. Du côté des agents classiques, il est facile de changer leurs configurations au gré des interactions, mais leurs nombres, lorsqu'ils sont présents en très grande quantité, peut devenir prohibitif pour la simulation, notamment du fait de leur empreinte mémoire. Du côté des *tokens*, la mise à jour de leur valeur est très simple, mais il est impossible de changer localement la configuration des composants qu'ils représentent. De ce fait, les *tokens* offrent une alternative aux agents classiques pour modéliser des composants avec peu de structure, mais pouvant intervenir en grande quantité. Il est toutefois possible de changer la configuration des composants représentés par des *token* en introduisant un *token* par configuration, mais cela devient vite fastidieux. Par ailleurs, seul le simulateur prend en charge les *tokens*, l'analyse statique et l'analyse causale les ignorent. Il est donc impossible d'augmenter la confiance et de comprendre les parties du modèle qui reposent sur l'utilisation de *tokens*. Du point de vue du concepteur de modèles, l'utilisation des *tokens* devrait être transparente. Il faudrait que l'utilisateur puisse décrire son modèle avec uniquement des agents classiques, c'est-à-dire des graphes à sites. Puis le compilateur de Kappa devrait choisir de lui-même quels composants doivent être représentés classiquement et quels composants doivent être encodés sous la forme de *tokens* en tenant compte du compromis entre la versalité, c'est-à-dire le nombre de configurations différentes de ces composants et la quantité, c'est-à-dire le nombre d'occurrences de ces composants.

L'utilisation des *counters* a constitué un second défi. Actuellement, l'usage de ces *counters* n'est pas très intuitif. En effet, dans la syntaxe actuelle, les règles d'interactions ne peuvent tester uniquement que si la valeur d'un *counter* est égale à une valeur ou si elle est plus grande qu'une valeur. Or, tester si la valeur d'un *counter* est inférieure à une valeur est également utile. Pour palier à ce manque, il est possible d'utiliser deux *counters* dont la somme reste constante. Aussi, pour tester si la valeur d'un *counter* est égale ou supérieure à une valeur, le premier *counter* sera utilisé, alors que pour tester si la valeur d'un *counter* est inférieure à une valeur, le second *counter* sera utilisé en traduisant la comparaison sous forme d'inégalité dans l'autre direction. Cet encodage demande de modifier les règles pour mettre à jour la valeur des deux *counters* quand la

valeur du *counter* initial est modifiée. C'est encodage fastidieux devrait être automatisé et être complètement transparent du point de vue de l'utilisateur. Par ailleurs, l'utilisation des *counters* sur de grandes valeurs n'est pas optimale, comme l'a montré notre tentative d'individualisation des récepteurs où le temps de calcul a explosé, que cela soit lié à la simulation elle-même ou au temps mis pour générer la carte des interactions. Une nouvelle perspective à ce travail consistera par conséquent à faire évoluer la syntaxe de Kappa pour faciliter l'usage de ces compteurs.

Enfin, nous avons été confrontés à la lourdeur du lancement de multiples simulations pour analyser la distribution des comportements observés et qui a nécessité de créer un script en l'absence d'outil développé dans l'écosystème de Kappa. Dans les perspectives à court terme, nous proposons d'améliorer notre pipeline permettant l'automatisation du lancement des simulations (scripts disponibles sur Github²) et de développer une API qui sera mise à disposition de la communauté. Actuellement, notre script utilise un fichier de paramètres *ini* en entrée et permet de modifier les paramètres du modèle et de lancer le nombre voulu de simulations. Ce pipeline est couplé avec un script permettant de générer les représentations graphiques des simulations sous différentes formes. À terme, ce pipeline facilitera le travail du biologiste pour analyser des modèles Kappa en modifiant facilement les paramètres et en visualisant les résultats des simulations de façon conviviale et intuitive.

Un modèle pour capturer l'hétérogénéité des cellules étoilées et le remodelage du microenvironnement

De récentes études de single-cell ont démontré la forte hétérogénéité des CEF et MFB durant le développement de la fibrose [153, 214, 288]. Ces travaux montrent aussi qu'en plus d'une forte diversité de phénotypes, la dynamique des CEF durant leur processus d'activation et de différenciation est loin d'être linéaire comme le montrent les analyses en pseudotime (Section 1.4). Notre modèle parvient à modéliser cette hétérogénéité en introduisant des étapes intermédiaires au processus d'activation et de différenciation, et en permettant aux cellules, dans chacun de ces états, de proliférer, de progresser dans le processus d'activation et de produire du COL1. L'implémentation de ces règles couplées à l'algorithme SSA (Stochastic Simulation Algorithm) permet d'avoir à chaque instant t

2. https://github.com/MBougueon/simulation_pipeline

de la simulation différents phénotypes et de conserver un historique pour chaque agent qui est "unique". Une cellule A peut par exemple ne jamais produire de COL1 avant d'être complètement transdifférenciée en MFB alors qu'une cellule B peut proliférer après chaque avancée dans le processus de transdifférenciation.

Cependant, la modélisation de la dynamique des CEF ne rend pas compte de toute la complexité du microenvironnement qui régit cette dynamique, par exemple la multitude d'événements régulés par les cellules immunitaires [111]. D'autres modèles multi-échelles incluent le rôle des cellules de Kupffer dans l'activation des CEF via le TNF- α [77, 271] ou la régulation antagoniste par les macrophages M1 et M2 [94]. De même, il est difficile d'intégrer tous les composants impliqués dans le remodelage de la matrice extracellulaire. Ce processus est finement régulé par l'équilibre entre la production de métalloprotéinases matricielles (MMP) et de leurs inhibiteurs, les Tissue Inhibitor of metalloproteinases (TIMP), tous deux sécrétés par les CEF, les MFB, mais aussi par les cellules immunitaires impliquées dans la résolution de la fibrose [149, 210, 76]. Dans leur modèle, Friedman & Hao [94] ont réduit le remodelage de la MEC à la seule contribution des macrophages et ne prennent pas du tout en compte le rôle connu des CEF. Pour surmonter cette complexité moléculaire et cellulaire, tout en intégrant cette notion de remodelage dans notre propre modèle, nous avons introduit des *tokens* pour représenter le COL1 à faible et à fort remodelage afin de modéliser respectivement l'accumulation de collagène pendant la stimulation du TGFB1 et sa régression lors du retrait du TGFB1. À noter que la contribution des MFB sénescents aux mécanismes de régression de la fibrose a été documenté [156] notamment par la production de MMP participant à la dégradation de la MEC et de cytokines qui potentialisent l'élimination des CEF sénescents par les cellules NK (Natural Killers) [207]. Actuellement, aucun modèle de fibrose n'inclut la sénescence des MFB dont l'évaluation quantitative durant le développement fibrose reste très limitée.

Une des perspectives de ce modèle serait évidemment d'intégrer d'autres informations biologiques. À titre d'exemple, le TGFB1 de notre modèle est actuellement sous sa forme active alors que son activation résulte en réalité d'un processus complexe qui implique de nombreux constituants de la MECs dont un grand nombre est produit par les CEF activées et MFB. Le modèle Kappa d'activation du TGFB1 précédemment développé par notre équipe Théret *et al.* [251] pourrait donc être adapté en utilisant des *tokens* pour être intégré à notre modèle de la dynamique des CEF. Cependant, des verrous restent à lever pour interface les deux modèles à partir des composants partagés et introduire les règles

de production/dégradation des différents composants matriciels à partir des données de la littérature (données de séquençage de cellules uniques et données de protéomique).

Un aspect important mis en lumière par le modèle d'activation du TGFB1[251] et qui n'a pas été évoqué dans nos travaux de thèse concerne l'architecture de la matrice extracellulaire qui est structurée principalement par des fibres de fibronectine et de collagène. Cette architecture possède un fort impact sur le signal mécanique, notamment durant le développement de la fibrose où l'environnement matriciel est constamment remodelé. Cette composante fibrillaire de la matrice est constituée d'agglomérats de macromolécules formant des polymères. En modélisation, l'implémentation des polymères est un défi pour plusieurs raisons. Tout d'abord, à cause des nombreuses possibilités d'interactions entre les éléments composant le polymère, entraînant une explosion de la combinatoire, ainsi que des comportements non souhaités (p.e. auto-liaison, compétition) [30]. Le coût de calcul peut aussi être fortement augmenté dans le cas où les éléments composant le polymère interagissent différemment avec d'autres éléments du système. Chaque composant du polymère ayant sa propre dynamique en plus de celle du polymère lui-même, le coût computationnel devient alors très important. Le langage Kappa permet de modéliser la formation et le comportement des polymères, quand ils peuvent être décrits par des règles d'interaction locale. L'analyse statique permet de mieux comprendre leurs propriétés structurelles [30] et tant qu'elles restent locales, la simulation de ces interactions ne pose guère de difficulté. Il est également possible d'utiliser des artifices pour prendre en compte des conditions globales sur les polymères, mais cela reste coûteux et compliqué. Un exemple consiste à vouloir borner la taille des polymères qui se forment, bien que cela soit possible en fixant des contraintes strictes comme en utilisant un *counter* pour compter le nombre de monomères constituant le polymère. Il est pour l'heure compliqué de le faire en préservant la dynamique du modèle. L'utilisation de *counters* pourrait permettre à la dynamique du système d'atteindre un équilibre dans la taille maximale des polymères en les utilisant cette fois pour modifier les taux d'application des règles afin de ralentir l'augmentation de la taille des fibres quand leur longueur augmente. Il est aussi important de noter que la longueur du polymère peut impacter l'affinité d'interaction avec d'autres éléments [198]. Ces exemples, loin d'être exhaustifs, constituent autant de paramètres à prendre en compte pour modéliser la structure de la matrice extracellulaire.

Un modèle prédictif

L'inactivation des MFB réactivés, un élément essentiel pour la progression de la fibrose

Le processus d'inactivation des MFB issus des CEF est présenté comme un élément clé de la réversion de la fibrose [143, 256] permettant l'élimination de $\sim 50\%$ des cellules MFB. Ces auteurs ont démontré que d'une part ces cellules inactivées peuvent être réactivées *in vivo* et *in vitro* en réponse à un nouveau stimulus et d'autre part qu'elles sont beaucoup plus réactives et davantage fibrogènes que les cellules activées initialement [143, 80]. Cependant, le devenir de ces cellules réactivées restait inconnu, ce qui nous a conduit à émettre l'hypothèse d'une potentielle ré-inactivation de ces cellules réactivées. Nos résultats ont montré que l'inactivation de ces cellules était indispensable au maintien de l'accumulation de COL1 en cas de lésions chroniques, mais que cette inactivation devait être mesurée. En effet, nos familles de modèles ont montré que les MFB réactivés devaient être inactivés dans une proportion d'environ 5% : au-delà, les observations expérimentales ne sont plus respectées. Cette boucle d'inactivation et réactivation induit un changement de la population de MFB. Plus les lésions s'accumulent, plus les cellules inactivées entre deux agressions sont réactivées, aboutissant à une population de MFB réactivés qui devient majoritaire et plus fibrogénique. À l'appui de ces prédictions, les données de séquençage sur cellules uniques ont identifié ces phénotypes de MFB producteurs de MEC indiquant une augmentation de l'activité fibrogène [214].

La réversion des cellules étoilées inactivées vers un état quiescent comme voie alternative à leur dégradation

Bien que l'existence de cellules étoilées inactivées soit étayée par plusieurs études, leur comportement n'est pas encore bien compris. Ces cellules ont un phénotype proche de celui des cellules quiescentes, mais conservent la mémoire de leur activation antérieure [143, 80]. Cependant, le devenir des cellules inactivées en l'absence de stimuli de réactivation reste inconnu. L'élaboration d'une famille de modèles nous a permis d'observer que la réversion de ces cellules inactivées vers un état quiescent, quel qu'en soit le pourcentage, n'affectait pas le comportement global des autres composants. Cette observation peut être due principalement aux temps d'élimination des cellules inactivées et de renouvellement des cellules quiescentes, qui sont particulièrement lents dans notre modèle. Cette

dynamique limite l'impact de cette réversion. Il est aussi possible que ces cellules aient une dynamique d'élimination plus rapide, mais qui serait compensée par une capacité à proliférer, ralentissant ainsi leur élimination et augmentant l'impact d'une potentielle réversion. En l'état, nos modèles suggèrent que la réversion complète des cellules inactivées vers l'état de quiescence est un moyen potentiel de les éliminer en l'absence de TGFB1. Un suivi expérimental de ces cellules sur des temps supérieurs à un mois de réversion permettrait d'être plus précis sur le devenir de ces cellules.

Les cellules étoilées inactivées, nouveaux marqueurs de l'évolution de la fibrose ?

La validation de l'accumulation des cellules inactivées au cours de la fibrose est un concept très important en termes cliniques, où la progression de la fibrose n'est pas du tout linéaire en fonction des dommages chroniques. Les cellules inactivées au sein des tissus pourraient agir comme une sorte de mémoire d'une lésion initiale, rendant le tissu plus sensible à la lésion suivante. Bien sûr, cette mémoire facilite une réparation rapide, mais la récurrence de ces épisodes de réparation, s'ils sont trop rapprochés, accélère la progression de la fibrose. Notre modèle n'est certes pas totalement adapté pour suivre la dynamique du modèle murin de NASH en raison de la libération lente du TGFB1 au cours du régime alimentaire. Nous avons pu cependant valider, grâce à leur signature transcriptomique, l'accumulation des cellules inactivées chez les patients NASH atteints de fibrose.

Identifier les cellules inactivées *in vivo* constituerait un véritable marqueur de l'histoire du patient. Cependant, cette identification reste une question complexe en raison du phénotype intermédiaire de ces cellules, partagé entre celui des cellules quiescentes et celui des cellules activées. Il n'existe pas encore de signature phénotypique établie pour les cellules inactivées et les signatures d'expression des gènes varient selon les modèles murins utilisés (NASH, CCl4). Ces signatures phénotypiques dépendent à la fois du temps de réversion et des méthodes utilisées pour les identifier, selon que leur expression soit comparée avec celle des cellules quiescentes et/ou des cellules activées. Les récentes avancées en matière de séquençage de cellules uniques permettront d'affiner ces signatures en caractérisant des groupes de gènes plus spécifiques.

Les apports de l'apprentissage automatique

Il est difficile de parler de biologie computationnelle sans évoquer les approches de machine learning (ML), neural network (NN) et de deep learning (DL) qui sont maintenant largement utilisées dans le domaine de la santé. Ces technologies ne font pas l'objet de notre travail de thèse, mais elles participent aujourd'hui à une meilleure compréhension des maladies chroniques hépatiques, au suivi des patients, mais sont surtout utilisées pour l'aide au diagnostic. De récentes revues permettent d'avoir un état de l'art assez complet ([19, 57, 242]) et parmi celles-ci, la revue de David Nam [190] qui classe les différentes approches ML appliquées à la fibrose selon trois grandes catégories : *Diagnostic, Prédiction et Segmentation*.

Si l'intelligence artificielle apporte sans conteste de solides avancées dans le domaine de la santé, il faut cependant rester prudent quant à son utilisation, notamment dans le domaine des maladies chroniques du foie. Comme le souligne Gennaro D'Amico [57], ces approches sont complexes et variées, elles nécessitent une formation pour les maîtriser et comprendre la "boite noire" si souvent citée dans les approches de ML/DL. Gennaro D'Amico *et al.* souligne aussi l'importance de considérer l'évolution de la pathologie dans le choix des données, un modèle capable de diagnostiquer un stade F1-F2 peut ne pas fonctionner aussi bien sur des stades F3-F4, les caractéristiques pathologiques étant différentes. Par ailleurs, il n'y a pour le moment aucune standardisation des processus de traitement des données dans ce domaine, que ce soit la sélection des données pour l'entraînement du modèle et celles pour le tester, même si certains outils comme PROBAST [279] permettent de mesurer le risque de biais du modèle en analysant les données et les choix faits aussi bien dans la sélection que dans les données écartées. Pour conclure, l'usage de l'intelligence artificielle sera de plus en plus utilisé dans le domaine de la santé, mais il est indispensable de développer une interdisciplinarité forte et de préserver une intelligence humaine.

ANNEXES

A.1 Différents formalismes en biologie des systèmes

Il existe divers formalismes adaptés pour chaque type de modèles, dans cette annexe, nous proposons d'en présenter quelques-uns. Pour chaque formalisme, il existe différents outils permettant leur implémentation, une très belle revue de Musilova et Sedlarfait fait une présentation de 50 outils utilisant des approches dynamiques en biologie des systèmes. Leur présentation de ces outils décrit en détails le type de données (qualitative, quantitative) ainsi que le type de simulation (déterministe, stochastique) [187].

Les réseaux booléens

Les réseaux booléens sont des modèles qualitatifs où le temps est discrétisé et où la quantité de chaque espèce est définie par deux valeurs : 0 ou *faux* (absent) et 1 ou *vrai*. La résolution de ces systèmes passe par une fonction f appliquée sur un espace booléen (A.1) où n est le nombre d'espèces du système.

$$f : \mathbb{B}^n \rightarrow \mathbb{B}^n \tag{A.1}$$

$$\tag{A.2}$$

Le graphe de cette fonction permet de représenter le système sous la forme d'une matrice de vérité, ou d'un graphe dirigé, où les espèces sont les nœuds et les arcs, qui peuvent être pondérées par des fonctions, les relations (dépendance, impact) entre les espèces. La mise à jour du réseau à chaque pas de temps peut se faire de différentes manières dont voici les 3 principales :

- Synchrones : toutes les réactions réalisables sont appliquées simultanément à chaque tour [99]

-
- Asynchrone : une seule réaction est appliquée à chaque tour [99]
 - Probabiliste : plusieurs réactions peuvent être réalisées simultanément en fonction de leur probabilité [233, 108]

En biologie, de nombreux processus observés par des approches quantitatives montrent que la production d'un élément n'est induite que lorsqu'une certaine quantité d'un élément régulateur est suffisante. La dynamique, suivant une courbe sigmoïde obtenue par l'ODE caractérisant l'équation, est telle que la valeur de seuil induisant un effet peut être abstraite à un *ON/OFF* [226]. C'est par exemple le cas pour certains coefficients de Hill, dont le sigmoïde décrit cette caractéristique [123, 105]. Ces approches sont donc particulièrement utilisées dans l'étude des réseaux de régulation des gènes et la production de protéines [43]. Un récent état de l'art, fait par Julian D. Schwab *et al.* en 2020, détaille davantage le fonctionnement des réseaux booléens ainsi que les différents outils permettant de les utiliser [226]. Les réseaux booléens permettent donc de décrire la topologie d'un système en ne nécessitant que les informations sur les relations entre les espèces de son système.

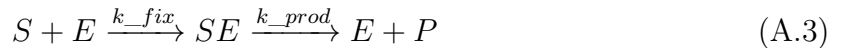
Les réseaux de Petri

Les modèles de réseaux de Petri sont des modèles de graphes, plus précisément des graphes bipartites où les espèces sont discrétisées. Un PN est un graphe qui est dirigé, pondéré et bipartite où les places sont des cercles et les transitions des rectangles. Les places correspondant aux espèces et les transitions sont les événements permettant de passer d'une espèce à une autre. Chaque cercle peut contenir entre zéro et plusieurs jetons, qui correspondent à l'énumération discrète du nombre d'individus pour une espèce. Les transitions sont appliquées si le nombre de jetons dans les cercles sont suffisants pour remplir les conditions pondérées par les arêtes. Si les conditions sont remplies, le jeton passera du cercle en amont de la transition à celui en aval de celle-ci [212]. Les réseaux de Petri sont une alternative aux réseaux réactionnels où chaque place correspond à une configuration d'espèce et chaque transition à une réaction. Ils peuvent également être munis de diverses sémantiques (non-déterministe, différentielle, stochastique). Des méta-syntaxes permettent d'exprimer de manière compacte des réseaux compliqués sous la forme de réseaux de Petri hiérarchiques ou colorés. Les réseaux de Petri sont des modèles idéaux pour la description de réseaux métaboliques [211, 22] dû à la représentation des transitions qui permet de décrire les réactions, mais aussi la stœchiométrie de celle-ci. De plus, leur approche qualitative (ou hybride) nécessite moins de données que les ODEs, mais reste dans un temps continu.

Les modèles basés sur des équations différentielles

La théorie des modèles dynamiques propose des approches mathématiques (dont font partie les ODE) qui permettent l'analyse de système biologique [61]. Les modèles basés sur des équations différentielles se divisent en trois catégories : ordinaires (ODE), partielles (PDE), stochastiques (SDE). Les deux dernières sont des variantes de la première. Les ODEs suivent dans un temps continue l'évolution de la concentration des espèces. Les ODEs suivent la loi d'action de Masse et parfois d'autres lois mathématiques bien connues en biochimie telles que Michaelis-Menten pour les réactions enzymatiques et Hill pour l'expression des gènes.

Chaque équation définira une réaction, et sera définie par un coefficient de réaction k et par la concentration des espèces. Ce formalisme mathématique permet donc d'associer une vitesse à chaque réaction qui permet de suivre les variations de concentration des différentes espèces du système. Par exemple, un processus enzymatique irréversible où une enzyme se fixe à son substrat pour générer le produit issu de ce substrat.



Ce processus sera défini par deux vitesses v_1 représentant la vitesse de fixation de l'enzyme à son substrat (A.4) et v_2 la production du produit et la dégradation du substrat (A.5). La réaction est définie par la concentration en enzyme $[E]$, en substrat $[S]$ et en produit P car suivant la loi d'action de Masse. Les coefficients de réactions de fixation k_{fix} et de production du produit k_{prod} doivent aussi être spécifiés.

$$v_1 = k_{fix}[S][E] \quad (\text{A.4})$$

$$v_2 = k_{prod}[ES] \quad (\text{A.5})$$

Partant de ces vitesses qui finalement définissent l'évolution des concentrations au cours du temps, il est nécessaire de définir l'évolution de la concentration de chacune des espèces impliquées. Les réactions impactant la concentration de chaque espèce doivent être précisées, celles consommant une espèce sont négatives et celle-là produisant sont positives :

$$\frac{d[S]}{dt} = -k_{fix}[S][E] \quad (\text{A.6})$$

$$\frac{d[E]}{dt} = k_{prod}[ES] - k_{fix}[S][E] \quad (\text{A.7})$$

$$\frac{d[ES]}{dt} = k_{fix}[S][E] \quad (\text{A.8})$$

$$\frac{d[P]}{dt} = k_{prod}[ES] \quad (\text{A.9})$$

Les exemples d'utilisations des ODEs sont très nombreux. En effet, il s'agit de l'une des approches de modélisation la plus utilisée [146] et servant fréquemment de référence [58, 48]. Leur utilisation a par exemple permis d'étudier les mécanismes de prolifération des hépatocytes [54], ou l'impact de l'accumulation acides gras sur le développement de la fibrose [221].

Les équations différentielles partielles (PDE) sont une version des ODEs qui, au lieu d'être définie par la dérivée d'un seul variable, sont définies par des dérivés partielles de plusieurs variables [240, 21]. Les PDEs sont utilisées pour étudier les flux, les vibrations, la diffusion et sont appliquées aussi bien en physique, en musique qu'en biologie. En modélisation, les PDEs sont notamment utilisés pour décrire le déplacement des molécules dans un environnement en deux dimensions, défini par l'équation A.10 de réaction-diffusion proposée par Euler en 1754 [84]. Dans cette équation D correspond à la matrice de diffusion appliquée a u et f la fonction modifiant la quantité de u .

$$\frac{\partial u}{\partial t} = D\nabla^2 u + f(u, x, t) \quad (\text{A.10})$$

Leur utilisation a permis par exemple dans les années 1950 à Turing de proposer le premier modèle de morphogenèse [258]. Les limites computationnelles de l'époque ne permettaient pas de résoudre ces équations, pour y parvenir, il fallait se placer dans un système avec un espace sans limites. Ce n'est qu'en 1990 qu'une solution fût trouvée permettant à partir de cet espace infini de redéfinir l'espace de Turing. Cette solution permit d'étudier le processus de régénération de l'épiderme après une lésion [232] et est encore utilisé aujourd'hui par exemple pour étudier la fibrose [94, 113, 112, 114].

Les équations différentielles stochastiques (SDE) sont des ODEs auquel s'ajoute une variable aléatoire qui définit la probabilité d'application d'un événement. Ces variables peuvent suivre différentes lois mathématiques, comme le processus de Poisson et le mou-

vement Brownien. La première est la plus simple des deux, il définit le nombre d'événements se réalisant dans un intervalle de temps donnés où la réalisation d'un événement est indépendante de l'état précédent du système sur un temps continu [206]. La seconde découle du processus de Markov (voir 2.1.2.3) continue, et permet de résoudre l'équation de Langevin découlant de ce processus. La résolution de cette équation par le processus de Brownien permet de définir la distribution de la réalisation de chaînes d'événement suivant une distribution Gaussienne [148, 206].

L'utilisation des ODEs dans l'analyse de voie de signalisation s'est démocratisée ces dernières années. L'apparition de nombreux outils a permis de faciliter leur utilisation, permettant de modifier les valeurs initiales et tester la sensibilité du modèle en modifiant ses paramètres [128]. L'utilisation des ODEs (PDE et SDE incluses) est cependant très dépendante de la disponibilité des données. La vitesse des réactions devant être connue, un grand nombre d'informations quantitative sont donc nécessaires pour écrire un modèle en ODEs. Elles sont aussi impossibles à utiliser pour représenter des systèmes avec des espèces possédant des nombreux états [49]. Il peut aussi être difficile de les utiliser sur des approches bottom-up, où l'ajout de nouvelles réactions demandera de mettre à jour les autres équations. Dans les modèles basés sur des ODEs, chaque espèce est définie par sa concentration et une équation décrivant son évolution au cours du temps. Cette équation devant spécifier toutes les réactions consommant ou produisant les espèces, ce qui rend compliquer la mise à jour du système.

Les automates cellulaires

Cette approche consiste à définir une matrice 2D ou 3D discrète définie par l'utilisateur, où chaque nœud (case) correspond à une *cellule* [45, 71]. Autrement dit, un automate cellulaire représente un espace cellulaire pouvant être composé de différentes espèces définies par des états discrets. La modification de ces *cellules* se fait dans une temporalité discrète où l'application de ces modifications peut être aussi bien stochastique que déterministe. La valeur d'une *cellule* à l'instant $t + 1$ dépend de sa propre valeur, mais aussi de la valeur des *cellules* adjacentes à t . Ainsi, la valeur d'une *cellule* peut être nulle ou détaillée en précisant la présence d'espèces. De cette manière, il est ainsi possible décrire le déplacement d'espèces cellulaires sur la grille en faisant se déplacer le "contenu" des *cellules*. Un automate cellulaire permet donc de décrire le comportement de cellules en fonction de leurs voisines, mais aussi de décrire leur déplacement au cours du temps. L'utilisation des automates cellulaires ne se limite pas à la biologie, bien que le nom de la

méthode puisse le laisser entendre. L'article de Yang Li *et al.* fait un état de l'art plutôt exhaustif (de 1998-2017) sur l'utilisation des automates cellulaires dans différentes disciplines comme la physique, l'économie, la biologie ou l'informatique [162]. La force des automates cellulaires réside dans la versatilité de leur utilisation et la possibilité d'appliquer une approche déterministe ou stochastique à un environnement spatial. Cependant, ces approches souffrent de la nécessité de déclarer l'ensemble des réactions, ce qui rend la mise à jour du modèle très compliquée.

Les modèles basés sur des Agents (ABM)

Il s'agit d'une approche discrète qui se concentre sur les interactions entre les éléments du système qui sont individualisés et définis comme des agents [121]. Un modèle ABM est défini par :

- les agents qui le composent, avec des caractéristiques propres à chaque type ;
- l'environnement, les agents ont des caractéristiques spatiales leur permettant de se déplacer ;
- les règles, éléments qui définissent tous les événements dans le système, permettant aux agents de se déplacer, mais aussi d'interagir si les conditions sont remplies.

Les agents peuvent être de différents types (p.e. différentes protéines) avec leurs propres caractéristiques (p.e. site de liaison, modification post-traductionnelle). Dans ces modèles, le système étudie le comportement de chaque agent individuellement, celui-ci possédant des caractéristiques et des comportements propres au groupe auquel il appartient. L'intérêt de cette approche est de capturer la diversité de comportement pour chaque agent. Les modèles ABM font partie des modèles spatio-temporels où chaque agent peut se déplacer sur une grille en 2D en utilisant une version simplifiée des PDE [106]. Ce n'est que lorsque deux agents sont sur la même *case* que l'application de la règle les impliquant est possible. Une règle est un événement, elle définit les conditions d'application de celui-ci et indique ce qui sera modifié à la suite de la réalisation de cet événement. Un avantage de ces modèles est leur flexibilité, ces modèles sont idéaux pour les approches bottom-up, car l'ajout de nouvelles règles et espèces ne demande pas de recalculer tout le reste du modèle. Les modèles ABM sont aussi stochastiques, ce qui permet aussi bien de réaliser aléatoirement les déplacements et les réactions. L'application des modèles ABM n'est plus à faire et fait partie des approches se démocratisant de plus en plus, notamment au niveau des modèles multi-échelles. Leur utilisation a, par exemple, permis d'étudier le développement de la fibrose [77] ou l'élaboration d'un modèle épidémiologique de la propagation de la covid-19

dans différents pays en fonction d'impact extérieurs [140].

A.2 Communications et Formations

Production scientifique

Article (en préparation)

Bouguéon, M., Legagneux, V., Bomo, J., Feret, J., Siegel, A., & Théret, N. (2023). A rule-based multiscale model of hepatic stellate cells plasticity : critical role of the inactivation loop in fibrosis progression

Chapitre de livre

Bouguéon, M., Boutillier, P., Feret, J., Hazard, O., & Théret, N. (2022). The Rule-Based Model Approach : A Kappa Model for Hepatic Stellate Cells Activation by TGFB1. Systems Biology Modelling and Analysis : Formal Bioinformatics Methods and Tools, 69-126. [26]

Contribution dans des conférences internationales

- Novembre 2021 ; CompSysBIO, Aussois France, A Kappa model for hepatic stellate cells activation by TGFB1 (poster) ;
- Septembre 2021 ; CMSB, Bordeaux, France, A Kappa model for hepatic stellate cells activation by TGFB1 (poster)

Contribution dans des conférences

Novembre 2021 GT BioSS, Lyon A Kappa model for hepatic stellate cells activation by TGFB1 (Présentation)

Enseignements

- 2022-2023 ; M1 AIRE, Université de Paris, Rules Based language (3 h CM and 3 h TD) ;
- 2022-2023 ; L2 OBI, Université de Rennes 1 python (12 h TP) ;
- 2022-2023 ; L1 BECV, Université de Rennes 1 python (36 h TP) ;

-
- 2022-2023 ; M2 bioinformatique, Université de Rennes 1, Rules Based language (4 h CM and 2 h TP) ;
 - 2021-2022 ; M1 AIRE, Université de Paris, Rules Based language (3 h CM and 3 h TD)

Encadrement

Durant mes travaux de thèse, j'ai eu l'opportunité d'encadrer les travaux de Master de Farah Diba Khan. Durant son stage, nous avons travaillé à l'élaboration d'un modèle sur la dynamique de gouttelette lipidique dans les CEF durant leur activation par le TGFB1. Bien que préliminaires, ces travaux pourraient se voir inclus dans notre modèle pour détailler les mécanismes intracellulaire qui ont lieu dans les CEFS durant leur activation.

Formations

Description de la formation	Trans	Scien	organisateur	Volume horaire
CompSysBio		X	Daniel Jost	21 h
Sciences en cour[t]s	X		Olivier Dennler	16 h
Recherche reproductible : principes méthodologiques pour une science transparente	X		FUN-MOOC	20 h
Introduction à l'éthique de la recherche et intégrité scientifique	X		EGAAL	4 h
Scientific english workshop	X		Sylvain Orain	10 h
Logiciel R / R Studio : Niveau 2 - Orientation Programmation		X	INSERM	21 h
Chemical Reaction Network Summer School		X	Politecnico di Torino	28 h
Sciences en cour[t]s(organisation)		X	Nicomaque	32 h

TABLEAU 6 – Tableau des formations réalisées. Trans : Transversale ; Scien : Scientifique

BIBLIOGRAPHIE

- [1] S. ABRAMOVITCH et al. « Vitamin D inhibits proliferation and profibrotic marker expression in hepatic stellate cells and decreases thioacetamide-induced liver fibrosis in rats ». In : *Gut* 60.12 (déc. 2011), p. 1728-1737.
- [2] M. ADLER et al. « Principles of Cell Circuits for Tissue Repair and Fibrosis ». In : *iScience* 23.2 (fév. 2020), p. 100841.
- [3] Aurélien ALFONSI et al. « Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems ». In : *ESAIM : proceedings*. T. 14. EDP Sciences. 2005, p. 1-13.
- [4] Martine Ben AMAR et Carlo BIANCA. « Towards a unified approach in the modeling of fibrosis : A review with research perspectives ». In : *Physics of life reviews* 17 (2016), p. 61-85.
- [5] Oana ANDREI et Hélène KIRCHNER. « A Rewriting Calculus for Multigraphs with Ports ». In : *Electr. Notes Theor. Comput. Sci.* 219 (2008), p. 67-82. URL : <https://doi.org/10.1016/j.entcs.2008.10.035>.
- [6] Steve S. ANDREWS et al. « Simulating macromolecular crowding with particle and lattice-based methods ». In : *Multiscale Spatial Computational Systems Biology*. 2014.
- [7] Geoffroy ANDRIEUX, Michel LE BORGNE et Nathalie THÉRET. « An integrative modeling framework reveals plasticity of TGF- β signaling ». In : *BMC systems biology* 8.1 (2014), p. 1-16.
- [8] Geoffroy ANDRIEUX et al. « Dynamic regulation of Tgf-B signaling by Tif1 γ : a computational approach ». In : *PloS one* 7.3 (2012), e33761.
- [9] J. ARAMBURU et al. « Computational Fluid Dynamics Modeling of Liver Radioembolization : A Review ». In : *Cardiovasc Intervent Radiol* 45.1 (jan. 2022), p. 12-20.
- [10] G. E. ARTEEL et A. NABA. « The liver matrisome - looking beyond collagens ». In : *JHEP Rep* 2.4 (août 2020), p. 100115.

-
- [11] Gavin ARTEEL et Alexandra NABA. « The liver matrisome, looking beyond collagens ». In : *JHEP Reports* 2 (avr. 2020), p. 100115. DOI : 10.1016/j.jhepr.2020.100115.
- [12] Gianluca ASCOLANI et Pietro LIÒ. « Modeling TGF- β in early stages of cancer tissue dynamics ». In : *PLoS One* 9.2 (2014), e88533.
- [13] *Association française pour l'étude du foie, Recommandation pour le diagnostic et le suivie non-invasif des maladies chroniques du foie.* <https://afef.asso.fr/wp-content/uploads/2020/07/DNI-VERSION-FINALE-RECO-2020.pdf>. Accessed : 2010-09-30.
- [14] Azeddine ATFI et al. « The disintegrin and metalloproteinase ADAM12 contributes to TGF- β signaling through interaction with the type II receptor ». In : *The Journal of cell biology* 178.2 (2007), p. 201-208.
- [15] Wenrui Hao AVNER FRIEDMAN. « Mathematical modeling of liver fibrosis ». In : *Mathematical Biosciences & Engineering* 14.1 (2017), p. 143-164.
- [16] M. G. BACHEM et al. « Activation of rat liver perisinusoidal lipocytes by transforming growth factors derived from myofibroblastlike cells. A potential mechanism of self perpetuation in liver fibrogenesis ». In : *J Clin Invest* 89.1 (jan. 1992), p. 19-27.
- [17] M.G. BACHEM et al. « The response of rat liver perisinusoidal lipocytes to polypeptide growth regulator changes with their transdifferentiation into myofibroblast-like cells in culture ». In : *Journal of hepatology* 18 (1 1993), p. 40-52.
- [18] Max G BACHEM et al. « The response of rat liver perisinusoidal lipocytes to polypeptide growth regulator changes with their transdifferentiation into myofibroblast-like cells in culture ». In : *Journal of hepatology* 18.1 (1993), p. 40-52.
- [19] Cristina BACIU et al. « Artificial intelligence applied to omics data in liver diseases : Enhancing clinical predictions ». In : *Frontiers in Artificial Intelligence* 5 (2022), p. 1050439.
- [20] G. BAFFY. « The impact of network medicine in gastroenterology and hepatology ». In : *Clin Gastroenterol Hepatol* 11.10 (oct. 2013), p. 1240-1244.
- [21] Ruth E BAKER, EA GAFFNEY et PK MAINI. « Partial differential equations for self-organization in cellular and developmental biology ». In : *Nonlinearity* 21.11 (2008), R251.

-
- [22] Paolo BALDAN et al. « Petri nets for modelling metabolic pathways : a survey ». In : *Natural Computing* 9 (2010), p. 955-989.
- [23] R. C. BENYON et M. J. ARTHUR. « Extracellular matrix degradation and the role of hepatic stellate cells ». In : *Semin Liver Dis* 21.3 (août 2001), p. 373-384.
- [24] Michael L BLINOV et al. « BioNetGen : software for rule-based modeling of signal transduction based on the interactions of molecular domains ». In : *Bioinformatics* 20.17 (2004), p. 3289-3291.
- [25] Nikolay M BORISOV et al. « Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains ». In : *Biosystems* 83.2-3 (2006), p. 152-166.
- [26] Matthieu BOUGUÉON et al. « The Rule-Based Model Approach : A Kappa Model for Hepatic Stellate Cells Activation by TGF β 1 ». In : *Systems Biology Modelling and Analysis : Formal Bioinformatics Methods and Tools* (2022), p. 69-126.
- [27] Katia BOURD-BOITTIN et al. « Protease profiling of liver fibrosis reveals the ADAM metallopeptidase with thrombospondin type 1 motif, 1 as a central activator of transforming growth factor beta ». In : *Hepatology* 54.6 (2011), p. 2173-2184.
- [28] Pierre BOUTILLIER, Ioana CRISTESCU et Jérôme FERET. « Counters in Kappa : Semantics, Simulation, and Static Analysis ». In : *Programming Languages and Systems - 28th European Symposium on Programming, ESOP 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings*. Sous la dir. de Luís CAIRES. T. 11423. Lecture Notes in Computer Science. Springer, 2019, p. 176-204. DOI : 10.1007/978-3-030-17184-1_7. URL : https://doi.org/10.1007/978-3-030-17184-1_7.
- [29] Pierre BOUTILLIER, Thomas EHRHARD et Jean KRIVINE. « Incremental Update for Graph Rewriting ». In : *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*. Sous la dir. d'Hongseok YANG. T. 10201. Lecture Notes in Computer Science. Springer, 2017, p. 201-228. DOI : 10.1007/978-3-662-54434-1_8. URL : https://doi.org/10.1007/978-3-662-54434-1_8.

-
- [30] Pierre BOUTILLIER, Aurélie Faure de PEBEYRE et Jérôme FERET. « Proving the absence of unbounded polymers in rule-based models ». In : *Electronic Notes in Theoretical Computer Science* 350 (2020), p. 33-56.
- [31] Pierre BOUTILLIER et al. « Kappa tools reference manual ». In : (2018).
- [32] Pierre BOUTILLIER et al. « KaSa : A Static Analyzer for Kappa ». In : *Computational Methods in Systems Biology - 16th International Conference, CMSB 2018, Brno, Czech Republic, September 12-14, 2018, Proceedings*. Sous la dir. de Milan CESKA et David SAFRÁNEK. T. 11095. Lecture Notes in Computer Science. Springer, 2018, p. 285-291. URL : https://doi.org/10.1007/978-3-319-99429-1%5C_17.
- [33] Pierre BOUTILLIER et al. « Kasa : a static analyzer for kappa ». In : *International Conference on Computational Methods in Systems Biology*. Springer. 2018, p. 285-291.
- [34] Pierre BOUTILLIER et al. « The Kappa platform for rule-based modeling ». In : *Bioinformatics* 34.13 (2018), p. i583-i592.
- [35] George EP BOX. « Robustness in the strategy of scientific model building ». In : *Robustness in statistics*. Elsevier, 1979, p. 201-236.
- [36] B. N. BROWN et al. « An agent-based model of inflammation and fibrosis following particulate exposure in the lung ». In : *Math Biosci* 231.2 (juin 2011), p. 186-196.
- [37] Erine H BUDI et al. « TGF- β as a driver of fibrosis : physiological roles and therapeutic opportunities ». In : *The Journal of pathology* 254.4 (2021), p. 358-373.
- [38] S. R. CALIARI et al. « Gradually softening hydrogels for modeling hepatic stellate cell behavior during fibrosis regression ». In : *Integr Biol (Camb)* 8.6 (juin 2016), p. 720-728.
- [39] Lara CAMPANA et John P IREDALE. « Regression of liver fibrosis ». In : *Seminars in liver disease*. T. 58. 01. Thieme Medical Publishers. 2017, p. 001-010.
- [40] Ferdinanda CAMPORESI, Jérôme FERET et Kim Quyên LÝ. « KaDE : A Tool to Compile Kappa Rules into (Reduced) ODE Models ». In : *Computational Methods in Systems Biology - 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27-29, 2017, Proceedings*. Sous la dir. de Jérôme FERET et Heinz KOEPL. T. 10545. Lecture Notes in Computer Science. Springer, 2017, p. 291-299.

DOI : 10.1007/978-3-319-67471-1_18. URL : https://doi.org/10.1007/978-3-319-67471-1%5C_18.

- [41] Ferdinanda CAMPORESI, Jérôme FERET et Kim Quyên LỖ. « KaDE : A Tool to Compile Kappa Rules into (Reduced) ODE Models ». In : *Computational Methods in Systems Biology : 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27–29, 2017, Proceedings 15*. Springer. 2017, p. 291-299.
- [42] Geraldine CELLIÈRE et al. « The plasticity of TGF- β signaling ». In : *BMC systems biology* 5 (2011), p. 1-13.
- [43] Lian En CHAI et al. « A review on the computational approaches for gene regulatory network construction ». In : *Computers in biology and medicine* 48 (2014), p. 55-65.
- [44] Ovijit CHAUDHURI et al. « Effects of extracellular matrix viscoelasticity on cellular behaviour ». In : *Nature* 584 (2020), p. 535-546.
- [45] Bastien CHOPARD et Michel DROZ. « Cellular automata ». In : *Modelling of Physical* (1998), p. 6-13.
- [46] B. CHRIST et al. « Computational Modeling in Liver Surgery ». In : *Front Physiol* 8 (2017), p. 906.
- [47] Seung-Wook CHUNG et al. « Quantitative modeling and analysis of the transforming growth factor β signaling pathway ». In : *Biophysical journal* 96.5 (2009), p. 1733-1750.
- [48] Lily A CHYLEK et al. « Modeling for (physical) biologists : an introduction to the rule-based approach ». In : *Physical biology* 12.4 (2015), p. 045007.
- [49] Lily A CHYLEK et al. « Rule-based modeling : a computational approach for studying biomolecular site dynamics in cell signaling systems ». In : *Wiley Interdisciplinary Reviews : Systems Biology and Medicine* 6.1 (2014), p. 13-36.
- [50] Federica CIOCCHETTA et Jane HILLSTON. « Bio-PEPA : A framework for the modelling and analysis of biological systems ». In : *Theoretical Computer Science* 410.33 – 34 (2009). Concurrent Systems Biology : To Nadia Busi (1968–2007), p. 3065-3084.
- [51] David C CLARKE, MD BETTERTON et Xuedong LIU. « Systems theory of Smad signalling ». In : *IEE Proceedings-Systems Biology* 153.6 (2006), p. 412-424.

-
- [52] Max COCKERILL, Michelle K RIGOZZI et Eugene M TARENTJEV. « Mechanosensitivity of the 2nd kind : TGF- β mechanism of cell sensing the substrate stiffness ». In : *PLoS One* 10.10 (2015), e0139959.
- [53] Paul R COHEN. « DARPA's Big Mechanism program ». In : *Physical biology* 12.4 (2015), p. 045008.
- [54] Daniel COOK, Babatunde A OGUNNAIKE et Rajanikanth VADIGEPALLI. « Systems analysis of non-parenchymal cell modulation of liver repair across multiple regeneration modes ». In : *BMC systems biology* 9 (2015), p. 1-24.
- [55] Jean COQUET et al. « Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering : Application to TGF- β Signaling ». In : *Computational Methods in Systems Biology - 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27-29, 2017, Proceedings*. Sous la dir. de Jérôme FERET et Heinz KOEPL. T. 10545. Lecture Notes in Computer Science. Springer, 2017, p. 91-107. DOI : 10.1007/978-3-319-67471-1_6. URL : https://doi.org/10.1007/978-3-319-67471-1%5C_6.
- [56] Marie CUVELLIER et al. « 3D culture of HepaRG cells in GelMa and its application to bioprinting of a multicellular hepatic model ». In : *Biomaterials* 269 (2021), p. 120611. ISSN : 0142-9612. DOI : <https://doi.org/10.1016/j.biomaterials.2020.120611>. URL : <https://www.sciencedirect.com/science/article/pii/S0142961220308577>.
- [57] Gennaro D'AMICO et al. « The potential role of machine learning in modelling advanced chronic liver disease ». In : *Digestive and Liver Disease* (2022).
- [58] Vincent DANOS et Cosimo LANEVE. « Formal molecular biology ». In : *Theor. Comput. Sci.* 325.1 (2004), p. 69-110. DOI : 10.1016/j.tcs.2004.03.065. URL : <https://doi.org/10.1016/j.tcs.2004.03.065>.
- [59] Vincent DANOS et Cosimo LANEVE. « Graphs for core molecular biology ». In : *International Conference on Computational Methods in Systems Biology*. Springer, 2003, p. 34-46.
- [60] Vincent DANOS et al. « Graphs, rewriting and pathway reconstruction for rule-based models ». In : *FSTTCS 2012-IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*. T. 18. 2012, p. 276-288.

-
- [61] Vincent DANOS et al. « Rule-based modelling of cellular signalling ». In : *International conference on concurrency theory*. Springer. 2007, p. 17-41.
- [62] Vincent DANOS et al. « Rule-based modelling, symmetries, refinements ». In : *International Workshop on Formal Methods in Systems Biology*. Springer. 2008, p. 103-122.
- [63] Vincent DANOS et al. « Scalable Simulation of Cellular Signaling Networks ». In : *Programming Languages and Systems, 5th Asian Symposium, APLAS 2007, Singapore, November 29-December 1, 2007, Proceedings*. Sous la dir. de Zhong SHAO. T. 4807. Lecture Notes in Computer Science. Springer, 2007, p. 139-157. DOI : 10.1007/978-3-540-76637-7\10. URL : https://doi.org/10.1007/978-3-540-76637-7%5C_10.
- [64] MIT Critical DATA et al. « Sensitivity analysis and model validation ». In : *Secondary analysis of electronic health records* (2016), p. 263-271.
- [65] Samuele DE MINICIS et al. « Gene Expression Profiles During Hepatic Stellate Cell Activation in Culture and In Vivo ». In : *Gastroenterology* 132 (juin 2007), p. 1937-46. DOI : 10.1053/j.gastro.2007.02.033.
- [66] V. DE SMET et al. « Initiation of hepatic stellate cell activation extends into chronic liver disease ». In : *Cell Death Dis* 12.12 (nov. 2021), p. 1110.
- [67] Charlotte DEBBAUT. « Multi-level modelling of hepatic perfusion in support of liver transplantation strategies ». In : *Annual Meeting of the Belgian Transplantation Society*. 2014, p. 16-19.
- [68] Tadeáš DĚD et al. « Formal biochemical space with semantics in Kappa and BNGL ». In : *Electronic Notes in Theoretical Computer Science* 326 (2016), p. 27-49.
- [69] Lorenzo DEMATTÉ, Corrado PRIAMI et Alessandro ROMANEL. « The BlenX Language : A Tutorial ». In : *Formal Methods for Computational Systems Biology : 8th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2008 Bertinoro, Italy, June 2-7, 2008 Advanced Lectures*. Sous la dir. de Marco BERNARDO, Pierpaolo DEGANO et Gianluigi ZAVATTARO. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008, p. 313-365. URL : http://dx.doi.org/10.1007/978-3-540-68894-5%5C_9.

-
- [70] Peter J DENNING et Ted G LEWIS. « Exponential laws of computing growth ». In : *Communications of the ACM* 60.1 (2016), p. 54-65.
- [71] Andreas DEUTSCH et Uwe BÖRNER. « Cellular automaton modeling of pattern formation in interacting cell systems ». In : *Advances in Parallel Computing*. T. 13. Elsevier, 2004, p. 695-704.
- [72] Bedair DEWIDAR et al. « TGF- β in hepatic stellate cell activation and liver fibrogenesis—updated 2019 ». In : *Cells* 8.11 (2019), p. 1419.
- [73] G. M. DI GUGLIELMO et al. « Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover ». In : *Nat Cell Biol* 5.5 (mai 2003), p. 410-421.
- [74] *Dictionnaire médical de l'Académie de Médecine – version 2023*. <https://www.academie-medecine.fr/le-dictionnaire/index.php>.
- [75] Steven DOOLEY et al. « Modulation of transforming growth factor β response and signaling during transdifferentiation of rat hepatic stellate cells to myofibroblasts ». In : *Hepatology* 31.5 (2000), p. 1094-1106.
- [76] J. S. DUFFIELD et al. « Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair ». In : *J Clin Invest* 115.1 (jan. 2005), p. 56-65.
- [77] Joyeeta DUTTA-MOSCATO et al. « A multiscale agent-based in silico model of liver fibrosis progression ». In : *Frontiers in bioengineering and biotechnology* 2 (2014), p. 18.
- [78] John W EATON et al. *GNU Octave version 4.0. 0 manual : a high-level interactive language for numerical computations* In : *CreateSpace Independent Publishing Platform*. 2015.
- [79] A. EHRLICH et al. « Challenges and Opportunities in the Design of Liver-on-Chip Microdevices ». In : *Annu Rev Biomed Eng* 21 (juin 2019), p. 219-239.
- [80] Adil EL TAGHDOUINI et al. « In vitro reversion of activated primary human hepatic stellate cells ». In : *Fibrogenesis & tissue repair* 8.1 (2015), p. 1-15.
- [81] Adil EL TAGHDOUINI et al. « In vitro reversion of activated primary human hepatic stellate cells ». In : *Fibrogenesis and tissue repair* 8 (août 2015), p. 14. DOI : 10.1186/s13069-015-0031-z.

-
- [82] A. M. ELSHARKAWY, F. OAKLEY et D. A. MANN. « The role and regulation of hepatic stellate cell apoptosis in reversal of liver fibrosis ». In : *Apoptosis* 10.5 (oct. 2005), p. 927-939.
- [83] Agner Krarup ERLANG. *Sandsynlighedsregning og Telefonsamtaler*. 1909. URL : <https://books.google.fr/books?id=xBJZQwAACAAJ>.
- [84] Leonhard EULER. *Harmonie entre les principes généraux de repos et de mouvement de M. de Maupertuis*. Haude & Spener, 1753.
- [85] James R FAEDER, Michael L BLINOV et William S HLAVACEK. « Rule-based modeling of biochemical systems with BioNetGen ». In : *Systems biology* (2009), p. 113-167.
- [86] James R. FAEDER et al. « Rule-based modeling of biochemical networks ». In : *Complexity* 10.4 (2005), p. 22-41. ISSN : 1099-0526. DOI : 10.1002/cplx.20074. URL : <http://dx.doi.org/10.1002/cplx.20074>.
- [87] M. FAROOQ et al. « Switching to Regular Diet Partially Resolves Liver Fibrosis Induced by High-Fat, High-Cholesterol Diet in Mice ». In : *Nutrients* 14.2 (jan. 2022).
- [88] Jérôme FERET. « An algebraic approach for inferring and using symmetries in rule-based models ». In : *Electronic Notes in Theoretical Computer Science* 316 (2015), p. 45-65.
- [89] Jérôme FERET. « Analyse statique et réduction de modèles de voies de signalisation intracellulaire ». In : *Informatique Mathématique Une photographie en 2023* (2023), p. 67.
- [90] Jérôme FERET. « Analyses des motifs accessibles dans les modèles Kappa ». In : (2020).
- [91] Jerome FERET, Heinz KOEPPL et Tatjana PETROV. « Stochastic fragments : A framework for the exact reduction of the stochastic semantics of rule-based models ». In : *International Journal of Software and Informatics* 7.ARTICLE (2013), p. 527-604.
- [92] Jérôme FERET et al. « Internal coarse-graining of molecular systems ». In : *Proceedings of the National Academy of Sciences of the United States of America* (avr. 2009). URL : <http://www.pnas.org/content/early/2009/04/02/0809908106.abstract>.

-
- [93] Xenia FICHT et Matteo IANNACONE. « Immune surveillance of the liver by T cells ». In : *Science immunology* 5.51 (2020), eaba2351.
- [94] Avner FRIEDMAN et Wenrui HAO. « Mathematical modeling of liver fibrosis ». In : *Mathematical Biosciences & Engineering* 14.1 (2016), p. 143-164.
- [95] Avner FRIEDMAN et Nourridine SIEWE. « Chronic hepatitis B virus and liver fibrosis : A mathematical model ». In : *PLoS One* 13.4 (2018), e0195037.
- [96] Scott L FRIEDMAN, Glenn YAMASAKI et Linda WONG. « Modulation of transforming growth factor beta receptors of rat lipocytes during the hepatic wound healing response. Enhanced binding and reduced gene expression accompany cellular activation in culture and in vivo. » In : *Journal of Biological Chemistry* 269.14 (1994), p. 10551-10558.
- [97] L. GAILHOUSTE et al. « Fibrillar collagen scoring by second harmonic microscopy : a new tool in the assessment of liver fibrosis ». In : *J Hepatol* 52.3 (mars 2010), p. 398-406.
- [98] R. GAO, H. TANG et J. MAO. « Programmed Cell Death in Liver Fibrosis ». In : *J Inflamm Res* 16 (2023), p. 3897-3910.
- [99] Abhishek GARG et al. « Synchronous versus asynchronous modeling of gene regulatory networks ». In : *Bioinformatics* 24.17 (2008), p. 1917-1925.
- [100] Daniel T GILLESPIE. « A diffusional bimolecular propensity function ». In : *The Journal of chemical physics* 131.16 (2009).
- [101] Daniel T GILLESPIE. « A general method for numerically simulating the stochastic time evolution of coupled chemical reactions ». In : *Journal of computational physics* 22.4 (1976), p. 403-434.
- [102] Daniel T GILLESPIE. « A rigorous derivation of the chemical master equation ». In : *Physica A : Statistical Mechanics and its Applications* 188.1-3 (1992), p. 404-425.
- [103] Daniel T GILLESPIE. « Approximate accelerated stochastic simulation of chemically reacting systems ». In : *The Journal of chemical physics* 115.4 (2001), p. 1716-1733.
- [104] Daniel T GILLESPIE. « Exact stochastic simulation of coupled chemical reactions ». In : *The journal of physical chemistry* 81.25 (1977), p. 2340-2361.

-
- [105] Leon GLASS et Stuart A KAUFFMAN. « The logical analysis of continuous, non-linear biochemical control networks ». In : *Journal of theoretical Biology* 39.1 (1973), p. 103-129.
- [106] Chad M GLEN, Melissa L KEMP et Eberhard O VOIT. « Agent-based modeling of morphogenetic systems : Advantages and challenges ». In : *PLoS computational biology* 15.3 (2019), e1006577.
- [107] O. GOVAERE et al. « Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis ». In : *Sci Transl Med* 12.572 (déc. 2020).
- [108] Melanie GRIEB et al. « Predicting variabilities in cardiac gene expression with a boolean network incorporating uncertainty ». In : *PloS one* 10.7 (2015), e0131832.
- [109] Benjamin M GYORI et al. « From word models to executable models of signaling networks using automated assembly ». In : *Molecular systems biology* 13.11 (2017), p. 954.
- [110] CALCULATOR.NET. 2008-2021. URL : <https://www.calculator.net/half-life-calculator.htm>.
- [111] L. HAMMERICH et F. TACKE. « Hepatic inflammatory responses in liver fibrosis ». In : *Nat Rev Gastroenterol Hepatol* (juill. 2023).
- [112] Wenrui HAO, Clay MARSH et Avner FRIEDMAN. « A mathematical model of idiopathic pulmonary fibrosis ». In : *PLoS One* 10.9 (2015), e0135097.
- [113] Wenrui HAO, Brad H ROVIN et Avner FRIEDMAN. « Mathematical model of renal interstitial fibrosis ». In : *Proceedings of the National Academy of Sciences* 111.39 (2014), p. 14193-14198.
- [114] Wenrui HAO et al. « Mathematical model of chronic pancreatitis ». In : *Proceedings of the National Academy of Sciences* 114.19 (2017), p. 5011-5016.
- [115] L. A. HARRIS et al. « BioNetGen 2.2 : advances in rule-based modeling ». In : *Bioinformatics* 32.21 (nov. 2016), p. 3366-3368.
- [116] Eric L HASELTINE et James B RAWLINGS. « Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics ». In : *The Journal of chemical physics* 117.15 (2002), p. 6959-6969.

-
- [117] Dieter HÄUSSINGER et Claus KORDES. « Space of Disse : a stem cell niche in the liver ». In : *Biological chemistry* 401.1 (2019), p. 81-95.
- [118] Paul L HERMONAT et al. « Mechanism of action and delivery possibilities for TGF β 1 in the treatment of myocardial ischemia ». In : *Cardiovascular research* 74.2 (2007), p. 235-243.
- [119] Boris HINZ. « The extracellular matrix and transforming growth factor- β 1 : Tale of a strained relationship ». In : *Matrix Biology* 47 (2015), p. 54-65. ISSN : 0945-053X. DOI : <https://doi.org/10.1016/j.matbio.2015.05.006>. URL : <https://www.sciencedirect.com/science/article/pii/S0945053X15001055>.
- [120] William S HLAVACEK et al. « Rules for modeling signal-transduction systems ». In : *Science's STKE* 2006.344 (2006), re6-re6.
- [121] John H HOLLAND et John H MILLER. « Artificial adaptive agents in economic theory ». In : *The American economic review* 81.2 (1991), p. 365-370.
- [122] J HONERKAMP. « Stochastic Dynamical Systems : Concepts ». In : *Numerical Methods, Data Analysis, VCH* (1994).
- [123] Martin HOPFENSITZ et al. « Attractors in Boolean networks : a tutorial ». In : *Computational Statistics* 28 (2013), p. 19-36.
- [124] R. HUANG et al. « Interferon-alpha plus adefovir combination therapy versus interferon-alpha monotherapy for chronic hepatitis B treatment : A meta-analysis ». In : *Hepatol Res* 43.10 (oct. 2013), p. 1040-1051.
- [125] Sui HUANG. « Non-genetic heterogeneity of cells in development : more than just noise ». In : *Development* 136.23 (2009), p. 3853-3862.
- [126] Michael HUCKA et al. « The Systems Biology Markup Language (SBML) : language specification for level 3 version 1 core ». In : *Journal of integrative bioinformatics* 12.2 (2015), p. 382-549.
- [127] Richard HYNES et Alexandra NABA. « Overview of the Matrisome—An Inventory of Extracellular Matrix Constituents and Functions ». In : *Cold Spring Harbor perspectives in biology* 4 (sept. 2011), a004903. DOI : 10.1101/cshperspect.a004903.
- [128] The MathWorks INC. *MATLAB version : 9.13.0 (R2022b)*. Natick, Massachusetts, United States, 2022. URL : <https://www.mathworks.com>.

-
- [129] Edward L INCE. *Ordinary differential equations*. Courier Corporation, 1956.
- [130] K. IWASAKO, D. A. BRENNER et T. KISSELEVA. « What's new in liver fibrosis? The origin of myofibroblasts in liver fibrosis ». In : *J Gastroenterol Hepatol* 27 Suppl 2.Suppl 2 (mars 2012), p. 65-68.
- [131] P. L. M. JANSEN et al. « Editorial : Systems Biology and Bioinformatics in Gastroenterology and Hepatology ». In : *Front Physiol* 10 (2019), p. 1438.
- [132] Won-Il JEONG, Ogyi PARK et Bin GAO. « Abrogation of the antifibrotic effects of natural killer cells/interferon- γ contributes to alcohol acceleration of liver fibrosis ». In : *Gastroenterology* 134.1 (2008), p. 248-258.
- [133] Won-Il JEONG et al. « STAT1 inhibits liver fibrosis in mice by inhibiting stellate cell proliferation and stimulating NK cell cytotoxicity ». In : *Hepatology* 44.6 (2006), p. 1441-1451.
- [134] Yu-Fang JIN et al. « Combining experimental and mathematical modeling to reveal mechanisms of macrophage-dependent left ventricle remodeling ». In : *BMC systems biology* 5 (mai 2011), p. 60. DOI : 10.1186/1752-0509-5-60.
- [135] Yaochu JIN. « Fuzzy modeling of high-dimensional systems : complexity reduction and interpretability improvement ». In : *IEEE Transactions on Fuzzy Systems* 8.2 (2000), p. 212-221.
- [136] Mathias JOHN et al. « Biochemical Reaction Rules with Constraints ». In : *Programming Languages and Systems - 20th European Symposium on Programming, ESOP 2011, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2011, Saarbrücken, Germany, March 26-April 3, 2011. Proceedings*. T. 6602. Lecture Notes in Computer Science. Springer, 2011, p. 338-357. URL : <http://dx.doi.org/10.1007/978-3-642-19718-5>.
- [137] Jerald B JOHNSON et Kristian S OMLAND. « Model selection in ecology and evolution ». In : *Trends in ecology & evolution* 19.2 (2004), p. 101-108.
- [138] Valerie G KALTER et Arnold R BRODY. « Receptors for Transforming Growth Factor-S (IGF-) on Rat Lung Fibroblasts Have Higher Affinity for IDF-, 81 than for IDF- ». In : *Am. J. Respir. Cell Mol. Biol.* 4 (1991), p. 397-407.
- [139] G. KARAGIANNIS, W. HAO et G. LIN. « Calibrations and validations of biological models with an application on the renal fibrosis ». In : *Int J Numer Method Biomed Eng* 36.5 (mai 2020), e3329.

-
- [140] Cliff C KERR et al. « Covasim : an agent-based model of COVID-19 dynamics and interventions ». In : *PLOS Computational Biology* 17.7 (2021), e1009149.
- [141] Shabnam KHATIBI et al. « Mathematical model of TGF- β signalling : feedback coupling is consistent with signal switching ». In : *BMC systems biology* 11 (2017), p. 1-15.
- [142] B. S. KIM et al. « Decellularized Extracellular Matrix-based Bioinks for Engineering Tissue- and Organ-specific Microenvironments ». In : *Chem Rev* 120.19 (oct. 2020), p. 10608-10661.
- [143] T. KISSELEVA et al. « Myofibroblasts revert to an inactive phenotype during regression of liver fibrosis ». In : *Proc Natl Acad Sci U S A* 109.24 (juin 2012), p. 9448-9453.
- [144] Tatiana KISSELEVA et David BRENNER. « Molecular and cellular mechanisms of liver fibrosis and its regression ». In : *Nature Reviews Gastroenterology & Hepatology* 18.3 (2021), p. 151-166.
- [145] Tatiana KISSELEVA et al. « Myofibroblasts revert to an inactive phenotype during regression of liver fibrosis ». In : *Proceedings of the National Academy of Sciences* 109.24 (2012), p. 9448-9453. ISSN : 0027-8424. DOI : 10.1073/pnas.1201840109. eprint : <https://www.pnas.org/content/109/24/9448.full.pdf>. URL : <https://www.pnas.org/content/109/24/9448>.
- [146] Hiroaki KITANO. « Computational systems biology ». In : *Nature* 420.6912 (2002), p. 206-210.
- [147] Hiroaki KITANO. « Systems biology : a brief overview ». In : *science* 295.5560 (2002), p. 1662-1664.
- [148] Edda KLIPP et al. " *Systems Biology : A Textbook*. 2009.
- [149] Yuzo KODA et al. « CD8+ tissue-resident memory T cells promote liver fibrosis resolution by inducing apoptosis of hepatic stellate cells ». In : *Nature communications* 12.1 (2021), p. 4474.
- [150] Agnes KÖHLER, Jean KRIVINE et Jakob VIDMAR. « A Rule-Based Model of Base Excision Repair ». In : *Computational Methods in Systems Biology*. Sous la dir. de Pedro MENDES, Joseph O. DADA et Kieran SMALLBONE. Cham : Springer International Publishing, 2014, p. 173-195. ISBN : 978-3-319-12982-2.

-
- [151] C. KORDES et al. « Hepatic stellate cells : current state and open questions ». In : *Biol Chem* 402.9 (août 2021), p. 1021-1032.
- [152] E. KOSTALLARI et al. « Stiffness is associated with hepatic stellate cell heterogeneity during liver fibrosis ». In : *Am J Physiol Gastrointest Liver Physiol* 322.2 (fév. 2022), G234-G246.
- [153] Oliver KRENKEL et al. « Single cell RNA sequencing identifies subsets of hepatic stellate cells and myofibroblasts in liver fibrosis ». In : *Cells* 8.5 (2019), p. 503.
- [154] Oliver KRENKEL et al. « Single Cell RNA Sequencing Identifies Subsets of Hepatic Stellate Cells and Myofibroblasts in Liver Fibrosis ». In : *Cells* 8 (mai 2019), p. 503. DOI : 10.3390/cells8050503.
- [155] Jean KRIVINE, Vincent DANOS et Arndt BENECKE. « Modelling Epigenetic Information Maintenance : A Kappa Tutorial ». In : *Computer Aided Verification*. Sous la dir. d'Ahmed BOUAJJANI et Oded MALER. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009, p. 17-32. ISBN : 978-3-642-02658-4.
- [156] V. KRIZHANOVSKY et al. « Senescence of activated stellate cells limits liver fibrosis ». In : *Cell* 134.4 (août 2008), p. 657-667.
- [157] D. LACHOWSKI et al. « Matrix stiffness modulates the activity of MMP-9 and TIMP-1 in hepatic stellate cells to perpetuate fibrosis ». In : *Sci Rep* 9.1 (mai 2019), p. 7299.
- [158] John Denholm LAMBERT. « Computational methods in ordinary differential equations ». In : *(No Title)* (1973).
- [159] Sébastien LÉGARÉ, Ievgeniia OSHURKO et Russel HARMER. « Biocuration and rule-based modelling of protein interaction networks in KAMI ». In : *Identification de marques épigénétiques marqueurs de prédisposition aux maladies métaboliques* (2018), p. 243.
- [160] J. LEONARD-DUKE et al. « Multi-scale models of lung fibrosis ». In : *Matrix Biol* 91-92 (sept. 2020), p. 35-50.
- [161] Huipeng LI et al. « Computational analysis reveals the coupling between bistability and the sign of a feedback loop in a TGF- β 1 activation model ». In : *BMC Systems Biology* 11.7 (2017), p. 91-99.
- [162] Yang LI et al. « A review of cellular automata models for crowd evacuation ». In : *Physica A : Statistical Mechanics and its Applications* 526 (2019), p. 120752.

-
- [163] Xiao LIU et al. « Reversibility of Liver Fibrosis and Inactivation of Fibrogenic Myofibroblasts ». In : *Current pathobiology reports* 1 (sept. 2013), p. 209-214. DOI : 10.1007/s40139-013-0018-7.
- [164] Xiao LIU et al. « Reversibility of liver fibrosis and inactivation of fibrogenic myofibroblasts ». In : *Current pathobiology reports* 1.3 (2013), p. 209-214.
- [165] Carlos F LOPEZ et al. « Programming biological models in Python using PySB ». In : *Molecular systems biology* 9.1 (2013), p. 646.
- [166] Pavel LOSKOT, Komlan ATITEY et Lyudmila MIHAYLOVA. « Comprehensive review of models and methods for inferences in bio-chemical reaction networks ». In : *Frontiers in genetics* (2019), p. 549.
- [167] Michael I LOVE, Wolfgang HUBER et Simon ANDERS. « Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 ». In : *Genome biology* 15.12 (2014), p. 1-21.
- [168] Philippe LUCARELLI et al. « Resolving the combinatorial complexity of smad protein complex formation and its link to gene expression ». In : *Cell systems* 6.1 (2018), p. 75-89.
- [169] Russette M LYONS et al. « Differential binding of transforming growth factor- β 1, β 2, and β 3 by fibroblasts and epithelial cells measured by affinity cross-linking of cell surface receptors ». In : *Molecular Endocrinology* 5.12 (1991), p. 1887-1896.
- [170] Daniel MACHADO et al. « Modeling formalisms in systems biology ». In : *AMB express* 1.1 (2011), p. 1-14.
- [171] Shev MACNAMARA et al. « Stochastic chemical kinetics and the total quasi-steady-state assumption : application to the stochastic simulation algorithm and chemical master equation ». In : *The Journal of chemical physics* 129.9 (2008).
- [172] Andrea MALANDRINO et al. « Non-Elastic Remodeling of the 3D Extracellular Matrix by Cell-Generated Forces ». In : *bioRxiv* (2017). DOI : 10.1101/193458. eprint : <https://www.biorxiv.org/content/early/2017/09/27/193458.full.pdf>. URL : <https://www.biorxiv.org/content/early/2017/09/27/193458>.
- [173] Luca MARCHETTI, Corrado PRIAMI et Vo Hong THANH. « HRSSA—Efficient hybrid stochastic simulation for spatially homogeneous biochemical reaction networks ». In : *Journal of Computational Physics* 317 (2016), p. 301-317.

-
- [174] A. MARDINOGLU et al. « Systems biology in hepatology : approaches and applications ». In : *Nat Rev Gastroenterol Hepatol* 15.6 (juin 2018), p. 365-377.
- [175] Simeone MARINO et al. « A methodology for performing global uncertainty and sensitivity analysis in systems biology ». In : *Journal of Theoretical Biology* 254.1 (2008), p. 178-196. ISSN : 0022-5193. DOI : <https://doi.org/10.1016/j.jtbi.2008.04.011>. URL : <https://www.sciencedirect.com/science/article/pii/S0022519308001896>.
- [176] F. J. MARTIN et al. « Ensembl 2023 ». In : *Nucleic Acids Res* 51.D1 (jan. 2023), p. D933-D941.
- [177] Joan MASSAGUE et Betsy LIKE. « Cellular receptors for type beta transforming growth factor. Ligand binding and affinity labeling in human and rodent cell lines. » In : *Journal of Biological Chemistry* 260.5 (1985), p. 2636-2645.
- [178] V. L. MASSEY et al. « The hepatic "matrisome" responds dynamically to injury : Characterization of transitional changes to the extracellular matrix in mice ». In : *Hepatology* 65.3 (mars 2017), p. 969-982.
- [179] Hector F MEDINA ABARCA. « A combinatorial view of canonical Wnt signaling ». Thèse de doct. 2020.
- [180] Martin MEIER-SCHELLERSHEIM, Iain DC FRASER et Frederick KLAUSCHEN. « Multiscale modeling for biologists ». In : *Wiley Interdisciplinary Reviews : Systems Biology and Medicine* 1.1 (2009), p. 4-14.
- [181] Pontus MELKE et al. « A rate equation approach to elucidate the kinetics and robustness of the TGF- β pathway ». In : *Biophysical journal* 91.12 (2006), p. 4368-4380.
- [182] Michael B MONAGAN et al. « Maple 10 Programming Guide Maplesoft ». In : *Waterloo, Ontario* (2005).
- [183] G. MONTOSI et al. « Spatial and temporal dynamics of hepatic stellate cell activation during oxidant-stress-induced fibrogenesis ». In : *Am J Pathol* 152.5 (mai 1998), p. 1319-1326.
- [184] V. K. MOOTHA et al. « PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes ». In : *Nat Genet* 34.3 (juill. 2003), p. 267-273.

-
- [185] Ehsanollah MORADI, Sasan JALILI-FIROOZINEZHAD et Mehran SOLATI-HASHJIN. « Microfluidic organ-on-a-chip models of human liver tissue ». In : *Acta Biomaterialia* 116 (2020), p. 67-83. ISSN : 1742-7061. DOI : <https://doi.org/10.1016/j.actbio.2020.08.041>. URL : <https://www.sciencedirect.com/science/article/pii/S1742706120305110>.
- [186] Jonathon M. MUNCIE et Valerie M. WEAVER. « Chapter One - The Physical and Biochemical Properties of the Extracellular Matrix Regulate Cell Fate ». In : *Extracellular Matrix and Egg Coats*. Sous la dir. d'Eveline S. LITSCHER et Paul M. WASSARMAN. T. 130. Current Topics in Developmental Biology. Academic Press, 2018, p. 1-37. DOI : <https://doi.org/10.1016/bs.ctdb.2018.02.002>. URL : <https://www.sciencedirect.com/science/article/pii/S0070215318300346>.
- [187] Jana MUSILOVA et Karel SEDLAR. « Tools for time-course simulation in systems biology : a brief overview ». In : *Briefings in Bioinformatics* 22.5 (2021), bbaa392.
- [188] MWJM MUSTERS et NAW van RIEL. « Analysis of the transforming growth factor- β pathway and extracellular matrix formation as a hybrid system ». In : *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. T. 2. IEEE. 2004, p. 2901-2904.
- [189] Jun NAKABAYASHI et Akira SASAKI. « A mathematical model of the stoichiometric control of Smad complex formation in TGF- β signal transduction pathway ». In : *Journal of theoretical biology* 259.2 (2009), p. 389-403.
- [190] David NAM et al. « Artificial intelligence in liver diseases : Improving diagnostics, prognostics and response prediction ». In : *JHEP Reports* 4.4 (2022), p. 100443.
- [191] Natasha A NEOGI. « Dynamic partitioning of large discrete event biological systems for hybrid simulation and analysis ». In : *International Workshop on Hybrid Systems : Computation and Control*. Springer. 2004, p. 463-476.
- [192] Daniel NICKLAS et Leonor SAIZ. « Computational modelling of Smad-mediated negative feedback and crosstalk in the TGF- β superfamily network ». In : *Journal of The Royal Society Interface* 10.86 (2013), p. 20130363.
- [193] Felipe NUNEZ et al. « A Rule-based Model for a Stochastic Simulation of a Zombie Outbreak ». In : ().

-
- [194] N. OHTANI. « The roles and mechanisms of senescence-associated secretory phenotype (SASP) : can it be controlled by senolysis ? » In : *Inflamm Regen* 42.1 (avr. 2022), p. 11.
- [195] Miles S OKINO et Michael L MAVROVOUNIOTIS. « Simplification of mathematical models of chemical reaction systems ». In : *Chemical reviews* 98.2 (1998), p. 391-408.
- [196] Brenda de OLIVEIRA DA SILVA, Letícia Ferreira RAMOS et Karen CM MORAES. « Molecular interplays in hepatic stellate cells : apoptosis, senescence, and phenotype reversion as cellular connections that modulate liver fibrosis ». In : *Cell biology international* 41.9 (2017), p. 946-959.
- [197] Abby L OLSEN et al. « Hepatic stellate cells require a stiff environment for myofibroblastic differentiation ». In : *American Journal of Physiology-Gastrointestinal and Liver Physiology* 301.1 (2011), G110-G118.
- [198] Andrés ORTIZ-MUÑOZ, Héctor F MEDINA-ABARCA et Walter FONTANA. « Combinatorial protein-protein interactions on a polymerizing scaffold ». In : *Proceedings of the National Academy of Sciences* 117.6 (2020), p. 2930-2937.
- [199] Rodrigo Alberto Santibáñez PALOMINOS. « Development of Rule-Based Models for Regulatory Mechanisms of Gene Expression and Bacterial Metabolism ». Thèse de doct. Pontificia Universidad Católica de Chile (Chile), 2020.
- [200] G. A. PAVLOPOULOS et al. « Using graph theory to analyze biological networks ». In : *BioData Min* 4 (avr. 2011), p. 10.
- [201] Tomas PEREZ-ACLE et al. « Stochastic simulation of multiscale complex systems with PISKaS : A rule-based approach ». In : *Biochemical and biophysical research communications* 498.2 (2018), p. 342-351.
- [202] Tatjana PETROV et al. « Long lived transients in gene regulation ». In : *Theoretical Computer Science* 893 (2021), p. 1-16.
- [203] Massimo PINZANI et Chandrashekhar R GANDHI. « History and early work ». In : *Stellate cells in health and disease*. Elsevier, 2015, p. 1-13.
- [204] Mark F. PITTENGER et al. « Mesenchymal stem cell perspective : cell biology to clinical progress. NPJ Regen Med ». In : (2019).

-
- [205] Carole J PROCTOR et Alison GARTLAND. « Simulated interventions to ameliorate age-related bone loss indicate the importance of timing ». In : *Frontiers in endocrinology* 7 (2016), p. 61.
- [206] Philip E PROTTER et Philip E PROTTER. *Stochastic differential equations*. Springer, 2005.
- [207] Svetlana RADAIEVA et al. « Natural killer cells ameliorate liver fibrosis by killing activated stellate cells in nkg2d-dependent and tumor necrosis factor-related apoptosis-inducing ligand-dependent manners ». In : *Gastroenterology* 130.2 (2006), p. 435-452.
- [208] Ovidiu RADULESCU, Sergei VAKULENKO et Dima GRIGORIEV. « Model reduction of biochemical reactions networks by tropical analysis methods ». In : *Mathematical Modelling of Natural Phenomena* 10.3 (2015), p. 124-138.
- [209] Ovidiu RADULESCU et al. « Reduction of dynamical biochemical reactions networks in computational biology ». In : *Frontiers in genetics* 3 (2012), p. 131.
- [210] P. RAMACHANDRAN et al. « Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis ». In : *Proc Natl Acad Sci U S A* 109.46 (nov. 2012), E3186-3195.
- [211] Venkatramana N REDDY, Michael L MAVROVOUNIOTIS, Michael N LIEBMAN et al. « Petri net representations in metabolic pathways. » In : *ISMB*. T. 93. 1993, p. 328-336.
- [212] Wolfgang REISIG. *Petri nets : an introduction*. T. 4. Springer Science & Business Media, 2012.
- [213] S. M. RIKARD et al. « Multiscale Coupling of an Agent-Based Model of Tissue Fibrosis and a Logic-Based Model of Intracellular Signaling ». In : *Front Physiol* 10 (2019), p. 1481.
- [214] Sara Brin ROSENTHAL et al. « Heterogeneity of HSC s in a Mouse Model of NASH ». In : *Hepatology* 74.2 (2021), p. 667-685.
- [215] A. D. ROUILLARD et J. W. HOLMES. « Coupled agent-based and finite-element models for predicting scar structure following myocardial infarction ». In : *Prog Biophys Mol Biol* 115.2-3 (août 2014), p. 235-243.

-
- [216] A. D. ROUILLARD et J. W. HOLMES. « Mechanical regulation of fibroblast migration and collagen remodelling in healing myocardial infarcts ». In : *J Physiol* 590.18 (sept. 2012), p. 4585-4602.
- [217] Michaël RUFF et al. « The disintegrin and metalloprotease ADAM12 is associated with TGF- β -induced epithelial to mesenchymal transition ». In : *PLoS One* 10.9 (2015), e0139179.
- [218] Takaoki SANEYASU, Riaz AKHTAR et Takao SAKAI. « Molecular Cues Guiding Matrix Stiffness in Liver Fibrosis ». In : *BioMed Research International* 2016 (jan. 2016), p. 1-11. DOI : 10.1155/2016/2646212.
- [219] Rodrigo SANTIBÁÑEZ, Daniel GARRIDO et Alberto JM MARTIN. « Atlas : automatic modeling of regulation of bacterial gene expression and metabolism using rule-based languages ». In : *Bioinformatics* 36.22-23 (2020), p. 5473-5480.
- [220] Y. SATO et al. « Resolution of fibrosis by siRNA HSP47 in vitamin A-coupled liposomes induces regeneration of chronically injured livers ». In : *J Gastroenterol Hepatol* 36.12 (déc. 2021), p. 3418-3428.
- [221] J. SCHLEICHER et al. « A theoretical study of lipid accumulation in the liver-implications for nonalcoholic fatty liver disease ». In : *Biochim Biophys Acta* 1841.1 (jan. 2014), p. 62-69.
- [222] Bernhard SCHMIERER et al. « Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system ». In : *Proceedings of the National Academy of Sciences* 105.18 (2008), p. 6608-6613.
- [223] B. SCHNABL et al. « Replicative senescence of activated human hepatic stellate cells is accompanied by a pronounced inflammatory but less fibrogenic phenotype ». In : *Hepatology* 37.3 (mars 2003), p. 653-664.
- [224] Bernd SCHNABL et al. « TAK1/JNK and p38 have opposite effects on rat hepatic stellate cells ». In : *Hepatology* 34.5 (2001), p. 953-963. ISSN : 0270-9139. DOI : <https://doi.org/10.1053/jhep.2001.28790>. URL : <https://www.sciencedirect.com/science/article/pii/S0270913901529845>.
- [225] D. SCHOLTEN et al. « The carbon tetrachloride model in mice ». In : *Lab Anim* 49.1 Suppl (avr. 2015), p. 4-11.

-
- [226] Julian D SCHWAB et al. « Concepts in Boolean network modeling : What do they all mean ? » In : *Computational and structural biotechnology journal* 18 (2020), p. 571-582.
- [227] Haruki SENOO, Yoshihiro MEZAKI et Mutsunori FUJIWARA. « The stellate cell system (vitamin A-storing cell system) ». In : *Anatomical science international* 92 (2017), p. 387-455.
- [228] E. SETTEN et al. « Understanding fibrosis pathogenesis via modeling macrophage-fibroblast interplay in immune-metabolic context ». In : *Nat Commun* 13.1 (oct. 2022), p. 6499.
- [229] Harish SHANKARAN et H Steven WILEY. « Smad signaling dynamics : insights from a parsimonious model ». In : *Science Signaling* 1.36 (2008), pe41-pe41.
- [230] Paul SHANNON et al. « Cytoscape : a software environment for integrated models of biomolecular interaction networks ». In : *Genome research* 13.11 (2003), p. 2498-2504.
- [231] Xinhao SHAO et al. « MatrisomeDB : the ECM-protein knowledge database ». In : *Nucleic Acids Research* 48.D1 (oct. 2019), p. D1136-D1144. ISSN : 0305-1048. DOI : 10.1093/nar/gkz849. eprint : <https://academic.oup.com/nar/article-pdf/48/D1/D1136/31697336/gkz849.pdf>. URL : <https://doi.org/10.1093/nar/gkz849>.
- [232] Jonathan Adam SHERRATT et JD MURRAY. « Mathematical analysis of a basic model for epidermal wound healing ». In : *Journal of mathematical biology* 29 (1991), p. 389-404.
- [233] Ilya SHMULEVICH et al. « Probabilistic Boolean networks : a rule-based uncertainty model for gene regulatory networks ». In : *Bioinformatics* 18.2 (2002), p. 261-274.
- [234] Giulia SIMONI et al. « A comparison of deterministic and stochastic approaches for sensitivity analysis in computational systems biology ». In : *Briefings in bioinformatics* 21.2 (2020), p. 527-540.
- [235] Giulia SIMONI et al. « Stochastic simulation algorithms for computational systems biology : Exact, approximate, and hybrid methods ». In : *Wiley Interdisciplinary Reviews : Systems Biology and Medicine* 11.6 (2019), e1459.
- [236] Oksana SOROKINA et al. « A simulator for spatially extended kappa models ». In : *Bioinformatics* 29.23 (2013), p. 3105.

-
- [237] Steven Nathaniel STEINWAY et al. « Combinatorial interventions inhibit TGF β -driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes ». In : *NPJ systems biology and applications* 1.1 (2015), p. 1-12.
- [238] Steven Nathaniel STEINWAY et al. « Network modeling of TGF β signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation ». In : *Cancer research* 74.21 (2014), p. 5963-5977.
- [239] Jette STRASEN et al. « Cell-specific responses to the cytokine TGF β are determined by variability in protein levels ». In : *Molecular systems biology* 14.1 (2018), e7733.
- [240] Walter A STRAUSS. *Partial differential equations : An introduction*. John Wiley & Sons, 2007.
- [241] K. STREET et al. « Slingshot : cell lineage and pseudotime inference for single-cell transcriptomics ». In : *BMC Genomics* 19.1 (juin 2018), p. 477.
- [242] T. H. SU, C. H. WU et J. H. KAO. « Artificial intelligence in precision medicine in hepatology ». In : *J Gastroenterol Hepatol* 36.3 (mars 2021), p. 569-580.
- [243] A. SUBRAMANIAN et al. « Gene set enrichment analysis : a knowledge-based approach for interpreting genome-wide expression profiles ». In : *Proc Natl Acad Sci U S A* 102.43 (oct. 2005), p. 15545-15550.
- [244] Ryan SUDERMAN et Eric J DEEDS. « Machines vs. ensembles : effective MAPK signaling through heterogeneous sets of protein complexes ». In : *PLoS computational biology* 9.10 (2013), e1003278.
- [245] Irina SUROVTSOVA et al. « Approaches to complexity reduction in a systems biology research environment (SYCAMORE) ». In : *Proceedings of the 2006 Winter Simulation Conference*. IEEE. 2006, p. 1683-1689.
- [246] Tamás SZÉKELY JR et Kevin BURRAGE. « Stochastic simulation in systems biology ». In : *Computational and structural biotechnology journal* 12.20-21 (2014), p. 14-25.
- [247] Pavel TAIMR et al. « Activated stellate cells express the TRAIL receptor-2/death receptor-5 and undergo TRAIL-mediated apoptosis ». In : *Hepatology* 37.1 (2003), p. 87-95.

-
- [248] Jose-Juan TAPIA et James R FAEDER. « The Atomizer : extracting implicit molecular structure from reaction network models ». In : *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*. 2013, p. 726-727.
- [249] Mike K TERKELSEN et al. « Transcriptional dynamics of hepatic sinusoid-associated cells after liver injury ». In : *Hepatology* 72.6 (2020), p. 2119-2133.
- [250] Vo Hong THANH, Corrado PRIAMI et Roberto ZUNINO. « Efficient rejection-based simulation of biochemical reactions with stochastic noise and delays ». In : *The Journal of chemical physics* 141.13 (2014).
- [251] Nathalie THERET et al. « Integrative models for TGF- β signaling and extracellular matrix ». In : *Extracellular Matrix Omics* (2020), p. 209-225.
- [252] Lorenzo TORTOLINA et al. « Advances in dynamic modeling of colorectal cancer signaling-network regions, a path toward targeted therapies ». In : *Oncotarget* 6.7 (2015), p. 5041.
- [253] Cole TRAPNELL et al. « The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells ». In : *Nature biotechnology* 32.4 (2014), p. 381-386.
- [254] George L TRIGG. *Mathematical tools for physicists*. Wiley Online Library, 2005.
- [255] P. TRIVEDI, S. WANG et S. L. FRIEDMAN. « The Power of Plasticity-Metabolic Regulation of Hepatic Stellate Cells ». In : *Cell Metab* 33.2 (fév. 2021), p. 242-257.
- [256] Juliane S TROEGER et al. « Deactivation of hepatic stellate cells during liver fibrosis resolution in mice ». In : *Gastroenterology* 143.4 (2012), p. 1073-1083.
- [257] Takuma TSUCHIDA et Scott L FRIEDMAN. « Mechanisms of hepatic stellate cell activation ». In : *Nature reviews Gastroenterology & hepatology* 14.7 (2017), p. 397-411.
- [258] Alan Mathison TURING. « The chemical basis of morphogenesis ». In : *Bulletin of mathematical biology* 52 (1990), p. 153-197.
- [259] Ondřej URBANOVSKÝ. « Comparative Analysis of Rule-Based Languages ». Thèse de doct. MASARYK UNIVERSITY, 2019.

-
- [260] Lakshmi VENKATRAMAN et al. « Plasmin triggers a switch-like decrease in thrombospondin-dependent activation of TGF- β 1 ». In : *Biophysical journal* 103.5 (2012), p. 1060-1068.
- [261] Jose VILAR, Ronald JANSEN et Chris SANDER. « Signal Processing in the TGF- β Superfamily Ligand-Receptor Network ». In : *PLoS computational biology* 2 (fév. 2006), e3. DOI : 10.1371/journal.pcbi.0020003.
- [262] Jose M G VILAR, Ronald JANSEN et Chris SANDER. « Signal processing in the TGF- β superfamily ligand-receptor network ». In : *PLoS computational biology* 2.1 (2006), e3.
- [263] Jose MG VILAR et Leonor SAIZ. « Trafficking coordinate description of intracellular transport control of signaling networks ». In : *Biophysical journal* 101.10 (2011), p. 2315-2323.
- [264] Pedro VIZÁN et al. « Controlling long-term signaling : receptor dynamics determine attenuation and refractory behavior of the TGF- β pathway ». In : *Science signaling* 6.305 (2013), ra106-ra106.
- [265] William WAITES. « Modelling and measurement in synthetic biology ». Thèse de doct. University of Edinburgh, 2020.
- [266] William WAITES et al. « Rule-based epidemic models ». In : *Journal of theoretical biology* 530 (2021), p. 110851.
- [267] K. WAKE. « Hepatic stellate cells : Three-dimensional structure, localization, heterogeneity and development ». In : *Proc Jpn Acad Ser B Phys Biol Sci* 82.4 (mai 2006), p. 155-164.
- [268] LM WAKEFIELD et al. « Recombinant latent TGF- β 1 has a longer plasma half life in rats than active TGF- β 1, and a different tissue distribution ». In : *J. Clin. Invest* 86 (1990), p. 1976-1984.
- [269] Joseph WALPOLE, Jason A PAPIN et Shayn M PEIRCE. « Multiscale computational models of complex biological systems ». In : *Annual review of biomedical engineering* 15 (2013), p. 137-154.
- [270] Junjie WANG et al. « The self-limiting dynamics of TGF- β signaling in silico and in vitro, with negative feedback through PPM1A upregulation ». In : *PLoS computational biology* 10.6 (2014), e1003573.

-
- [271] Yu WANG et Jingfeng JIANG. « A two-dimensional (2d) systems biology-based discrete liver tissue model : A simulation study with implications for ultrasound elastography of liver fibrosis ». In : *Computers in biology and medicine* 104 (2019), p. 227-234.
- [272] Hayley C WARSINSKE et al. « Identifying mechanisms of homeostatic signaling in fibroblast differentiation ». In : *Bulletin of mathematical biology* 77.8 (2015), p. 1556-1582.
- [273] Katja WEGNER et al. « Dynamics and feedback loops in the transforming growth factor β signaling pathway ». In : *Biophysical chemistry* 162 (2012), p. 22-34.
- [274] G. WEI et al. « Comparison of murine steatohepatitis models identifies a dietary intervention with robust fibrosis, ductular reaction, and rapid progression to cirrhosis and cancer ». In : *Am J Physiol Gastrointest Liver Physiol* 318.1 (jan. 2020), G174-G188.
- [275] Rebecca WELLS. « The Role of Matrix Stiffness in Hepatic Stellate Cell Activation and Liver Fibrosis ». In : *Journal of clinical gastroenterology* 39 (mai 2005), S158-61. DOI : 10.1097/01.mcg.0000155516.02468.0f.
- [276] Douglas Brent WEST et al. *Introduction to graph theory*. T. 2. Prentice hall Upper Saddle River, 2001.
- [277] Darren J WILKINSON. « Stochastic modelling for quantitative description of heterogeneous biological systems ». In : *Nature Reviews Genetics* 10.2 (2009), p. 122-133.
- [278] J. WILSON-KANAMORI et al. « Kappa rule-based modeling in synthetic biology ». In : *Methods Mol Biol* 1244 (2015), p. 105-135.
- [279] Robert F WOLFF et al. « PROBAST : a tool to assess the risk of bias and applicability of prediction model studies ». In : *Annals of internal medicine* 170.1 (2019), p. 51-58.
- [280] Inc. WOLFRAM RESEARCH. *Mathematica*. Wolfram Research, Inc., 2017.
- [281] S. WU et al. « An update on animal models of liver fibrosis ». In : *Front Med (Lausanne)* 10 (2023), p. 1160053.
- [282] Emilia M WYSOCKA et al. « Comparison of rule-and ordinary differential equation-based dynamic model of DARPP-32 signalling network ». In : *PeerJ* 10 (2022), e14516.

-
- [283] WU YANG et al. « Single-cell transcriptomic analysis reveals a hepatic stellate cell-activation roadmap and myofibroblast origin during liver fibrosis in mice ». In : *Hepatology* 74.5 (2021), p. 2774-2790.
- [284] S. C. YANGUAS et al. « Experimental models of liver fibrosis ». In : *Arch Toxicol* 90.5 (mai 2016), p. 1025-1048.
- [285] Misa YOSHIZAWA et al. « Computational simulation of liver fibrosis dynamics ». In : *Scientific reports* 12.1 (2022), p. 14112.
- [286] Angela C ZEIGLER et al. « Computational modeling of cardiac fibroblasts and fibrosis ». In : *Journal of molecular and cellular cardiology* 93 (2016), p. 73-83.
- [287] Jingyu ZHANG et al. « Pathway crosstalk enables cells to interpret TGF- β duration ». In : *NPJ systems biology and applications* 4.1 (2018), p. 18.
- [288] Wenjun ZHANG et al. « Heterogeneity of hepatic stellate cells in fibrogenesis of the liver : insights from single-cell transcriptomic analysis in liver injury ». In : *Cells* 10.8 (2021), p. 2129.
- [289] Z. ZI. « Sensitivity analysis approaches applied to systems biology models ». In : *IET Syst Biol* 5.6 (nov. 2011), p. 336-336.
- [290] Z. ZI et E. KLIPP. « Constraint-based modeling and kinetic analysis of the Smad dependent TGF-beta signaling pathway ». In : *PLoS One* 2.9 (sept. 2007), e936.
- [291] Zhike ZI. « Molecular engineering of the TGF- β signaling pathway ». In : *Journal of molecular biology* 431.15 (2019), p. 2644-2654.
- [292] Zhike ZI, Douglas A CHAPNICK et Xuedong LIU. « Dynamics of TGF- β /Smad signaling ». In : *FEBS letters* 586.14 (2012), p. 1921-1928.
- [293] Zhike ZI et al. « Quantitative analysis of transient and sustained transforming growth factor- β signaling dynamics ». In : *Molecular systems biology* 7.1 (2011), p. 492.



Titre : Modélisation en Kappa de la dynamique des cellules étoilées hépatiques durant le développement et la réversion de la fibrose

Mot clés : Fibrose hépatique, CEF, TGFB1, KAPPA, Modélisation

Résumé : La fibrose hépatique est une réponse de cicatrisation excessive induite par des lésions chroniques. Elle se caractérise par une accumulation de matrice extracellulaire (MEC), principalement constituées de collagène 1 (COL1), qui augmente la rigidité des tissus et entraîne un grave dysfonctionnement du foie. L'activation des cellules étoilées du foie (CEF), induite par le facteur de croissance TGFB1, est le principal processus à l'origine de la fibrose hépatique. Afin d'étudier la dynamique des CEF durant le développement et la réversion de la fibrose, nous avons développé un modèle multi-échelle intégrant les différents états des CEF ainsi que leur production

de COL1, sous influence du TGFB1. Ce modèle est implémenté avec le langage Kappa, qui est un langage de réécriture de graphes à sites. En plus d'être le premier modèle multi-échelle de Kappa, ce modèle nous a permis de capturer la plasticité des cellules étoilées durant le développement et la réversion la fibrose. Les prédictions du modèle montrent que l'état d'inactivation des CEF joue un rôle essentiel dans le développement de la fibrose. Le modèle a été validé par de nouvelles expériences chez la souris et les prédictions ont été validées avec des données de RNAseq chez des patients fibrotiques.

Title: Kappa modelling of hepatic stellate cell dynamics during fibrosis development and reversion

Keywords: Liver Fibrosis, HSC, TGFB1, KAPPA, Modelling

Abstract: Hepatic fibrosis is an excessive scarring response induced by chronic injury. It is characterised by an accumulation of extracellular matrix (ECM), mainly composed of collagen 1 (COL1), which increases tissue rigidity and leads to severe liver dysfunction. Activation of hepatic stellate cells (HSCs), induced by growth factor TGFB1, is the main process underlying liver fibrosis. In order to study the dynamics of HSC during the development and reversion of fibrosis, we have developed a multi-scale model integrating the different states of HSC as well as their produc-

tion of COL1, under the influence of TGFB1. This model is implemented using the Kappa language, which is a rewriting language for site graphs. As well as being the first multi-scale Kappa model, this model allowed us to capture the plasticity of star cells during the development and reversion of fibrosis. The model's predictions show that the inactivation state of HSC plays an essential role in the development of fibrosis. The model was validated by new experiments in mice and the predictions were validated with RNAseq data from fibrotic patients.