



**HAL**  
open science

# Intégrer les échelles moléculaires et cellulaires dans l'inférence de réseaux métaboliques. Application aux xénobiotiques.

Victorien Delannée

## ► To cite this version:

Victorien Delannée. Intégrer les échelles moléculaires et cellulaires dans l'inférence de réseaux métaboliques. Application aux xénobiotiques.. Bio-Informatique, Biologie Systémique [q-bio.QM]. Univ. Rennes 1, 2017. Français. NNT: . tel-01659375v1

**HAL Id: tel-01659375**

**<https://univ-rennes.hal.science/tel-01659375v1>**

Submitted on 8 Dec 2017 (v1), last revised 22 Jan 2018 (v2)

**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 / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Bretagne Loire*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Informatique*

**École doctorale Matisse**

présentée par

**Victorien DELANNÉE**

préparée à l'unité de recherche IRISA – UMR6074 et IRSET – UMR1085  
Institut de Recherche en Informatique et Système Aléatoires et Institut  
de Recherche en Santé, Environnement et Travail  
ISTIC

**Intégrer les  
échelles molé-  
culaires et cellu-  
laires dans l'infé-  
rence de réseaux  
métaboliques.  
Application aux  
xénobiotiques**

**Thèse soutenue à Rennes  
le 8 Novembre 2017**

devant le jury composé de :

**Gilles BERNOT**

Professeur à l'Université de Nice Sophia  
Antipolis / *Rapporteur*

**Fabien JOURDAN**

Directeur de Recherche INRA / *Rapporteur*

**Karine AUDOUZE**

Maitre de conférences à l'Université Paris  
Diderot / *Examinatrice*

**Jérémie BOURDON**

Professeur à l'Université de Nantes /  
*Examineur*

**Olivier DAMERON**

Maitre de conférences à l'Université Rennes 1  
/ *Examineur*

**Anne SIEGEL**

Directrice de Recherche CNRS /  
*Directrice de thèse*

**Nathalie THÉRET**

Directrice de Recherche INSERM /  
*Co-directrice de thèse*



L'Homme se découvre quand il se mesure à l'obstacle  
par Antoine de Saint-Exupéry





# Remerciements

Il est parfois coutume d'entamer cette section par des remerciements formels, mais je commencerai par une touche philosophique en remerciant toutes les personnes qui ont croisé mon chemin, que ce soit l'instant d'un moment ou d'une période plus longue. En effet, la vie est faite de rencontres, d'opportunités, de moments qui peuvent paraître anodins, mais qui peuvent marquer à jamais. Tous ces instants passés, éphémères ou plus durables, malheureux ou heureux, désagréables ou agréables font de chaque jour une journée unique qui permet d'en apprendre toujours plus, contribuer aux succès futurs et à faire ce que nous sommes. Et puis, comme disait W. Churchill, le succès c'est d'aller d'échec en échec sans perdre son enthousiasme. C'est aussi en regardant des mauvais films qu'on apprécie les bons (c'était ici l'occasion de placer cette très célèbre citation qui n'est pas de W. Churchill). J'adresse ici donc un grand merci à ceux qui ont croisé et qui croiseront mon chemin.

Je remercie en premier lieu Gilles Bernot et Fabien Jourdan de m'avoir fait l'honneur d'accepter le rôle de rapporteur. Puis, je tiens aussi à remercier les autres membres du jury : Karine Audouze, Jérémie Bourdon et Olivier Dameron.

Je tiens aussi à remercier Nathalie Théret à la fois de m'avoir recruté en stage de M2, qui a abouti à cette thèse et pour ton rôle de supervision. Je remercie aussi Anne Siegel et Sophie Langouët qui ont aussi contribué à l'encadrement de cette thèse. Merci à vous trois.

Mes remerciements se portent aussi à l'égard de toutes les personnes ayant divagué vers de nouveaux horizons, ou encore présentes dans le groupe symbiose constitué des équipes Dyliss, Genscale et la plateforme Genouest. Merci aussi à François Moreews (anciennement dans l'équipe Genscale) de m'avoir encadré en stage de M1.

Je remercie tout autant tous les membres de l'équipe 5 de l'IRSET. Il serait trop long de citer tout le monde, alors merci aux personnes notamment en photo sur ces sites web : <http://www.irisa.fr/dyliss/members>, <https://team.inria.fr/genscale/team-members/>, <https://www.genouest.org/people/>, [https://www.irset.org/?page\\_id=1542&lang=fr](https://www.irset.org/?page_id=1542&lang=fr) (en espérant que les URL ne changent pas :-\).

Une pensée va aussi à l'équipe d'enseignement de l'UFR pharmacie en mathématiques et physique et plus particulièrement à Eric, Giulio, Fanny et Olivier avec qui j'ai eu la chance d'enseigner et avec qui il fut très agréable de travailler.

La thèse a aussi été l'occasion de participer à de nombreuses activités associatives fort enrichissantes à la fois humainement et professionnellement. Ainsi, je remercie toutes les personnes avec qui j'ai pu travailler au sein des activités de Nicomaque notamment sur les projets forum Docteurs & Entreprises et Sciences en cour[t]s. Ce fut un réel plaisir que d'avoir pu échanger avec vous. Je tiens à remercier en particulier Aymeric Antoine-Lorquin et Jean Coquet pour la réalisation de notre super film. Je remercie aussi en complément Fanny et Seb.

Comme promis, je dédie ce paragraphe à ma co-bureau Stéphanie qui s'est prêtée à la recherche de fautes pour une bonne partie de ce manuscrit.

Merci aux amis locaux Thomas et Jean, anciens du Master BIG, ainsi qu'à Panos, Mathilde et Julie. Je remercie aussi mes grands copains d'enfance Nicolas, Gwen, Ben et Jimmy pour les moments passés ensemble et qui continuent toujours plus de 15 années plus tard. Les temps deviennent maintenant un peu difficiles en raison des projets différents de chacun qui nous occupent tous énormément, mais il reste encore plein de délires à vivre.

Enfin, je conclus cette partie en remerciant mes parents Isabelle (merci maman pour ta participation dans l'élimination des fautes) et Patrice, qui sans leur soutien indéfectible sur tous les plans possibles aussi bien moral que financier, m'ont permis aujourd'hui d'en arriver à ce parcours et m'offrir la chance de m'épanouir dans des secteurs professionnels qui me plaisent.





# Table des matières

Remerciements	1
Table des matières	3
Introduction	7
1 De la modélisation gros grain à la modélisation fine du métabolisme	11
1.1 La biologie des systèmes	12
1.2 Les approches modèles pour l'étude du métabolisme	15
1.2.1 Généralités	15
1.2.2 Les voies métaboliques : Nomenclatures et identifiants des composés chimiques	15
1.2.2.1 Les métabolites	15
1.2.2.2 Encodage d'un métabolite	16
1.2.3 Les réactions biochimiques	20
1.2.3.1 Les enzymes	20
1.2.3.2 Encodage d'une réaction	22
1.3 Modélisation du métabolisme	23
1.3.1 Les approches à l'échelle de l'organe, du tissu et de l'organisme	24
1.3.2 Les approches à l'échelle cellulaire	27
1.3.2.1 Les méthodes basées sur les graphes	27
1.3.2.2 Les méthodes basées sur les contraintes	28
1.3.2.3 Les méthodes cinétiques	29
1.3.2.4 Résumé	31
1.3.3 Les approches à l'échelle moléculaire	31
1.3.3.1 Prédiction des sites de métabolisme	31
1.3.3.2 Prédiction de réactions biochimiques	33
1.3.3.3 Résumé	35
1.3.4 Cas particulier des processus de toxicité	35

2	Les xénobiotiques : métabolisme et cancer	39
2.1	Généralités	39
2.2	Métabolisme des xénobiotiques	40
2.3	Les xénobiotiques de type AHA	42
2.3.1	Le métabolisme général des AHA	44
2.3.2	Le métabolisme des AHA par les enzymes de Phase I	46
2.3.3	Le métabolisme des AHA par les enzymes de Phase II	51
2.4	Enjeux de modélisation : un challenge informatique et biologique	53
2.5	Résumé et principaux apports	54
3	Construction automatique <i>de novo</i> de graphes métaboliques combinant fortes sensibilités et précisions	57
3.1	Introduction	58
3.2	Description des méthodes implémentées au sein d'un pipeline	60
3.2.1	Prédiction des réactions métaboliques	60
3.2.2	Identification de la sélectivité enzymatique et prédiction des SOM	62
3.2.3	Prédiction de la réactivité vis-à-vis de l'ADN	63
3.2.4	Implémentation de filtres pour limiter la combinatoire	63
3.3	Procédure de calibration	64
3.3.1	Base de données biologiques pour l'optimisation	64
3.3.2	Optimisation des méthodes de prédictions des SOM et des SOR	64
3.4	Validation et analyses statistiques	67
3.5	Discussion et conclusion	72
3.6	Résumé	77
4	Cas d'étude : Prédiction des métabolismes des AHA	79
4.1	Prédictions des graphes métaboliques d'une famille de xénobiotique : les AHA	80
4.1.1	Matériel biologique d'étude	80
4.1.2	Prédictions des graphes métaboliques d'une famille de xénobiotique : les AHA	80
4.2	Implémentation logicielle	84
4.3	Conclusion	91
4.4	Résumé	91
5	Développement d'un modèle dynamique pour étudier le métabolisme du MeIQx	93
5.1	Introduction	94
5.2	Un modèle numérique des compétitions enzymatiques du métabolisme	95
5.2.1	Connaissances biologiques	95
5.2.2	Données expérimentales	95
5.2.3	Modélisation des compétitions enzymatiques	99
5.3	Evaluation des dynamiques enzymatiques	101
5.3.1	Ajustement de modèles, estimation des paramètres et filtration <i>a posteriori</i>	102

5.3.2	La régiosélectivité est dépendante de la concentration de MeIQx	104
5.3.3	73 ensembles de paramètres expliquent les données . . . . .	106
5.3.4	Validation du modèle . . . . .	107
5.4	Discussion et conclusion . . . . .	107
5.5	Résumé . . . . .	113
6	Cas d'étude : Application au métabolisme du MeIQx	115
6.1	Introduction . . . . .	115
6.2	Influence de la biphasicité sur la réponse biologique . . . . .	116
6.2.1	Effet de biphasicité sur l'équilibre des voies métaboliques . . . . .	116
6.2.2	La régulation des voies métaboliques opère à faible concentration de MeIQx . . . . .	117
6.2.3	Preuve de l'effet biphasique et évaluation de la génotoxicité . . . . .	120
6.2.4	Influence de la régulation du CYP1A2 sur l'équilibre des voies métaboliques . . . . .	120
6.3	Discussion et conclusion . . . . .	122
6.4	Résumé . . . . .	126
	Conclusion et perspectives	127
	Bibliographie	133
	Table des figures	159
	Publications	163
	Annexes	165



# Introduction

La biologie des systèmes est une science intégrative à l'interface de nombreuses disciplines telles que l'informatique, les mathématiques, la physique, etc. L'objectif de la biologie des systèmes est d'appréhender la complexité cellulaire par l'intégration des connaissances biologiques à travers la modélisation de réseaux. La modélisation est parfois fortement limitée par l'absence ou la quantité limitée de connaissances et données biologiques. Il est donc impératif d'implémenter des méthodes prédictives en biologie des systèmes, permettant à la fois d'intégrer la connaissance limitée, mais aussi générer du nouveau savoir même en l'absence de données.

Le métabolisme d'une cellule décrit l'ensemble des transformations de composés biologiques via des modifications moléculaires contrôlées par des enzymes. Classiquement, les approches de modélisation de l'état de l'art se focalisent sur une échelle spécifique qui peut être au niveau des organes (modèle numérique sous la forme d'équations différentielles ordinaires), des cellules (approches de programmation par contraintes sur des graphes à grande échelle) ou des molécules (approches d'apprentissage s'appuyant sur des bases de connaissances). Néanmoins, l'article "DREAMS of metabolism" de Soh et Hatzimanikatis [204] exposait l'importance et les enjeux du développement d'approches *in silico* multi-échelles, afin de prédire et d'étudier le métabolisme d'une cellule dans sa globalité par la combinaison de différentes démarches de modélisation. Le travail de cette thèse s'inscrit dans l'optique d'employer une approche multi-échelles combinant des méthodes informatiques permettant de prendre en compte simultanément les échelles cellulaires et moléculaires du métabolisme. L'objectif d'une telle approche est une meilleure compréhension de métabolisme général et des mécanismes complexes qui le régissent.

Ainsi, l'enjeu de cette thèse est double en développant et implémentant une méthode permettant la construction automatique *de novo* de graphes de biotransformations de substances chimiques chez l'homme à partir de sa structure 2D uniquement ; et une autre approche exploitant un modèle sous-déterminé intégrant des données biologiques limitées pour l'étude de la dynamique de métabolisme de toxiques à travers la modélisation de mécanismes biologiques complexes.

Le premier objectif est la construction automatique *de novo* de graphes de biotransformations de substances chimiques chez l'homme. L'enjeu est donc de prédire correctement le métabolisme complet et la toxicité d'une molécule dont on connaît uniquement la structure 2D. Outre la prédiction des graphes métaboliques, une difficulté majeure des approches implémentées réside dans l'explosion de la combinatoire inhérente aux méthodes de prédictions actuelles du métabolisme de composés chimiques. Pour atteindre ce but, le premier travail de cette thèse a consisté à s'appuyer sur des informations relatives aux caractéristiques moléculaires pour prédire les réactions métaboliques que peut subir une molécule en limitant significativement la combinatoire des transformations possibles.

Lors de la modélisation dynamique d'un système, on peut parfois être confronté à un problème de taille majeure : la limitation des données biologiques. Pour remédier à ce problème, la seconde partie du travail s'est penchée à développer une méthodologie permettant la prédiction de la dynamique tout en exploitant des données biologiques restreintes. Cette dernière a été effectuée par l'identification de modèles sous-déterminés en considérant une famille d'ensembles de paramètres décrivant la dynamique d'un système. De plus, cette approche a aussi permis d'élucider un mécanisme complexe de compétitions enzymatiques à travers le développement d'un modèle simple modélisant ces effets de compétitions.

Cette thèse constitue donc un travail préliminaire pour la modélisation de la complexité biologique à travers l'implémentation d'un ensemble de techniques multi-échelles qui trouvent tout leur sens dans la problématique de l'étude de molécules toxiques. A travers six chapitres, elle dévoilera aussi l'intérêt et la nécessité de continuer à développer des méthodes informatiques multi-échelles et non focalisées sur un aspect du métabolisme afin de le comprendre dans son intégralité et ne pas passer à côté de certains mécanismes, métabolites ou voies métaboliques majeurs qui pourraient ne jamais être mis en lumière par l'utilisation d'approches avec *à priori*. Aussi, les différentes approches implémentées dans cette thèse seront aussi rapprochées afin d'étayer ces propos et montrer leur complémentarité par l'étude du contaminant alimentaire MeIQx.

En effet, l'étude approfondie du MeIQx servira de fil directeur tout au long de ce document afin d'illustrer l'intérêt et la complémentarité des approches développées dans le cadre de cette thèse. Il sera par exemple introduit dès l'état de l'art lors de la présentation de la formulation informatique des molécules biochimiques (voir chapitre 1 section 1.2.2). Puis, son métabolisme sera décrit dans le chapitre 2. Il sera aussi un des candidats adoptés pour la validation de la construction automatique *de novo* de graphes métaboliques permettant notamment de prédire son métabolisme et sa génotoxicité (chapitres 3 et 4). Enfin, il sera utilisé une dernière fois pour appréhender la complexité métabolique de certains xénobiotiques à travers des mécanismes de régulation particuliers médiant la toxicité (chapitres 5 et 6).

Le premier chapitre de cette thèse vise à présenter la biologie des systèmes et ses approches majeures pour modéliser un système biologique. Puis, ce chapitre est suivi d'une présentation du contexte biologique dans lequel sont présentés le modèle d'application (les xénobiotiques) et ses subtilités. Ce dernier est utilisé pour la validation des méthodes informatiques développées et la génération de nouvelles connaissances biologiques. Le chapitre 3 porte sur le développement d'une méthode originale de construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions, tout en annotant ces graphes avec des informations moléculaires (plus précisément, leur génotoxicité). Cette méthode est ensuite appliquée dans le chapitre 4 afin de prédire le métabolisme et évaluer la génotoxicité d'une même famille de contaminants environnementaux et alimentaires dont 90% sont inconnus. Le chapitre 5 expose un modèle générique pour la modélisation dynamique de mécanismes complexes de compétitions enzymatiques à travers le développement d'une méthodologie basée sur l'exploitation d'un modèle sous-déterminé permettant l'utilisation de données biologiques limitées. Cette approche est appliquée dans le chapitre 6 au MeIQx, qui est un contaminant alimentaire majeur afin d'appréhender l'influence de mécanismes complexes de compétitions enzymatiques sur la génotoxicité. Enfin, la conclusion et les perspectives de cette thèse terminent ce document.





# Chapitre 1

## De la modélisation gros grain à la modélisation fine du métabolisme

Ce premier chapitre a pour vocation d'introduire les contextes informatiques de cette thèse. Il illustrera les principales approches de modélisation d'un système biologique employées par la communauté scientifique s'intéressant notamment au métabolisme des xénobiotiques. Cette présentation sera décomposée en fonction des différentes échelles d'étude du métabolisme. Un accent sur l'application de ces méthodes sera mis en évidence tout au long de ce chapitre.

Le plan de ce premier chapitre s'articule autour d'une courte introduction à la biologie des systèmes, suivi d'une présentation générale du métabolisme. Puis, les différentes méthodes de modélisation du métabolisme à travers trois niveaux d'étude seront exposées. Enfin, cette partie sera complétée par un bref descriptif de la prédiction *in silico* de la toxicité de systèmes biologiques.

---

<b>1.1</b>	<b>La biologie des systèmes</b>	<b>12</b>
<b>1.2</b>	<b>Les approches modèles pour l'étude du métabolisme</b>	<b>15</b>
1.2.1	Généralités	15
1.2.2	Les voies métaboliques : Nomenclatures et identifiants des composés chimiques	15
1.2.3	Les réactions biochimiques	20
<b>1.3</b>	<b>Modélisation du métabolisme</b>	<b>23</b>
1.3.1	Les approches à l'échelle de l'organe, du tissu et de l'organisme	24
1.3.2	Les approches à l'échelle cellulaire	27
1.3.3	Les approches à l'échelle moléculaire	31
1.3.4	Cas particulier des processus de toxicité	35

---

## 1.1 La biologie des systèmes

La biologie des systèmes, qui a pour objectif de modéliser des systèmes biologiques complexes, peut être définie comme une science intégrative multi-disciplinaire à l'interface de nombreux domaines tels que l'informatique, les mathématiques, la physique, etc. Bien que l'essor de cette science soit souvent associé au développement des ressources informatiques, ses prémices pourraient remonter à l'Antiquité lorsqu'Aristote indiquait la nécessité de voir l'ensemble avant la partie ("the whole is of necessity prior to the part") [56]. Cette pensée a émergé entre les XIV<sup>ème</sup> et le XIX<sup>ème</sup> siècles avec notamment N. of Cusa, G.W. Leibniz, et J.W. von Goethe, qui sont considérés par certains comme les pères de la philosophie de la biologie des systèmes [56]. Néanmoins, la biologie des systèmes moderne trouve ses fondements avec l'émergence de la théorie générale des systèmes qui modélise la croissance individuelle. Cette théorie proposée par L. von Bertalanffy et publiée en 1968 [239] a fortement impacté sur l'essor de la biologie des systèmes. Cette dernière trouve toujours son application dans des modèles démographiques ou parfois pour modéliser la biomasse.

Le développement de la biotechnologie et de nouvelles techniques analytiques a permis une accumulation importante de connaissances dans les domaines allant de la génomique à la métabolomique. Ce nouveau savoir généré a donc nécessité l'invention de nouvelles techniques analytiques pour comprendre la complexité cellulaire. En effet, comme exposée par Z.N. Oltvai, la complexité cellulaire est inter-connectée et dépendante à la fois du génome, du transcriptome, du protéome et du métabolome [166]. Ainsi il convient de considérer les fonctions cellulaires comme réparties au sein de différents composés hétérogènes qui interagissent entre eux au moyen de réseaux inter-connectés (réseaux métaboliques, réseaux d'interactions protéines-protéines et réseaux de régulation génique) (figure 1.1) [166]. C'est à ce niveau que la biologie des systèmes trouve tout son sens pour étudier les différents réseaux et leurs interactions afin d'explorer et appréhender la complexité cellulaire.

La biologie des systèmes est donc une science variée intervenant à différents niveaux organisationnels afin de comprendre la complexité biologique. Selon Kitano, la compréhension de cette complexité à l'échelle d'un système biologique repose sur l'étude de quatre propriétés majeures [116] :

- **L'analyse de la structure d'un système** visant à étudier les réseaux métaboliques, les réseaux d'interactions, ou les réseaux de régulation et d'en comprendre leurs propriétés ;
- **L'analyse de la dynamique d'un système** afin de déchiffrer son comportement en fonction du temps dans différentes conditions expérimentales ;
- **Le contrôle du système** qui vise à identifier les acteurs clés du système en le perturbant et à permettre ainsi le développement de nouvelles cibles thérapeutiques ;
- **Concevoir un nouveau système biologique** répondant à des fonctions biologiques contrôlées.

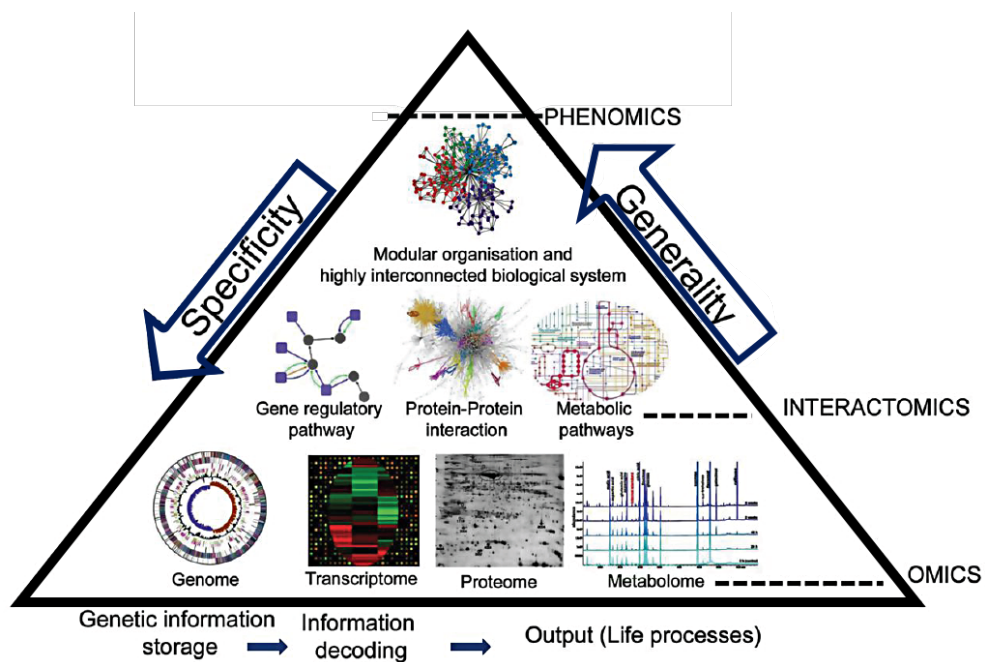


FIGURE 1.1 – **Pyramide de la complexité de la vie autour de laquelle les problématiques en biologie des systèmes s’articulent (adaptée de [67])**

Les fonctions biologiques sont dépendantes de différents matériels génétiques tels que l’ADN, les différents ARN, les protéines et les métabolites. Ces différentes entités sont souvent organisées au sein de catégories distinctes qui sont la génomique, la transcriptomique, la protéomique et la métabolomique. Ces entités biologiques ne sont pas dépendantes, mais interconnectées entre elles formant ainsi des réseaux : réseaux métaboliques, réseaux d’interactions protéine-protéine et réseaux de régulation. De plus, ces différents réseaux ne sont pas indépendants et peuvent "dialoguer" entre eux afin d’affecter un processus biologique qui pourra impacter sur le phénotype.

Ergo, de nombreuses méthodes pour l'intégration et l'analyse de données biologiques ont été implémentées afin d'évaluer la toxicité d'un composé biologique, d'identifier un biomarqueur d'une maladie, d'appréhender des mécanismes complexes à l'origine de maladies graves comme le cancer, d'identifier de nouvelles cibles thérapeutiques, ... Pour mener à bien ces différentes études, les approches développées peuvent se focaliser sur une échelle et cibler une réaction majeure, sur l'étude d'un réseau métabolique, ou d'un réseau d'intégration, ou soit être multi-échelles en combinant plusieurs approches, ou plusieurs réseaux, ou encore dans le cadre de l'homme en étudiant par exemple plusieurs organes. Plus encore, les approches intégratives conçues sont parfois multi-échelles et utilisées dans le cadre d'une population afin de déchiffrer la complexité inter-individuelle pour par exemple évaluer l'effet d'une nouvelle molécule thérapeutique (figure 1.2).

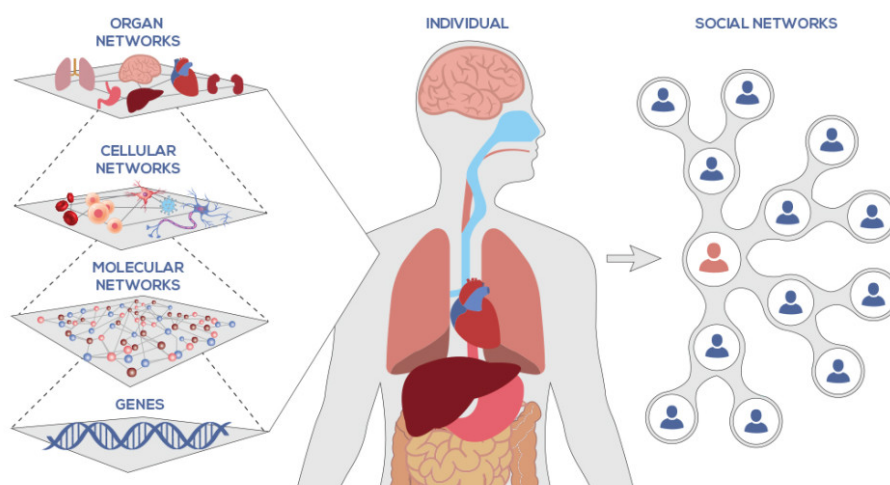


FIGURE 1.2 – Les différents niveaux organisationnels d'étude de la biologie des systèmes [1]

La biologie des systèmes s'attache à étudier une partie, ou l'intégralité de réseaux. Les niveaux d'études peuvent être focalisés sur un type de réseaux, ou au contraire être inter-connectés afin d'étudier des sur-structures comme un réseau d'organes. De plus, ces études peuvent aussi se faire à l'échelle d'un individu ou d'une population.

Malgré l'intérêt de développer des méthodes combinant plusieurs niveaux biologiques organisationnels, il est parfois avant tout nécessaire d'appréhender spécifiquement une fonction à un niveau organisationnel précis comme le métabolisme avant d'extrapoler à très grande échelle. Ainsi, cette thèse se focalisera sur le développement de méthodes pour la prédiction et l'analyse du métabolisme de xénobiotiques chez l'homme.

## **1.2 Les approches modèles pour l'étude du métabolisme**

Le métabolisme est une fonction majeure de tout organisme vivant régissant son fonctionnement au cours de toutes les grandes étapes de la vie et assurant ainsi son développement, sa reproduction et son adaptation à l'environnement. Ce processus désigne l'ensemble des réactions biochimiques extra et intracellulaires qui transforment des composés chimiques en molécules appelées métabolites. Ce chapitre présentera succinctement les bases du métabolisme nécessaires à la compréhension de cette thèse.

### **1.2.1 Généralités**

Le métabolisme est un processus qui permet d'assurer deux fonctions indispensables à la survie d'un organisme vivant : la production d'énergie et la synthèse de molécules. Il est donc habituellement divisé en deux classes qui sont le catabolisme et l'anabolisme. L'anabolisme désigne l'ensemble des réactions endergoniques (nécessitant un apport d'énergie) consommant l'énergie issue du catabolisme ou de l'environnement et assurant la synthèse de molécules complexes telles que les protéines, les glucides, les lipides et les acides nucléiques. Cette fonction est aussi connue sous le terme de biosynthèse. D'autre part, le catabolisme désigne l'ensemble des réactions spontanées produisant de l'énergie par la dégradation de molécules chimiques complexes, provenant par exemple de l'alimentation ou des réserves cellulaires, en molécules chimiques plus simples. Ces deux fonctions sont indispensables au bon fonctionnement d'un organisme, mais aussi dans sa protection par la détoxification via le métabolisme de substances étrangères provenant de l'environnement communément connues sous le terme de xénobiotiques. Ces substances sont biotransformées par une succession ordonnée de réactions chimiques appelée voie métabolique.

### **1.2.2 Les voies métaboliques : Nomenclatures et identifiants des composés chimiques**

Une voie métabolique représente l'enchaînement de réactions biochimiques qui biotransforment une substance chimique initiale en un ensemble de métabolites intermédiaires. Ces réactions sont généralement catalysées par des enzymes.

#### **1.2.2.1 Les métabolites**

Les métabolites sont de petites molécules différenciées des macromolécules en raison de leur faible masse moléculaire et produites par la biotransformation d'une molécule initiale lors du métabolisme. Afin de caractériser leur structure chimique, plusieurs moyens permettent de les identifier. Une première façon de les identifier est l'utilisation du numéro CAS (Chemical Abstracts Service), qui est un numéro d'enregistrement d'une molécule chimique dans la base de données CAS de l'American Chemical Society [244]. Néanmoins, ce numéro est propriétaire et permet d'identifier uniquement les métabolites présents dans cette base de données. Cet identifiant

ne constitue donc pas une façon générale d'identifier la structure chimique. Le second moyen d'identifier une substance chimique est l'utilisation de la nomenclature de l'IUPAC (International Union of Pure and Applied Chemistry) [60]. Cette nomenclature est utilisée pour définir la structure 2D d'un composé organique en fonction de règles basées sur la structure carbonée principale et de ses groupes fonctionnels (par exemple une fonction alcool) qui la composent. Cette dernière est largement utilisée par les chimistes pour décrire une molécule.

### 1.2.2.2 Encodage d'un métabolite

Afin d'utiliser ces molécules chimiques au sein d'applications informatiques, plusieurs formats génériques ou propriétaires encodant la structure 2D de la molécule ont été créés. Parmi les plus connus et open source, il convient de citer les plus utilisés que sont : le SMILES (Simplified Molecular-Input Line-Entry System) [13, 242], l'InChI (International Chemical Identifier) [88] et les fichiers de format MDL [50].

Inventé dans les années 1980 par D. Weininger, le langage SMILES permet de définir simplement la structure 2D d'une molécule chimique par la combinaison de chaînes de caractères ASCII représentant le squelette et les greffons de la molécule. L'encodage au format SMILES est simple et réalisé en deux étapes (figure 1.3). La première étape consiste à supprimer les hydrogènes, ouvrir les cycles au niveau de liaison simple et numéroter ces ouvertures. Enfin, l'identifiant est construit en partant d'une extrémité et en y ajoutant ses branches par l'utilisation de parenthèses. Ce système permet donc d'identifier toutes molécules chimiques et la génération de leur structure 2D à partir de cet identifiant. Il présente cependant le désavantage de produire potentiellement différentes solutions. Dans l'exemple donné, on peut commencer par décrire la molécule en partant de l'atome de carbone jusqu'à l'atome d'azote comme cela a été fait, mais il est tout aussi exact de faire l'inverse. On aurait pu aussi tout autant couper les cycles à des endroits différents. Dans un but d'uniformisation des solutions générées, de nombreuses librairies ont été développées afin de générer un unique format SMILES nommé sous le terme SMILES canonique. De plus, ce format peut aussi décrire la stéréochimie (conformation spatiale) et l'isotopie (même élément chimique partageant un nombre de neutrons différents) de la molécule par l'implémentation des SMILES isomériques. Des librairies JAVA telles que la très réputée CDK (Chemistry Development Kit) implémentent ces fonctionnalités [246].

En 2005, un nouveau format permettant de décrire la structure 2D d'une molécule a vu le jour à l'initiative de l'IUPAC et du NIST (National Institute of Standards and Technology). L'objectif de ce format est d'incorporer plus d'informations qu'une "simple" chaîne SMILES. A la différence de l'identifiant SMILES, le format InChi se veut être un identifiant unique pour la molécule. Cet identifiant commence par "In-ChI=" suivi de la version utilisée, ainsi que son format (standard ou non standard) et d'une à six couches décrivant la molécule. La première couche est la couche principale

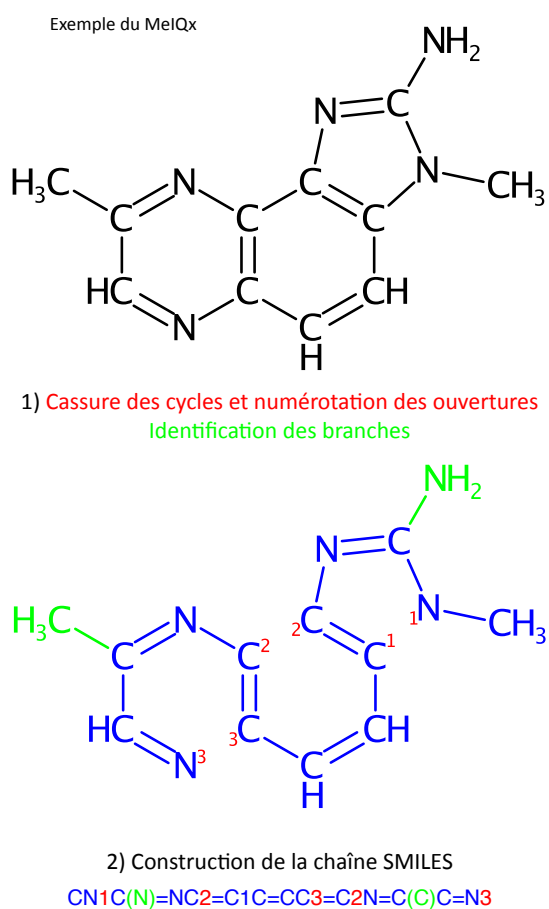


FIGURE 1.3 – Construction d'une chaîne SMILES pour décrire une structure 2D d'une molécule chimique

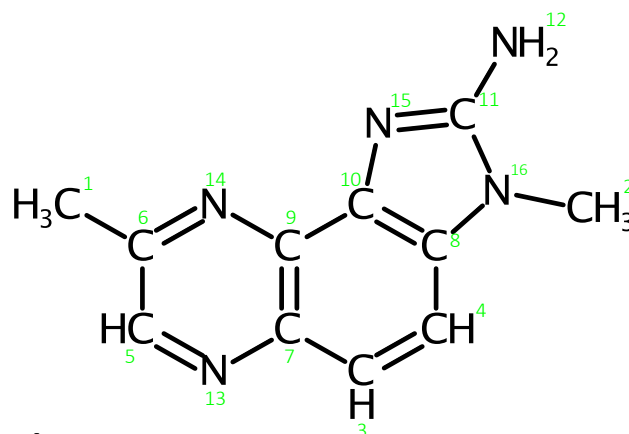
L'encodage au format SMILES est simple et réalisé en deux étapes. La première étape consiste à supprimer les hydrogènes, ouvrir les cycles et numérotter ces ouvertures. Enfin, l'identifiant est construit en partant d'une extrémité et en y ajoutant ses branches par l'utilisation de parenthèses.



contenant la formule chimique brute, la connectivité interatomique et une dernière décrivant le nombre d'atomes d'hydrogène connectés aux autres atomes. Les 5 autres couches contiennent notamment les informations sur les charges de la molécule, la stéréochimie et l'isotopie. En 2007, afin de faciliter la recherche de molécules chimiques, la création d'une clé InChI (InChIKey) à partir de l'identifiant unique InChI a vu le jour. Cette clé est générée par une fonction de hachage SHA256 et contient 27 caractères. Les 14 premiers contiennent les informations de la couche principale. Ceux-ci sont suivis par un trait d'union, puis de 10 caractères contenant les informations des autres couches. Les deux avant-derniers caractères identifient le format et la version. Enfin, un dernier caractère séparé du bloc précédent par un trait d'union indique l'état de protonation de la molécule.

Le dernier format couramment utilisé pour encoder une molécule chimique est le format de fichier MolFile. Ce fichier contient un en-tête contenant des informations comme le nom de la molécule, le propriétaire, et d'autres commentaires libres. La seconde partie du fichier contient la table de connections (CT) qui référence la liste des atomes de la molécule, les coordonnées x-y-z des atomes, leurs liaisons entre eux et diverses propriétés comme l'isotopie, les charges atomiques, etc.

En raison de leur facilité à manipuler, les formats SMILES et MDL sont les plus couramment utilisés dans les solutions chemoinformatiques. Contrairement aux SMILES, les InChI ne sont pas adaptés pour des utilisations de routine en chemoinformatique, à savoir l'application de réactions chimiques sur une molécule, ou encore la recherche de sous-structures ... De plus, la prise en compte de la mobilité des hydrogènes (notamment liée à des effets de conjugaison lors de l'alternance de liaisons simples et doubles) dans le format InChI peut poser des soucis lors de la génération de la structure moléculaire associée. Cette information est signalée dans la couche d'hydrogène par un bloc bordé par des parenthèses indiquant les atomes affectés par ce processus. Ainsi, l'identifiant InChI de la molécule MeIQx tel qu'il est présenté dans la figure 1.4 et validé dans les différentes bases de données donne une structure chimique différente de celle dessinée. La conversion de cet identifiant délocalise la double liaison entre l'atome 11 et 15 vers les atomes 11 et 12 changeant ainsi le nombre d'hydrogène des atomes 12 et 15. Pour pouvoir retrouver cette structure, il convient de fixer une double liaison en rajoutant une couche sur la stéréochimie et de supprimer les informations sur la mobilité des hydrogènes. L'identifiant "InChI=1S/C11H11N5/c1-6-5-13-7-3-4-8-10(9(7)14-6)15-11(12)16(8)2/h3-5H,12H2,1-2H3/b15-11+" permet de générer la structure chimique telle que présentée dans la figure 1.4. Ainsi, toutes ces contraintes appliquées à l'InChI font du SMILES le format par excellence lors de la manipulation de molécule chimique. Ce format SMILES a donc été utilisé pour ce travail de thèse où des réactions biochimiques sont prédites.



### Nom usuel

MeIQx

### CAS

77500-04-0

### IUPAC

3,8-dimethylimidazo[4,5-f]quinoxalin-2-amine

### SMILES

CN1C(N)=NC2=C1C=CC3=C2N=C(C)C=N3

### InChI

InChI=1S/C11H11N5/c1-6-5-13-7-3-4-8-10(9(7)14-6)15-11(12)16(8)2/h3-5H,1-2H3,(H2,12,15)  
 Couche principale  
 Formule chimique brute      Connectivité inter-atomes      Mobilité des Hydrogènes entre les atomes 12 et 15  
 Connectivité des Hydrogènes

### InChIKey

DVCCCQNKIYNKKB-UHFFFAOYSA-N  
 version (A : 1)      Nombre de protons (N : neutre)  
 Connectivité inter-atomes      Information des 5 autres couches      Format (standard / non standard)

FIGURE 1.4 – Formulation de la molécule MeIQx avec les principaux identifiants moléculaires biologiques et informatiques

### 1.2.3 Les réactions biochimiques

Une réaction biochimique est un processus métabolique modifiant une molécule initiale appelée substrat en un composé nommé produit. On représente usuellement cette réaction par son équation bilan contenant les réactifs, c'est à dire les espèces qui vont être modifiées, les produits, le sens de la réaction, ainsi que la stœchiométrie indiquant les quantités de réactifs et produits nécessaires à l'équilibre réactionnel (figure 1.5). Ces réactions pouvant être informatiquement encodées sont très souvent régies par un catalyseur nommé enzyme qui n'est pas altéré lors du processus réactionnel.

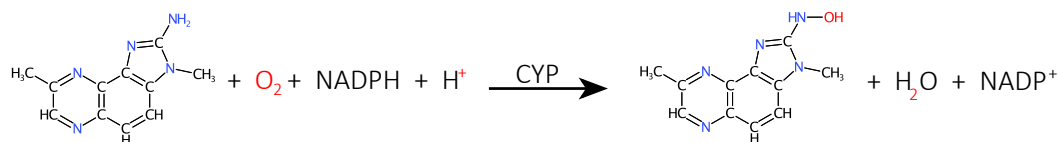


FIGURE 1.5 – Equation bilan de l'hydroxylation du MeIQx catalysé par l'enzyme cytochrome P450 (CYP)

#### 1.2.3.1 Les enzymes

Les enzymes sont des protéines qui agissent en baissant l'énergie nécessaire à l'activation d'une réaction biochimique par la stabilisation de l'état de transition (point irréversible d'une réaction) (figure 1.6). Ce mécanisme permet de faciliter la synthèse de produits. Les enzymes sont de par leurs structures tridimensionnelles très spécifiques vis-à-vis des substrats et des réactions qu'elles catalysent. En d'autres termes, les enzymes catalysent des réactions spécifiques (hydroxylation, glucuronidation sulfatation, etc) et ne peuvent prendre en charge qu'un nombre restreint de substrats. De plus, elles sont capables de distinguer des molécules fortement similaires et de catalyser une réaction sur un atome spécifique. Cette préférence de catalyse d'une réaction sur un atome spécifique du substrat est appelée régiosélectivité.

Les différentes phases de complexation et décomplexation du complexe Enzyme-Substrat (ES) permettent de définir la cinétique d'une réaction. En 1913, L. Michaelis et M. Menten postulèrent une modélisation de la cinétique enzymatique d'une réaction biochimique en deux étapes [13, 45, 104]. La première étape est la fixation réversible du substrat (S) sur l'enzyme (E) pour former le complexe ES. Puis ce complexe est dissocié, libérant ainsi le produit (P) (Figure 1.6). Ce modèle permet de décrire la cinétique enzymatique par une équation connue sous le nom d'équation de Michaelis-Menten.



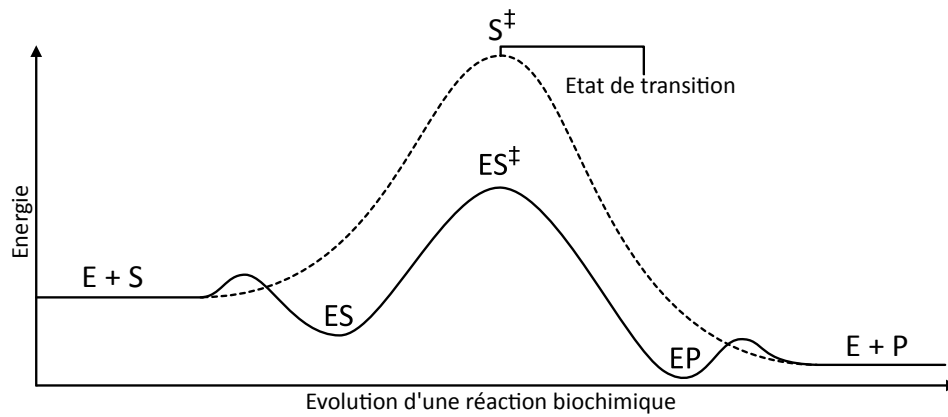


FIGURE 1.6 – **Evolution énergétique d'une réaction biochimique**

Une réaction peut se faire soit spontanément (pointillés) ou soit catalysée par une enzyme (ligne pleine). Lors d'une réaction enzymatique, l'enzyme (E) fixe le substrat (S) dans son site actif et stabilise le complexe nouvellement formé vers un état de transition qui abaisse l'énergie nécessaire à l'activation d'une réaction. Enfin, le produit est libéré de l'enzyme.

Pour décrire la cinétique enzymatique, Michaelis et Menten postulèrent la quasi-stationnarité du complexe ES. C'est à dire :

$$\frac{d[ES]}{dt} = 0 = k_1 \times [E][S] - k_{-1} \times [ES] - k_2 \times [ES]$$

Sachant aussi que la concentration totale d'enzyme dans le modèle est égale à  $[E]_0 = [E] + [ES]$ , le complexe ES équivaut alors à

$$[ES] = \frac{k_1 \times [E]_0 \times [S]}{k_{-1} + k_2 + k_1 \times [S]}$$

La vitesse de la réaction étant dépendante de la formation du produit, alors

$$v = \frac{d[P]}{dt} = k_2 \times [ES]$$

Après intégration :

$$v = \frac{k_2 \times [E]_0 \times [S]}{\frac{k_{-1} + k_2}{k_1} + [S]}$$

Dans cette équation exprimant la cinétique généralement en  $\mu\text{M}$  ( $\mu\text{mol/L}$ ), le terme  $k_2 \times [E]_0$  a été défini comme la vitesse maximale ( $V_m$ ) de la réaction exprimée généralement en  $\mu\text{mol/min}$ . Cette dernière est mesurée quand la concentration de substrat est saturante ( $[S] \gg [E]_0$ ). Le terme  $\frac{k_{-1} + k_2}{k_1}$  désigne la constante de Michaelis ( $K_m$ ) qui est sans unité et spécifique à l'enzyme.

Bien que cette équation date de plus d'un siècle, elle est de nos jours largement utilisée dans la communauté pour modéliser une cinétique de biotransformation [45]. Néanmoins, cette équation ne permet de modéliser que les cas "généraux" simples et n'est pas adaptée lors de cas complexes comme la multiplicité des substrats et/ou des sites catalytiques, des effets de compétitions, l'intervention de régulateurs, ou encore des cinétiques biphasiques.

### 1.2.3.2 Encodage d'une réaction

Tout comme les molécules, les réactions peuvent être encodées en utilisant le langage SMIRKS (SMiles ReaKtion Specification) dérivant des langages SMILES et SMARTS (SMiles ARbitrary Target Specification) [53]. Ce dernier est conçu pour la recherche de sous-structures dans une chaîne SMILES. Le langage SMIRKS permet d'appliquer simplement une réaction chimique sur une molécule par la modification d'une sous-structure de celle-ci (figure 1.7). Chaque expression SMIRKS contient deux parties séparées par le symbole "»". La partie située à gauche du symbole correspond au motif à rechercher dans la molécule et celle à droite, au motif à obtenir.

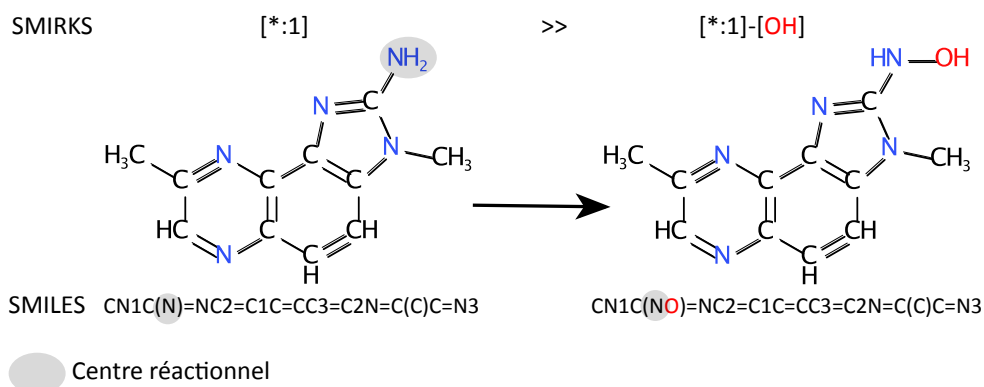


FIGURE 1.7 – Application d’une réaction chimique grâce à l’utilisation du langage SMIRKS

Dans cet exemple, une réaction d’hydroxylation est appliquée sur l’atome d’azote marqué par un cercle gris. Le motif [\* :1] donne l’instruction de rechercher n’importe quel atome. Le motif [\* :1]-[OH] indique que l’atome précédemment sélectionné doit être modifié de manière à être lié à un groupement OH.

### 1.3 Modélisation du métabolisme

La modélisation des systèmes métaboliques a été développée pour répondre à des questions posées aussi bien à l’échelle moléculaire, cellulaire, tissulaire, ou d’un organisme entier [127, 250]. A l’échelle moléculaire (réaction ou métabolite), les objectifs sont multiples et consistent notamment à étudier les interactions enzyme-substrat, à prédire les modifications qu’une molécule peut subir, à étudier l’effet de changements structuraux d’une molécule sur son activité biologique, à comprendre le devenir d’un métabolite, etc. Ces approches regroupent les méthodes de relation structure-activité (QSAR), le docking, les méthodes de prédiction de sites de métabolisme (SOM), et les méthodes de prédiction de réactions biochimiques. Ces dernières peuvent dans une autre mesure prédire des réseaux métaboliques [29, 115]. Les approches à l’échelle cellulaire s’attachent notamment à modéliser le réseau métabolique dans son ensemble. L’étude au niveau cellulaire trouve son importance dans l’étude de la structure d’un système biologique, de l’influence de facteurs sur ce système, et de la dynamique d’un système [117]. Trois types d’approches majeures permettent d’étudier le métabolisme au niveau cellulaire : les approches basées sur les graphes, celles sur les contraintes et les approches cinétiques [141]. Enfin, à l’échelle d’un organe/tissu ou même d’un organisme entier, les approches de modélisation sont très utilisées en pharmacologie dans le cadre de la modélisation du métabolisme d’un médicament, ou toute autre substance chimique dans un organe, une série d’organes ou le corps entier. A titre d’exemple, les modèles pharmacocinétiques physiologiques (PBPK) constituent une famille de modèles des plus utilisés [194, 262]. La diversité des méthodes de modélisation est grande et l’utilisation de l’une d’entre elles dépend généralement de la pro-

blématique. Dans ce travail de thèse, un positionnement à la fois cellulaire et moléculaire a été employé via l'implémentation de méthodes cinétiques et le développement d'une méthode de prédiction moléculaire des xénobiotiques basée sur la prédiction de réactions, combinée à l'utilisation de méthodes de prédiction de sites de métabolisme (SOM) et sites de réactivité (SOR).

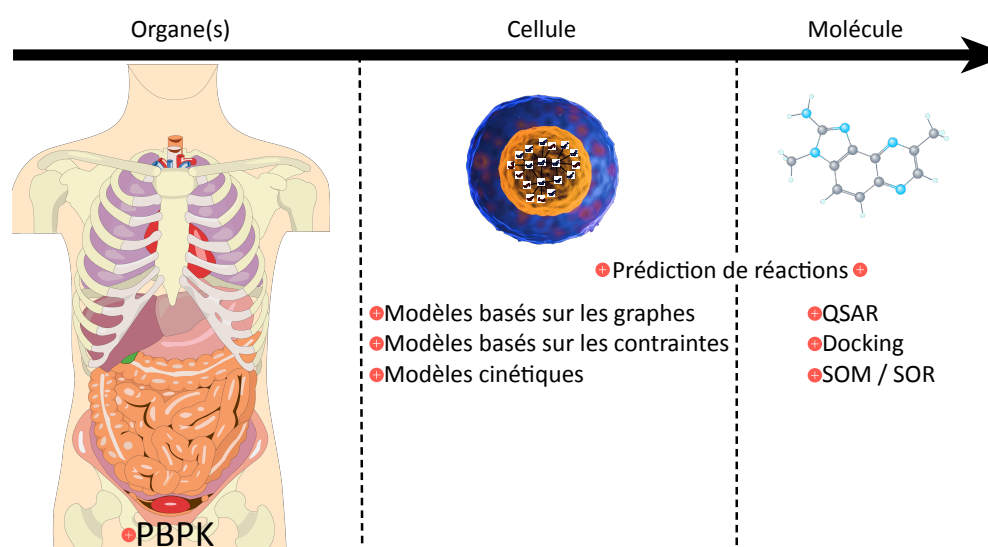


FIGURE 1.8 – Les différentes approches de modélisation en fonction de leur échelle d'étude

Trois échelles de modélisation peuvent être définies : la modélisation à l'échelle d'un ou plusieurs organes, la modélisation à l'échelle cellulaire et la modélisation à l'échelle moléculaire. La figure présente les différentes approches de modélisation en fonction des différentes échelles.

### 1.3.1 Les approches à l'échelle de l'organe, du tissu et de l'organisme

Les modèles PBPK (Physiologically Based Pharmacokinetic) sont des modèles mathématiques dynamiques modélisant l'évolution d'une substance chimique lors des phases d'absorption (introduction de la substance dans un compartiment), de distribution (diffusion de la substance au sein des compartiments), de métabolisme (biotransformation d'une substance chimique), et d'élimination à l'échelle d'un organe, d'un ensemble d'organes ou dans le corps entier [105, 126, 157, 262, 186]. Ces modèles ont vu le jour pour la première fois en 1937 grâce à T. Teorell [218], mais la complexité du modèle rendait à l'époque les équations insolubles. Ce n'est que dans les années 2010 que l'utilisation de ce type d'approche a explosé [194]. Bien que ce type d'approche soit largement utilisé dans le développement de médicaments, il l'est tout autant pour l'étude du devenir de xénobiotiques non médicamenteux, ainsi que pour l'étude de leur toxicologie [38, 37, 90, 123, 137, 144, 256, 236].

Ces modèles sont construits sous forme de compartiments définis par le modélisateur en fonction de la problématique biologique. Chaque compartiment peut correspondre à un organe ou un groupe d'organes, à un tissu, voire même à un composant biologique comme les urines. Ces compartiments sont reliés entre eux par deux compartiments : le compartiment artériel correspondant au sang artériel et le compartiment vasculaire correspondant au sang veineux. De plus, les sites d'absorption, d'ingestion ou d'administration d'une substance chimique doivent être ajoutés au compartiment correspondant (figure 1.9).

Chaque compartiment est modélisé par une équation différentielle (ODE) décrivant l'évolution d'une substance chimique en fonction du temps dans ce compartiment. Ainsi, en cas de présence de plusieurs substances, il y a autant d'ODE par organe que de substances circulantes dans les compartiments à étudier [90]. En d'autres termes, si l'on veut étudier la mise en circulation d'un métabolite dans l'organisme, il faut lui attribuer une équation par organe. Cette étape n'est pas réalisée dans le cas où les produits métaboliques restent confinés dans le compartiment où ils ont été formés. L'ODE correspondant à l'évolution de la concentration d'un composé chimique en fonction du temps dans un organe ( $C_T$ ) comprend les entrées, les sorties et les caractéristiques de l'organe. Parmi les entrées, on trouve le flux sanguin à l'entrée de l'organe ( $f_S$ ), la concentration sanguine de la substance entrant dans le tissu ( $C_i$ ) et le volume sanguin artériel ( $V_A$ ). En fonction des organes, une fonction d'absorption correspondant au lieu d'administration de la substance est ajoutée ( $f(t)$ ). Les sorties comprennent le volume sanguin veineux ( $V_V$ ), la concentration sanguine de la substance sortant dans le tissu ( $C_o$ ), ainsi que parfois son élimination ( $k_E C_T$ ). Les caractéristiques spécifiques au tissu sont : la concentration de la substance dans le tissu ( $C_T$ ), le volume sanguin du tissu ( $V_T$ ) et son métabolisme souvent modélisé par une équation de Michaelis-Menten ( $\frac{V_m \times C_T}{K_m + C_T}$ ). Après intégration, l'ODE correspond à :

$$\frac{dC_T}{dt} = f_S(C_i - C_o) \frac{1}{V_T} - k_E C_T - \frac{V_m \times C_T}{K_m + C_T} + f(t)$$

En définissant le coefficient de partage (P) d'une substance entre l'organe et le sang comme  $P_T = \frac{C_T}{C_o}$ , on obtient l'équation utilisée en PBPK pour modéliser un organe (le gris représente la partie spécifique à certains organes) :

$$\frac{dC_T}{dt} = f_S(C_i - \frac{C_T}{P_T}) \frac{1}{V_T} - k_E C_T - \frac{V_m \times C_T}{K_m + C_T} + f(t)$$

Il est aussi possible de regrouper les compartiments veineux et artériels lorsque les poumons qui permettent de faire le lien entre les deux systèmes sont absents. Cela est effectué lorsqu'ils ne sont pas en mesure d'éliminer la substance chimique. Dans ce cas, l'équation suivante modélise ce nouveau compartiment :

$$\frac{dC_S}{dt} = \frac{1}{V_S} \left( \sum_i^n F_{S_i} C_i - \frac{C_T}{P_T} - F_T C_S + f(t) \right)$$



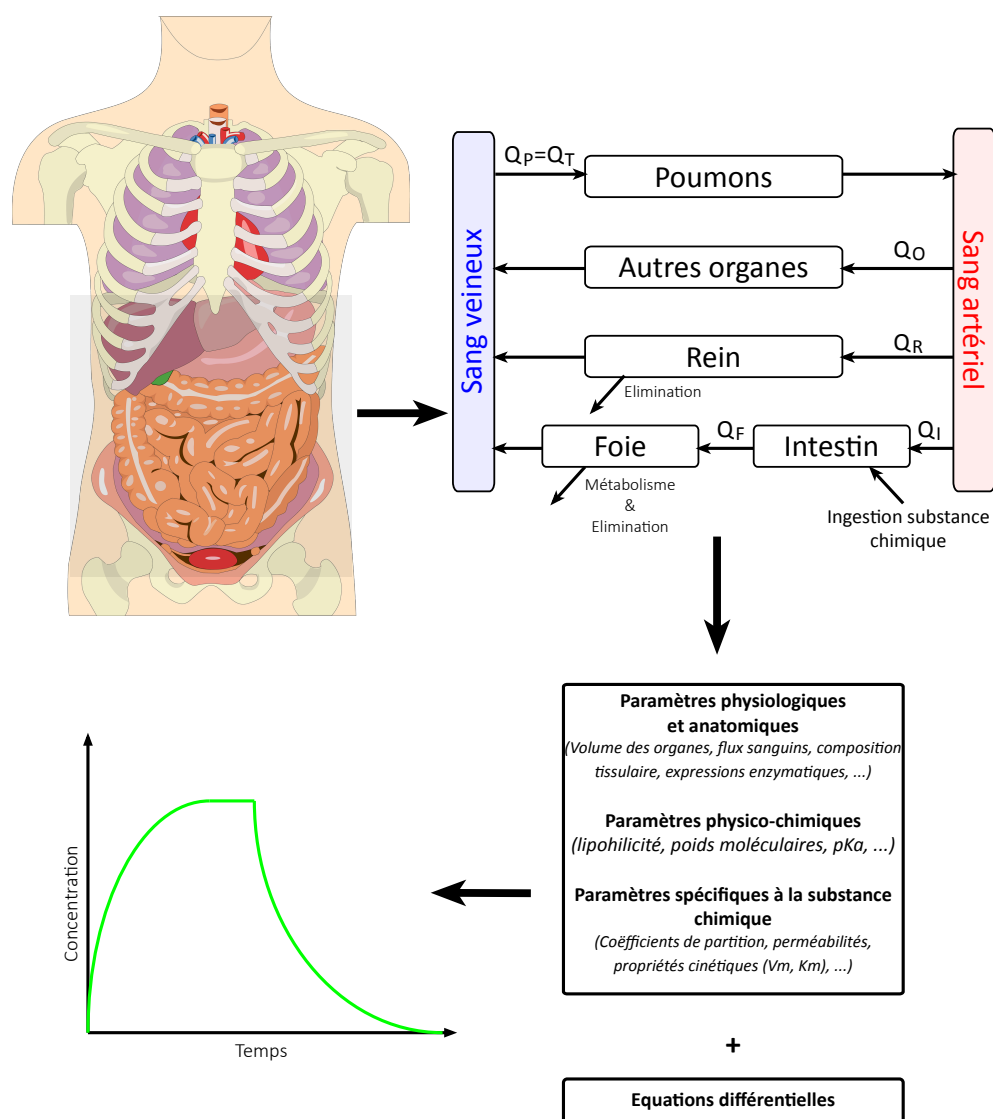


FIGURE 1.9 – Modélisation PBPK

L'approche PBPK consiste à modéliser un ensemble de compartiments incluant différents paramètres de nature physiologique, anatomique, physico-chimique ou encore spécifique à la substance afin de modéliser son évolution lors des phases d'absorption, de distribution, de métabolisme et d'élimination.

où  $C_S$  est la concentration sanguine de la substance,  $V_S$  le volume sanguin artériel et veineux,  $F_{SO}$  le débit sanguin à la sortie de l'organe,  $C_T$  la concentration sortante de la substance,  $P_T$  le coefficient de partition,  $F_T$  le flux sanguin total, et  $f(t)$  la fonction d'administration de la substance [47].

En conclusion, les approches PBPK permettent de modéliser la dynamique de biotransformation d'une substance en fonction du temps, à une échelle pouvant atteindre celle d'un organisme entier. Outre la dynamique, un des avantages majeurs de ces méthodes est la possibilité de modéliser une population spécifique grâce à la combinaison de paramètres anthropométriques, anatomiques et physiologiques [247, 197]. Bien qu'il y ait une flexibilité dans l'utilisation des paramètres comme la possibilité d'intégrer des données d'expression [153], de prédire certains paramètres comme les  $K_m$  [118], ou encore d'extrapoler des paramètres à partir de ceux d'autres espèces [219], ces méthodes sont très paramétrisées, et par conséquent, elles demandent beaucoup de données biologiques, ce qui les rend parfois compliquées à utiliser lorsque la connaissance biologique est incomplète.

### 1.3.2 Les approches à l'échelle cellulaire

Trois types d'approches informatiques principales se distinguent pour étudier un réseau métabolique à l'échelle cellulaire. Ces approches se différencient en fonction de la précision des prédictions et de la complexité des modèles (qualitatives ou quantitatives). Elles sont souvent utilisées en fonction de la taille des graphes, du type de données biologiques présentes et de la problématique d'étude (figure 1.10) [207].

#### 1.3.2.1 Les méthodes basées sur les graphes

Ces méthodes basées sur la théorie des graphes essaient de faire le lien entre la structure d'un réseau et ses fonctions biologiques. L'avantage de ces méthodes est la possibilité de travailler sur de très grands graphes de réseaux métaboliques sans information sur les stœchiométries ou les cinétiques. En revanche, ces méthodes sont des méthodes statiques qui étudient la topologie des graphes en recherchant des critères tels que les coefficients de clustering, les chemins les plus courts, etc [141]. Ces modèles peuvent apporter des informations multiples sur la robustesse et la stabilité d'un réseau, représenter des réseaux métaboliques en modélisant un processus biologique, ou les concurrences de réactions biochimiques au sein de celui-ci [64]. Ils sont très utilisés pour des questions biologiques qui peuvent être : l'identification de voies métaboliques majeures dans le cancer [92], la modélisation des interactions protéine-protéine pour étudier la croissance de cellules tumorales [61], la différenciation entre les patients sains de ceux atteints d'un cancer par l'étude de la courbure des réseaux métaboliques [196], etc.

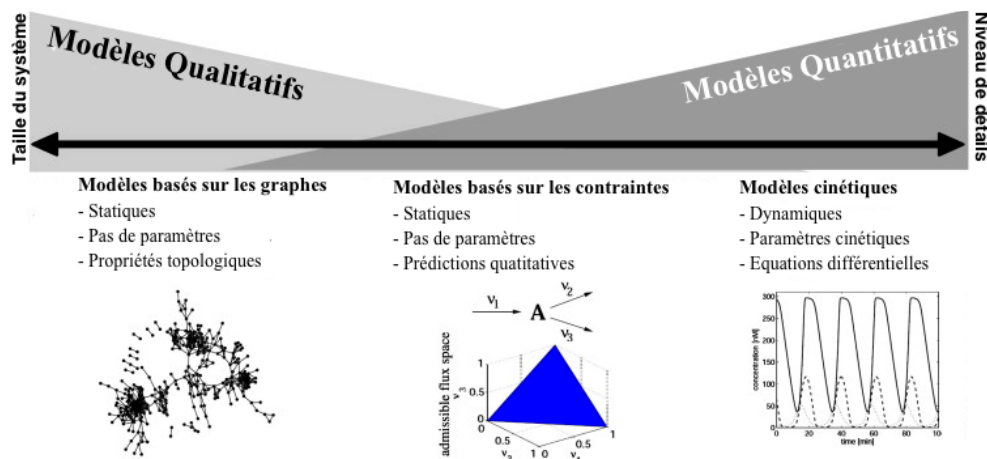


FIGURE 1.10 – **Modélisation à l'échelle cellulaire (adaptée de [207])**

Pour étudier les réseaux métaboliques à l'échelle cellulaire, trois types d'approches sont couramment utilisées en fonction de la taille du système et du niveau de détail. Les modèles qualitatifs basés sur les graphes permettent de faire le lien entre la structure d'un réseau et sa fonction biologique. Les modèles quantitatifs basés sur les contraintes offrent la possibilité d'étudier un réseau en s'affranchissant des données cinétiques. Les modèles quantitatifs cinétiques sont employés pour évaluer la dynamique d'un réseau.

### 1.3.2.2 Les méthodes basées sur les contraintes

Ces méthodes qui s'intéressent à prédire les flux métaboliques combinent les avantages de pouvoir travailler sur de grands graphes, tout en apportant des prédictions quantitatives et s'affranchissant de données cinétiques ou de distributions métaboliques. Pour ce faire, ces méthodes supposent que les cellules atteignent rapidement un état pseudo-stationnaire, c'est-à-dire que les concentrations métaboliques soient constantes. Fort de ces hypothèses, ces méthodes permettent de s'affranchir de paramètres cinétiques. Néanmoins, les modèles basés sur les contraintes nécessitent :

- la stoechiométrie, c'est à dire les contraintes d'équilibre de masse qui maintiennent le système dans un état stationnaire ;
- les contraintes thermodynamiques indiquant le sens des réactions ;
- les capacités enzymatiques régissant les débits maximaux des flux des réactions métaboliques. [180, 184]

Une des méthodes phares dans ce domaine est l'émergence de la FBA (Flux Balance Analysis) dans les années 80 [62, 172], qui s'attache à analyser les flux de métabolites d'un réseau biologique [167]. La FBA consiste à optimiser un phénotype par l'emploi d'une fonction objective qui représente le poids de chaque réaction dans le phénotype optimisé [167]. La formulation d'un problème en FBA se fait en 5 étapes. Dans un premier temps, il convient de définir un réseau métabolique en un ensemble de réactions métaboliques équilibrées (figure 1.11a). Puis, ces réactions sont transposées en

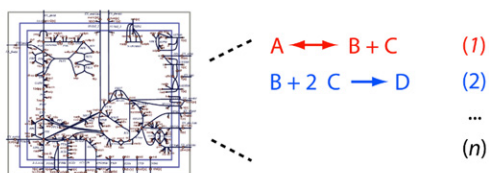
une matrice stœchiométrique  $S$  de taille  $m \times n$ , où  $m$  désigne les métabolites en ligne et  $n$  les réactions du réseau métabolique en colonne. Chaque métabolite consommé à un coefficient stœchiométrique négatif et inversement pour les composés produits qui ont un coefficient positif. Un métabolite ne participant pas à la réaction possède un coefficient nul (figure 1.11b). La troisième étape consiste à définir le vecteur  $v$  de taille  $n$  modélisant le flux qui passe à travers toutes les réactions du système. Les concentrations de tous les métabolites sont aussi initialisées dans un vecteur  $x$  de taille  $m$ . Ainsi, à l'état d'équilibre, le système des équations de masse est défini comme  $Sv = 0$  (figure 1.11c). On notera que ce système d'équations de masse ne contient pas forcément d'unique solution dans le cas, où un réseau métabolique contient plus de réactions que de métabolites. La quatrième étape définit la fonction objective  $Z = cv$  à maximiser ou à minimiser, où  $c$  représente un vecteur de poids indiquant la contribution de chaque réaction dans la fonction objective (figure 1.11d). Enfin, la fonction objective est optimisée par programmation linéaire afin d'identifier une distribution de flux qui maximise ou minimise cette fonction (figure 1.11e).

Fort de son succès, de nombreuses variantes de FBA ont émergé au début des années 2000. C'est ainsi qu'une dérivation de la FBA a abouti à l'élaboration de la dFBA (dynamicFBA) qui s'attache à modéliser la dynamique d'un système en fonction du temps et peut permettre l'intégration de paramètres cinétiques [142]. Ainsi, la FBA a été appliquée avec succès dans le domaine de la santé en contribuant notamment à l'élaboration de nouvelles stratégies thérapeutiques [65, 114], à la prédiction de biomarqueurs [201], à la modélisation de pathologies au sein d'un compartiment [25] ou de plusieurs compartiments [26].

### 1.3.2.3 Les méthodes cinétiques

Cette dernière approche permet d'avoir le plus haut niveau de précision, mais avec la contrainte de ne pouvoir être appliquée que sur des réseaux métaboliques de petite échelle [46, 15, 136]. La modélisation cinétique consiste à prédire l'évolution dynamique des espèces à travers un système d'équations différentielles (ODE) paramétrisées par des paramètres cinétiques [101]. Ces paramètres cinétiques peuvent provenir de la littérature ou être déterminés à l'aide de données expérimentales comme des distributions métaboliques [42]. Dans le cadre de la modélisation du métabolisme d'une substance chimique, il est possible d'utiliser des équations de masse, ou des cinétiques michaelienne en conditions saturantes ou non [42]. Ces approches semblent maintenant délaissées au profit d'approches PBPK par la communauté scientifique s'attachant à répondre à des questions toxicologiques [118, 186, 236]. Néanmoins, lors de la modélisation grande échelle de phénomènes biochimiques complexes difficilement intégrables dans un système plus simpliste, ces méthodes présentent tout leur intérêt en offrant de nouvelles perspectives de modélisation [251].

### a Curate metabolic reactions



### b Formulate S matrix

		Reactions			
		1	2	...	n
Metabolites	A	-1			
	B	1	-1		
	C	1	-2		
	D			1	
	...				
m		S			

### c Apply mass balance constraints

$$\begin{matrix} \mathbf{S} (m \times n) \\ \begin{matrix} -1 & & & \\ 1 & -1 & & \\ 1 & -2 & & \\ & & & 1 \\ & & & & \dots \\ & & & & & 1 \end{matrix} \end{matrix} * \begin{matrix} \mathbf{v} (n \times 1) \\ \begin{matrix} v_1 \\ v_2 \\ \dots \\ v_n \end{matrix} \end{matrix} = 0 \rightarrow \begin{matrix} m \text{ mass balance} \\ \text{equations} \\ -v_1 + \dots = 0 \\ v_1 - v_2 + \dots = 0 \\ v_1 - 2v_2 + \dots = 0 \\ v_2 + \dots = 0 \\ \dots \end{matrix}$$

### d Define objective function Z

$$\mathbf{Z} = \begin{matrix} \mathbf{c}^T (1 \times n) \\ \begin{matrix} 1 & 0 & \dots & 0 \end{matrix} \end{matrix} * \begin{matrix} \mathbf{v} (n \times 1) \\ \begin{matrix} v_1 \\ v_2 \\ \dots \\ v_n \end{matrix} \end{matrix}$$

sets reaction 1 as the objective

### e Optimize Z using linear programming

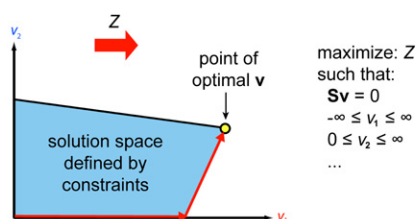


FIGURE 1.11 – Définition d'un problème FBA [167]

La définition d'un problème FBA est réalisée en 5 étapes : définition du réseau métabolique (a) ; formulation de la matrice S associée (b) ; application des contraintes d'équilibre de masse (c) ; définition de la fonction objective Z (d) ; optimisation de la fonction objective Z par programmation linéaire (e).

#### 1.3.2.4 Résumé

Les approches en biologie des systèmes pour l'étude d'un système biologique au niveau cellulaire sont dépendantes de la taille du système et du niveau de précision requis. Néanmoins, le choix d'une méthode dépend en pratique principalement du type et de la quantité de données biologiques disponibles. Ainsi, les approches basées sur les graphes seront privilégiées lors de l'étude de grands réseaux métaboliques en l'absence de données de stœchiométries, tandis que les méthodes basées sur les contraintes seront appliquées à l'étude de larges graphes métaboliques, lorsqu'un aspect quantitatif est nécessaire et que les informations stœchiométriques sont accessibles. Enfin, les approches cinétiques sont favorisées lorsque des données biologiques quantitatives sont disponibles et qu'un niveau de précision important est requis.

Les données peuvent parfois contraindre l'emploi d'une de ces méthodes, alors qu'une étude plus fine du métabolisme est nécessaire. Ainsi, il sera présenté dans le chapitre 5 une méthodologie permettant l'usage de données biologiques limitées dans le cadre de l'implémentation d'une approche cinétique à base d'ODE. Puis, il sera montré dans le chapitre 6 que cette méthode permet la compréhension fine du métabolisme malgré l'utilisation de données biologiques restreintes.

### 1.3.3 Les approches à l'échelle moléculaire

Les méthodes situées au niveau moléculaire se concentrent sur l'étude des réactions biochimiques que peut subir une molécule. Dans le cadre de la prédiction du métabolisme, on distingue principalement deux types de méthodes : les méthodes de prédiction des sites de métabolisme (SOM) d'une molécule et les méthodes de prédiction des réactions que peut subir une substance chimique.

#### 1.3.3.1 Prédiction des sites de métabolisme

La prédiction de sites de métabolisme (SOM) consiste à déterminer les atomes d'une molécule qui peuvent interagir avec une enzyme afin d'être modifiés lors d'une réaction biochimique. Malgré une très grande diversité d'approches implémentées, les méthodes actuelles prédisent les SOM en postulant la présence d'une relation entre la structure chimique d'une molécule et ses propriétés métaboliques à l'aide de l'exploitation de variables décrivant cette structure chimique appelées descripteurs moléculaires. Ces descripteurs sont des propriétés décrivant les caractéristiques moléculaires de la structure d'une molécule chimique [222]. Ils peuvent être de nature constitutionnelle en considérant la composition chimique de la molécule (masse moléculaire, nombre d'atomes, nombre de cycles aromatiques, ...), topologique par l'utilisation d'indices basés sur la structure 2D (indice de Wiener pour le volume moléculaire et le taux de ramification, l'indice de Randic caractérisant l'aire de la molécule accessible au solvant, ...), géométrique définie à partir de la structure 3D (Volume, surface moléculaire, distances, ...) et quantique (énergie de dissociation d'une liaison covalente, descripteurs de réactivité pour caractériser les propriétés de réactivités de composés

chimiques, ...). Ainsi, de nombreuses approches pour caractériser les SOM ont été proposées ces dernières années [165, 115]. On peut classer ces approches au sein de trois catégories : les approches basées sur la réactivité des composants d'une réaction biochimique, les approches utilisant des méthodes d'apprentissage et les méthodes basées sur la structure du substrat et/ou de l'enzyme.

Les approches basées sur la réactivité s'attachent à étudier le changement de réactivité que peut subir le substrat lors d'une réaction biochimique. En effet, lors de la catalyse d'une réaction biochimique dans le site catalytique des CYP, un atome d'hydrogène est supprimé du substrat affectant sa réactivité et permettant ainsi d'estimer l'énergie d'activation. Ainsi, l'énergie d'abstraction d'un hydrogène est un critère important pour déterminer si un atome peut subir une transformation chimique. Il a été démontré en 2003 par Singh et al. qu'un atome possédant une énergie d'abstraction d'hydrogène inférieur à  $27 \text{ kcal.mol}^{-1}$  est plus susceptible de pouvoir être métabolisé et donc d'être un SOM [202]. Cette méthodologie a été adaptée par des outils comme QMBO [6] estimant l'énergie d'abstraction d'un hydrogène, CypScore [89] utilisant des descripteurs de réactivité atomique, Metasite [49] tenant compte de la variation d'énergie d'interaction entre son substrat et une enzyme et le calcul des orbitales moléculaires (les différents états moléculaires d'une molécule), SMARTCyp [192, 193] associant les énergies d'activation avec un descripteur d'accessibilité d'un atome au centre catalytique d'une enzyme.

Des méthodes concurrentes basées sur la structure sont aussi très utilisées pour prédire les SOM. Ces méthodes sont focalisées sur l'utilisation de techniques de machine learning telles que les SVM (Support Vector Machine), les arbres de décision, les réseaux de neurones, etc. Pour ce faire, elles utilisent comme supports, soit des sous-structures de molécules chimiques, soit des descripteurs constitutionnels (type d'atome, l'état d'hybridation, le nombre d'atomes d'hydrogène, ...), ou encore quantiques (énergie d'ionisation, affinité électronique, électronégativité, ...) de la molécule. Ces méthodes ont abouti au développement de logiciels comme RS-Predictor [259] basé sur des modèles SVM et combinant l'utilisation de descripteurs topologiques et quantiques pour prédire la réactivité de sites de réactions métaboliques, Way2Drug SOMP [191, 189] basé sur l'exploitation d'un algorithme bayésien, et Xenosite Metabolism 1.0 [260] exploitant les réseaux de neurones pour prédire les SOM.

Enfin, des approches automatisées ou non focalisées sur la structure du ligand ou du complexe ligand-protéine ont été développées. Elles utilisent, soit les similarités de structures des ligands pour prédire l'interaction avec une enzyme, soit des techniques de docking s'intéressant aux interactions ligand-protéine. Le docking est une technique qui calcule l'orientation privilégiée d'une molécule dans le site catalytique d'une enzyme. Les scores résultant de ces calculs permettent d'identifier les conformations possibles et donc de prédire si un atome est un SOM en fonction de sa proximité au centre catalytique de l'enzyme. Parmi ces approches, on peut notamment citer ID-Site [138] basé sur un workflow pour prédire les SOM, IMPACTS [35] regroupant le

docking avec la modélisation de l'état de transition d'une réaction (voir figure 1.6) et la prédiction de la réactivité du substrat basée sur des règles, MLite combinant l'utilisation de docking et l'estimation d'énergie d'activation [164], et le docking basé sur Ensemble [36] prenant en compte la dynamique moléculaire associée à un ensemble de conformations enzymatiques.

Malgré le nombre important d'approches permettant de prédire les SOM, les applications ont l'inconvénient d'être spécialisées par leurs prédictions focalisées principalement sur les enzymes CYP. Ce choix peut être justifié par l'implication des cytochromes P450 à 75% dans le métabolisme total des xénobiotiques [78]. De plus, la diversité des approches peut rendre le choix d'une méthode par rapport à une autre compliqué, d'autant plus que des analyses comparatives ont montré les qualités individuelles des différentes méthodes implémentées [216, 205]. Enfin, la prédiction des SOM aboutit généralement à un classement de chaque atome en fonction de sa probabilité d'être un SOM, ou sous forme d'un classement des 3 ou 5 meilleurs atomes candidats pour être un SOM. Il peut donc être difficile de distinguer et d'identifier les SOM réels. En effet, certaines molécules peuvent subir une modification par une enzyme sur un seul ou sur de multiples atomes.

### 1.3.3.2 Prédiction de réactions biochimiques

Alors que de nombreuses méthodes ont été développées ces dernières années pour prédire les SOM, celles consistant à prédire les réactions biochimiques sont plus restreintes. En effet, la littérature référence deux approches majeures : les systèmes experts et celle basée sur des empreintes moléculaires (Figure 1.12).

Lors de la prédiction de réactions métaboliques, les systèmes experts sont basés sur l'exploitation de règles de transformations métaboliques extraites de l'analyse de la littérature et d'expériences biologiques. Ces règles stockées dans un dictionnaire consistent à associer une sous-structure chimique cible à une sous-structure produite. Ainsi, un métabolite est prédit par la recherche d'une sous-structure de la molécule d'intérêt dans le dictionnaire de réactions métaboliques et, lorsque celle-ci est retrouvée, ce fragment est remplacé par le fragment produit du dictionnaire. Pour la règle  $\text{CH}_3\text{-NH}_2 \rightarrow \text{CH}_3\text{-NH-OH}$ , toutes les sous-structures  $\text{CH}_3\text{-NH}_2$  dans une molécule requête seront remplacées par le fragment  $\text{CH}_3\text{-NH-OH}$ . Ainsi, dès qu'une règle est trouvée dans le dictionnaire, son métabolite est généré, ce qui peut en fonction de la taille du dictionnaire, générer une combinatoire très importante. Parmi les logiciels utilisant les systèmes experts, MetabolExpert qui a été créé en 1985 était le pionnier pour la prédiction de réactions biochimiques [52]. Puis, des approches prédictives comme META [119, 214, 120], Meteor-Nexus [33, 148], SyGMA [185], MetaDrug [59], TIMES [152], UM-PPS [72] et plus récemment PROXIMAL [257] ont émergé. Ces méthodes se différencient principalement par la taille de leurs dictionnaires (nombre de règles réactionnelles enregistrées), ainsi que par leur capacité à diminuer la combinatoire.



La deuxième approche implémentée au sein de MetaPrint2D-React est basée sur l'utilisation d'empreintes moléculaires [5]. Elle consiste à générer une empreinte moléculaire pour chaque atome d'une molécule et rechercher sa similarité avec celles contenues dans une base de données. Le métabolite sera générée en fonction du succès de la correspondance de l'empreinte créée avec celles présentes dans la base de données. Cette base de données est construite à partir de la base de données anciennement nommée Accelrys Metabolites Database, maintenant rachetée par Dassault. En 2008, cette dernière contenait 87446 réactions métaboliques [5]. La construction de cette base de données consiste à encoder en une empreinte moléculaire chaque SOM de chaque couple substrat-métabolite contenu dans la base de données Accelrys Metabolites Database.

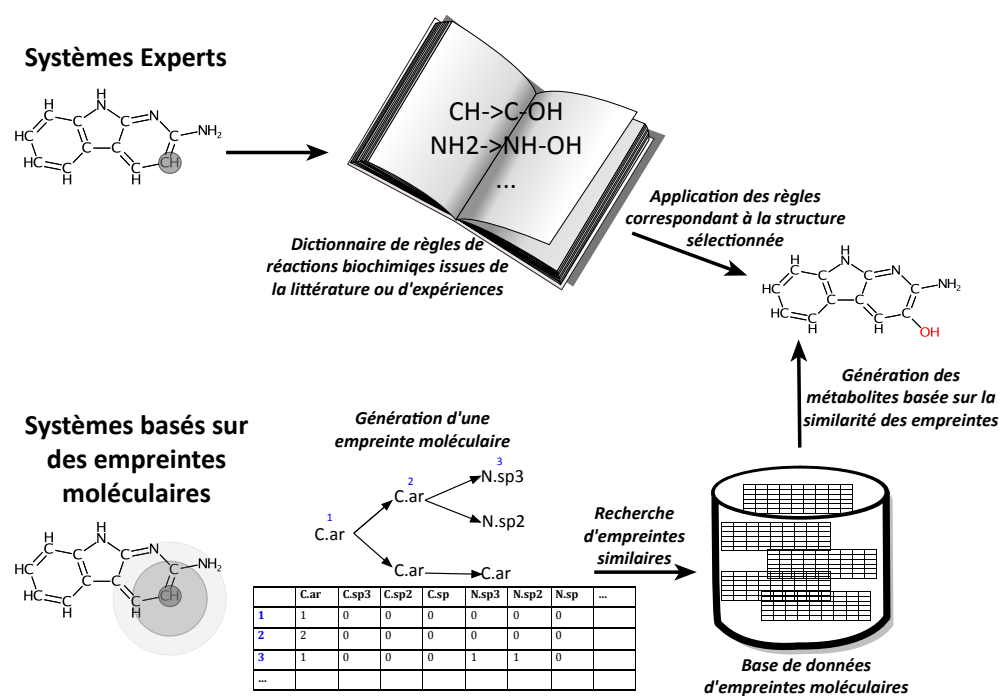


FIGURE 1.12 – Prédiction des métabolites

Deux types d'approches distinctes permettent de prédire le métabolisme des xénobiotiques : les systèmes experts et les systèmes basés sur les empreintes moléculaires. Les systèmes experts génèrent des métabolites en fonction de règles réactionnelles contenues dans un dictionnaire, alors que les systèmes basés sur les empreintes moléculaires utilisent des empreintes de la molécule qui sont comparées avec celles contenues dans une base de données afin de générer les éventuels métabolites correspondants.

Ces systèmes permettent donc de prédire des métabolites et sont capables de prédire un réseau métabolique par itérations successives. Néanmoins, un défaut majeur

de ces systèmes est la combinatoire importante générée. Il est donc important de développer des méthodes pour améliorer la filtration de ces prédictions afin de les rendre plus précises et en meilleure adéquation avec la réalité biologique [115].

### 1.3.3.3 Résumé

Pour étudier une réaction biochimique à des fins de prédictions, deux approches majeures à l'échelle moléculaire sont usuellement employées. Ainsi, les méthodes de prédiction des sites de métabolisme se focalisent à déterminer les atomes pouvant être modifiés par une enzyme donnée en prenant en considération un ensemble de propriétés moléculaires. Les méthodes de prédiction de réactions biochimiques se focalisent à identifier la présence de motifs et y appliquer les réactions correspondantes, indépendamment des propriétés moléculaires de la substance chimique. Alors que les approches basées sur les SOM sont restreintes à une quantité limitée d'enzymes, celles permettant la prédiction des réactions biochimiques d'un large panel de modifications chimiques sont associées à une explosion de la combinatoire. Ainsi, le chapitre 3 vise à développer une méthode de construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions par l'association des deux approches précédemment présentées. De ce fait, ce chapitre aura aussi pour vocation d'approfondir l'usage de ces méthodes tout en les détaillant plus précisément.

### 1.3.4 Cas particulier des processus de toxicité

La toxicité désigne l'ensemble des effets néfastes subis par un organisme à la suite d'une exposition à une substance chimique et en corrélation avec la dose et le temps d'exposition. Cette toxicité peut provenir de l'effet pharmacologique d'une substance (exagération de l'effet d'un médicament, ou déclenchement d'effets secondaires) ou lors de la formation d'intermédiaires métaboliques réactifs avec des protéines, ou l'ADN. Dans le contexte des xénobiotiques, il est donc important d'identifier les métabolites réactifs notamment vis-à-vis de l'ADN pouvant être à l'origine de cancers (voir chapitre 2 section 2.2). Pour se faire, trois méthodes majeures sont utilisées : les approches basées sur la structure et l'utilisation de références croisées, celles basées sur les relations structure-activité (QSAR) et celles basées sur l'apprentissage par deep learning [182, 93] (figure 1.13)

La prédiction de la toxicité basée sur la structure procède par l'utilisation de références croisées pour prédire la toxicité d'une molécule inconnue à partir de molécules analogues et/ou d'utilisation de règles (SI *structure\_chimique* EST *présente/absente* ALORS *toxicité/non toxicité*). Ces méthodologies sont implémentées dans des logiciels comme Derek Nexus [148], HazardExpert [158], AMBIT [100], Toxtree [174].

Le deuxième type d'approche majeure très utilisée est l'approche basée sur les modèles QSAR. Dans le cadre d'une étude toxicologique, elle peut être connue dans la littérature sous l'appellation QSTR (Quantitative Structure-Toxicity Relationship).

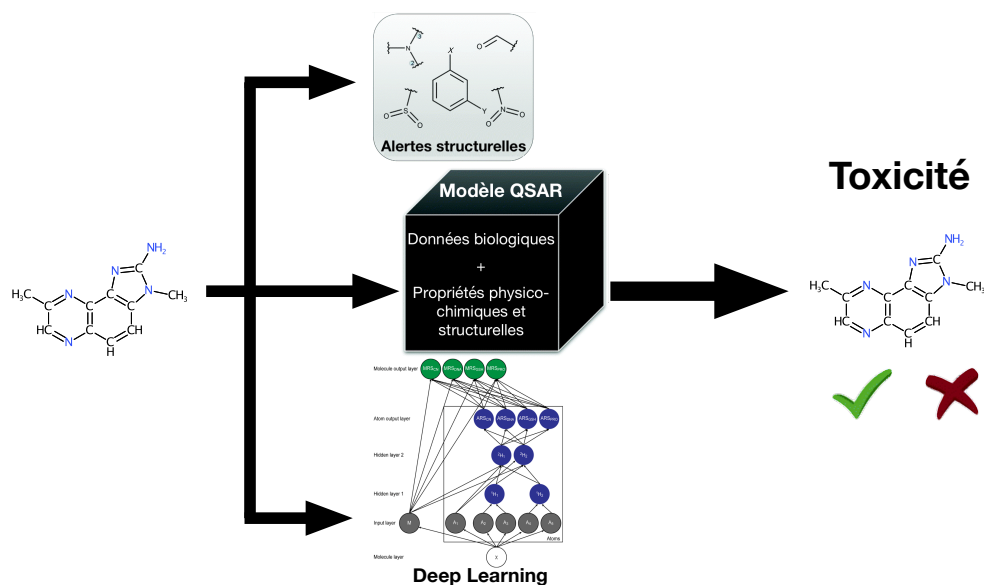


FIGURE 1.13 – Les trois approches majeures pour la prédiction *in silico* de la toxicité (adaptée de [11, 93])

Trois types d'approches majeures permettent de prédire la toxicité d'une substance chimique : les approches basées sur la structure et l'utilisation de références croisées, celles basées sur les relations structure-activité (QSAR) et celles basées sur l'apprentissage par deep learning

Cette approche consiste à relier l'activité toxique à la structure d'un composé à partir de descripteurs moléculaires décrivant cette activité. Ainsi, la toxicité est une fonction mathématique de ces descripteurs telle que :  $\text{Toxicité} = f(\text{descripteurs})$ . Les modèles QSAR sont principalement des méthodes de régression linéaire ou non linéaire. Cette approche est notamment implémentée au sein de TOPKAT [4], Derek Nexus [148], HazardExpert [158], AMBIT [100], et MetaTox [190]

Quelques approches utilisant des réseaux de neurones ont été développées pour prédire la toxicité d'un composé chimique. Parmi celles-ci, ADMET Predictor [179], Xenosite Reactivity 2.0 [93, 94] permettent de prédire la toxicité. De plus, Xenosite Reactivity 2.0 offre aussi la possibilité d'identifier les atomes d'une molécule chimique réactifs vis-à-vis de l'ADN.



# Chapitre 2

## Les xénobiotiques : métabolisme et cancer

Ce deuxième chapitre d'exposition du contexte de cette thèse se focalisera sur la partie biologique sur laquelle les méthodes informatiques développées seront appliquées. Tout d'abord, le contexte biologique sera présenté à travers la présentation du métabolisme d'une famille de xénobiotiques : les amines hétérocycliques aromatiques (AHA). Puis, les challenges associés à leurs métabolisme et toxicité chez l'homme seront introduits. Enfin ce chapitre sera conclu par un résumé général des méthodes informatiques précédemment introduites et par l'introduction du positionnement de ce travail de thèse.

---

<b>2.1 Généralités</b>	<b>39</b>
<b>2.2 Métabolisme des xénobiotiques</b>	<b>40</b>
<b>2.3 Les xénobiotiques de type AHA</b>	<b>42</b>
2.3.1 Le métabolisme général des AHA	44
2.3.2 Le métabolisme des AHA par les enzymes de Phase I	46
2.3.3 Le métabolisme des AHA par les enzymes de Phase II	51
<b>2.4 Enjeux de modélisation : un challenge informatique et biologique</b>	<b>53</b>
<b>2.5 Résumé et principaux apports</b>	<b>54</b>

---

### 2.1 Généralités

Le cancer est l'une des causes principales de mortalités dans le monde provoquant chaque année environ 8.2 millions de morts [99]. Bien que le cancer puisse être induit par des facteurs endogènes, c'est à dire des facteurs génétiques, il est admis que les facteurs environnementaux et alimentaires sont très largement responsables de sa genèse. Ainsi, certains scientifiques estiment que jusqu'à 80% des cancers observés

chez l'homme sont imputés aux contaminants environnementaux et alimentaires [21] connus sous le terme des xénobiotiques lorsqu'ils sont présents dans un organisme vivant. Les xénobiotiques sont donc des substances chimiques étrangères artificielles présentes dans un organisme vivant, et qui sont issues ni de l'alimentation naturelle ni produites par ce corps vivant. Ils sont souvent assimilés aux médicaments et pesticides, bien qu'ils ne soient pas exclusifs à ces deux classes de xénobiotiques.

## 2.2 Métabolisme des xénobiotiques

Le métabolisme général des xénobiotiques (Figure 2.1), substances étrangères à l'organisme, désigne l'ensemble des voies métaboliques qui modifient la structure biochimique de ces composés par l'action successive de diverses enzymes. Cette activation métabolique agit de façon à éliminer et détoxifier les contaminants en trois phases successives : la phase de fonctionnalisation (Phase I), la phase de conjugaison (Phase II) et la phase d'excrétion (Phase III) [140]. Néanmoins, certains intermédiaires du métabolisme de ces substances carcinogènes peuvent s'avérer toxiques. En effet, certains métabolites intermédiaires sont très réactifs et peuvent interagir avec des molécules comme l'ADN, l'ARN et les protéines [199]. Dans le cas de fixation covalente à l'ADN, il y a formation d'adduits à l'ADN qui, s'ils ne sont pas réparés, sont à l'origine de mutations après réplication cellulaire, et ainsi à l'origine du développement du cancer. Deux particularités du métabolisme des xénobiotiques concernent leurs restrictions en termes de profondeur des voies métaboliques et des familles enzymatiques limitées impliquées dans les réactions.

Le métabolisme des xénobiotiques est sujet à de nombreuses variabilités dépendantes de facteurs physico-chimiques, endogènes et exogènes [48]. Ainsi, les xénobiotiques peuvent se différencier sur un plan physico-chimique par des différences chimiques telles que leur composition chimique, leur polarité, leur électrophilicité, leur nucléophilicité, et leur lipophilicité, leurs différences d'encombrement stérique entravant le métabolisme de ces molécules. De plus, des facteurs endogènes comme l'âge, le sexe, l'espèce, l'état pathologique, les différences génétiques inter-individus sont d'autres causes sujettes à de nombreuses variabilités. A ces facteurs, la complexité de la variabilité affectant le métabolisme est encore augmentée par des facteurs exogènes comme la dose d'exposition, la voie d'administration ou les apports alimentaires. Tous ces exemples illustrent parfaitement la complexité du métabolisme et de sa variabilité conséquente.

Afin, d'exposer plus en détail quelques problèmes de variabilité affectant le métabolisme, les sections suivantes porteront sur la description du métabolisme de la famille des xénobiotiques des AHA, qui est une famille adaptée à la résolution informatique de problématiques affectant les xénobiotiques en général.

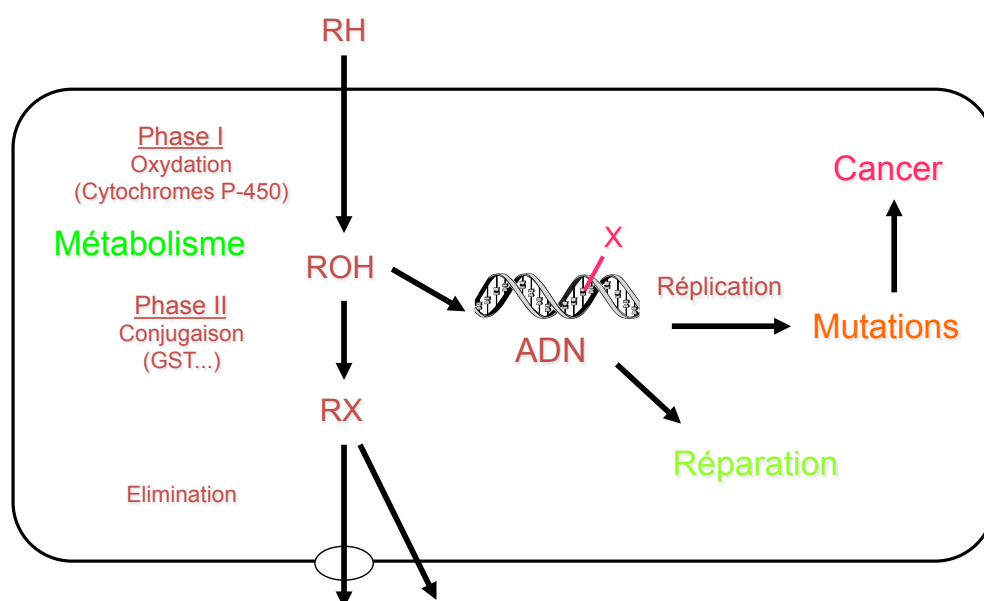


FIGURE 2.1 – Métabolisme des carcinogènes chimiques

Les xénobiotiques RH sont tout d'abord bioactivés par les enzymes de phase I contenant notamment les CYP, puis le produit résultant ROH peut être transformé par les enzymes de phase II (UGT, SULT, NAT, GST) pour permettre son élimination (RX) lors de la phase III, ou encore, il peut être modifié de façon à produire un métabolite réactif (X) vis-à-vis de l'ADN. La fixation d'une molécule réactive (X) sur l'ADN forme un composé connu sous le terme adduit à l'ADN. La formation d'adduits à l'ADN est à l'origine de carcinogénèse et de modifications d'expression génétique, s'ils ne sont pas réparés par l'organisme.



### 2.3 Les xénobiotiques de type AHA

Les amines hétérocycliques aromatiques (AHA) sont une famille de xénobiotiques considérées comme potentiellement cancérigènes pour l'homme [96, 95], et principalement issues des cuissons de viandes et poissons bien cuits [63, 213, 212], dans une moindre mesure de la fumée de cigarette [175, 147, 151, 215, 95] et d'huile de cuisson [255, 43]. Les prémices de l'étude des xénobiotiques de type AHA remontent à 1939, où le chimiste suédois E. M. P. Widmark fut le premier à démontrer la présence de substances cancérigènes dans la viande bien cuite [245]. Il a mis en évidence la formation de tumeur dans les glandes mammaires de souris après expositions cutanées multiples de ces substances contenues dans la viande bien cuite. Néanmoins, ce n'est qu'à partir de 1977 que les études sur les AHA ont réellement débuté grâce à la découverte de composés hautement mutagéniques dans les parties bien cuites de poissons [212]. Depuis cette découverte par Sugumira et al., l'évaluation de l'impact des AHA sur la cancérogenèse est une problématique de santé publique. Les recherches sur leurs identifications et caractérisations, l'étude de leurs mécanismes de formation, leurs métabolismes et leurs génotoxicités ont fait l'objet de nombreuses publications au cours de ces dernières années. En 2004, plus de 2100 articles sur les AHA avaient été publiés [121]

Les différentes études sur cette famille de xénobiotiques ont conduit à l'identification d'une trentaine d'AHA [73, 160, 161, 171, 211, 223, 234]. Il a été aussi démontré que les quantités d'AHA présentes dans les viandes surcuites sont fonction de l'importance de la température et du temps de cuisson [234]. Ainsi, les AHA ont pu être séparées en deux familles en fonction de la température à laquelle ils se forment. La première classe est les AHA de type pyrolytique (Table 2.1), aussi connu sous le terme amino-carbolines qui regroupent les carbolines et qui ont la particularité d'être généralement synthétisés à des températures supérieures à 250°C [234], bien qu'ils puissent être aussi formés à plus faible température, mais en très faible quantité [234]. 2-Amino-9H-pyrido[2,3-*b*]indole (A $\alpha$ C) qui est présent en grande quantité dans la fumée de cigarette est un représentant majeur de cette catégorie. La seconde famille d'AHA regroupe les molécules formées principalement par la réaction de Maillard, réaction chimique entre les acides aminés et la réduction des sucres, à des températures comprises entre 150 et 250°C [234] et sont aussi connues sous le nom d'aminoimidazoazaarènes (AIA) (Table 2.2). Ces dernières étant habituellement produites à l'issue de cuisson d'aliments à des températures usuelles comportent deux AHA très présentes dans les viandes bien cuites et populaires dans la communauté scientifique : le 2-amino-3,8-diméthylimidazo[4,5-*f*]quinoxaline (MeIQx) et le 2-amino-1-méthyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP). Grâce à la mise en place d'études épidémiologiques, le risque de cancer lié à l'ingestion d'AHA est aujourd'hui avéré [30, 41, 163]. D'après, ces études, il est admis que ce risque est dépendant du degré de cuisson de la viande ainsi que de sa fréquence d'exposition.

---

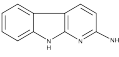
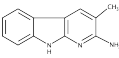
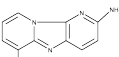
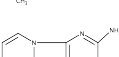
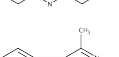
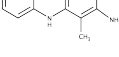
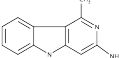
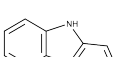
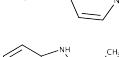
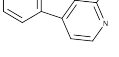
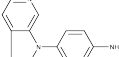
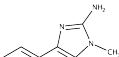
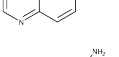
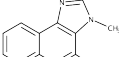
Image	Name	IUPAC Name
Pyrolysis Heterocyclic Aromatic Amines		
	AαC	9H-pyrido[2,3-b]indol-2-amine
	MeAαC	3-methyl-9H-pyrido[2,3-b]indol-2-amine
	Glu-P-1	10-methyl-1,3,8-triazatricyclo[7.4.0.0 <sup>2,7</sup> ]trideca-2(7),3,5,8,10,12-hexaen-4-amine
	Glu-P-2	1,3,8-triazatricyclo[7.4.0.0 <sup>2,7</sup> ]trideca-2(7),3,5,8,10,12-hexaen-4-amine
	Tr-P-1	1,4-dimethyl-5H-pyrido[4,3-b]indol-3-amine
	Tr-P-2	1-methyl-5H-pyrido[4,3-b]indol-3-amine
	NorHarman	9H-pyrido[3,4-b]indole
	Harman	1-methyl-9H-pyrido[3,4-b]indole
	APNH	4-{9H-pyrido[3,4-b]indol-9-yl}aniline
	AMPNH	2-methyl-4-9H-pyrido[3,4-b]indol-9-ylaniline
	Phe-P-1	5-phenylpyridin-2-amine

TABLE 2.1 – Liste des principales AHA pyrolytiques actuellement identifiées

Image	Name	IUPAC Name
Aminoimidazoarene Heterocyclic Aromatic Amines		
	IQ	3-methyl-3H-imidazo[4,5-f]quinolin-2-amine
	MeIQ	3,4-dimethyl-3H-imidazo[4,5-f]quinolin-2-amine
	IQx	3-methyl-3H-imidazo[4,5-f]quinoxalin-2-amine

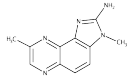
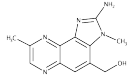
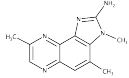
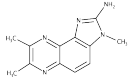
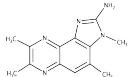
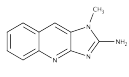
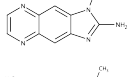
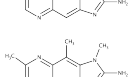
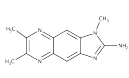
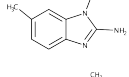
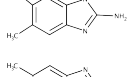
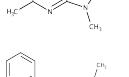
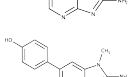
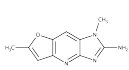
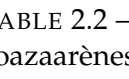

	MeIQx	3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine
	4-CH <sub>2</sub> OH-8-MeIQx	{2-amino-3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-4-yl}methanol
	4,8-DiMeIQx	3,4,8-trimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine
	7,8-DiMeIQx	3,7,8-trimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine
	4,7,8-TriMeIQx	3,4,7,8-tetramethyl-3H-imidazo[4,5-f]quinoxalin-2-amine
	IQ[4,5-b]	1-methyl-1H-imidazo[4,5-b]quinolin-2-amine
	IgQx	1-methyl-1H-imidazo[4,5-g]quinoxalin-2-amine
	7-MeIgQx	1,7-dimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine
	7,9-DiMeIgQx	1,7,9-trimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine
	6,7-DiMeIgQx	1,6,7-trimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine
	1,6-DMIP	1,6-dimethyl-1H-1,3-benzodiazol-2-amine
	1,5,6-TMIP	1,5,6-trimethyl-1H-1,3-benzodiazol-2-amine
	3,5,6-TMIP	3,5,6-trimethyl-3H-imidazo[4,5-b]pyridin-2-amine
	PhIP	1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine
	4'-OH-PhIP	4-{2-amino-1-methyl-1H-imidazo[4,5-b]pyridin-6-yl}phenol
	IFP	6,11-dimethyl-10-oxa-2,4,6-triazatricyclo[7.3.0.0.3.7]dodeca-1(9),2,4,7,11-pentaen-5-amine

TABLE 2.2 – Liste des principales AHA actuellement identifiées de type aminoimidazozaarènes (AIA)

### 2.3.1 Le métabolisme général des AHA

Le métabolisme des AHA est assuré par les actions successives enzymes de phase I et II. La première étape du métabolisme aboutit à la modification de la fonction amine

des AHA, qui est généralement hydroxylée par les enzymes de phase I (Figure 2.2). Cette phase de fonctionnalisation de l'AHA est grandement assurée par cytochromes P450 (CYP) [24]. La phase de conjugaison assure la prise en charge des métabolites hydroxylés lors de la phase I. Cette dernière voie tend à augmenter le poids moléculaire et à réduire l'activité du métabolite produit. Ainsi, les métabolites hydroxylés sont transformés par les enzymes de phase II, qui regroupent les N-acétyltransférases (NAT), les sulfotransférases (SULT), et les UDP-glucuronosyl transférases (UGT). Durant cette phase, les UGT agissent comme agents de détoxification grâce à l'ajout d'un groupement glucuronide sur les AHA N-Hydroxylées. Une fois les métabolites glucuronidés, ils sont excrétés par des protéines de transport membranaire en dehors de l'hépatocyte, afin d'être éliminés. Les NAT [85] et les SULT [44] sont quant à elles responsables de l'estérification des AHA N-Hydroxylées, qui une fois estérifiées, sont réactives vis-à-vis de l'ADN et engendrent conséquemment la formation d'adduits à l'ADN [199].

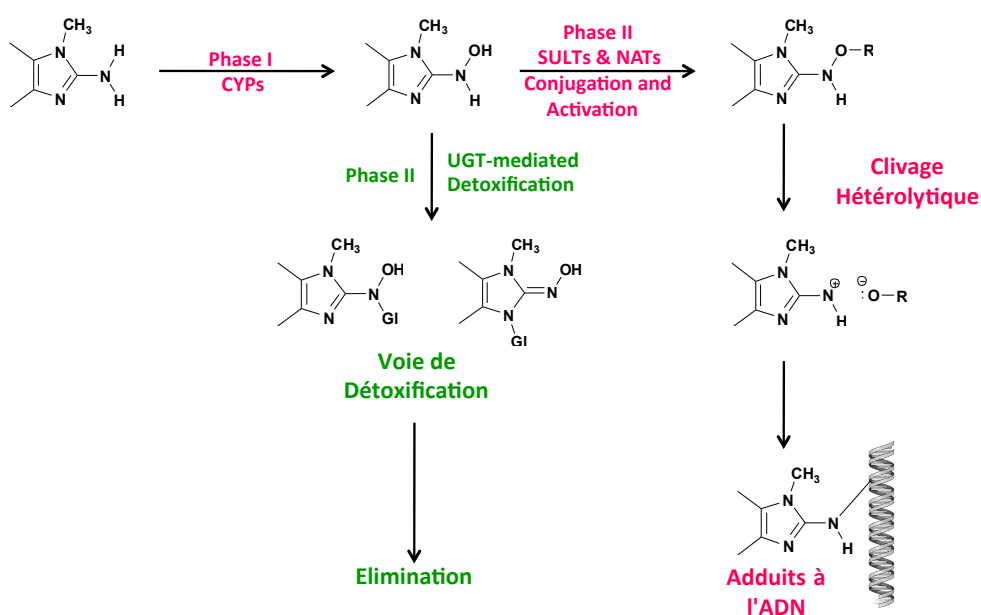


FIGURE 2.2 – Voies de bioactivation des AHA

Lors de la phase I, les AHA sont d'abord modifiées généralement par une hydroxylation (greffe d'un groupement alcool OH) médiée par les cytochromes P450 (CYP). Le composé résultant est métabolisé pendant la phase II, soit par les UDP-glucuronyltransférases (UGT) qui détoxifient le métabolite par l'ajout d'un groupement glucuronide ( $C_6H_{10}O_7$ ), ou soit par les sulfotransférases (SULT) greffant un groupement sulfate ( $H_2SO_4$ ) ou les N-acétyltransférases (NAT) greffant un groupement acétyl ( $COCH_3$ ) sur une fonction amine. Une prise en charge par ces deux dernières conduit à l'activation du composé chimique qui après clivage hétérolytique ira se fixer à l'ADN pour former un adduit à l'ADN.

Le métabolisme des AHA aboutissant à la formation d'adduits à l'ADN a été montré dans tous les tissus investigués tels que le colon, le rectum, le rein et le foie même s'il n'est pas vecteur de cancérogenèse dans tous les tissus [14, 213, 212, 210]. Néanmoins, il a été démontré expérimentalement que le foie est le principal organe assurant la biotransformation et la bioactivation des AHA [253, 106, 112], tout en étant aussi le lieu privilégié de la carcinogenèse via une formation d'adduits à l'ADN pour la majorité des AHA [199, 213].

### 2.3.2 Le métabolisme des AHA par les enzymes de Phase I

Les AHA sont en premier lieu transformées par les cytochromes P450 (CYP) catalysant l'oxydation d'une fonction amine ou sur un carbone [19, 133, 132]. Cette première étape du métabolisme est très majoritairement assurée dans le foie par le cytochrome 1A2 (CYP1A2) [199], qui représente environ 15% du contenu hépatique de CYP [77]. Ainsi, cette enzyme qui est régiosélective pour la N-oxydation des AHA assure notamment la transformation de A $\alpha$ C, MeIQx, et PhIP en A $\alpha$ C-HN<sup>2</sup>-OH (Figure 2.3), HON-MeIQx (Figure 2.4) et HON-PhIP (Figure 2.5). Bien que la bioactivation médiée par les CYP1A2 par l'oxydation des fonctions amines soit la voie métabolique privilégiée par cette enzyme, elle a aussi été caractérisée pour catalyser des réactions non régiosélectives impliquées dans la détoxification [233]. Ainsi, il a été démontré que le CYP1A2 peut aussi médier l'oxydation du carbone C8 du MeIQx qui aboutit à la formation du 8-CH<sub>2</sub>OH-MeIQx, qui est à nouveau réoxydé sur le même carbone pour former le IQx-COOH identifié comme un produit de détoxification [132]. Il a tout d'abord été démontré en 1994 que le CYP1A2 contribuait à métaboliser 91% du MeIQx et 70% du PhIP consommés [24]. Puis plus récemment, il été prouvé que la formation de HON-MeIQx et ses conjugués glucuronés, HON-PhIP et ses conjugués glucuronés, 8-CH<sub>2</sub>OH-MeIQx et IQx-COOH était réduite de jusqu'à 95% [133, 132], ce qui souligne le rôle majeur du CYP1A2 dans le métabolisme de phase des AHA par les enzymes de phase I. Toutefois, A $\alpha$ C et PhIP sont aussi connues pour pouvoir subir une oxydation d'un cycle aromatique formant un composé inoffensif. L'oxydation d'A $\alpha$ C sur les carbones C3 et C6 conduit à la formation de deux métabolites hépatiques majeurs : 3-OH-A $\alpha$ C et 6-OH-A $\alpha$ C (Figure 2.3). Le PhIP peut quant à lui être biotransformé en un métabolite mineur qui est le 4'HO-PhIP (Figure 2.5).

Un aspect majeur à prendre en compte dans le métabolisme des AHA par les enzymes de phase I concerne les variabilités multiples inter et intra-espèces du sentier métabolique catalysé par le CYP1A2. Un premier niveau de variabilité de cette voie a été montré dans les différences inter-espèces qui jouent un rôle dans la régiosélectivité vis-à-vis des AHA, conduisant à la formation privilégiée différente de composés chimiques, ainsi que des différences au niveau des concentrations enzymatiques et des activités catalytiques [228, 231]. Il a été ainsi identifié que les concentrations enzymatiques de CYP1A2 étaient significativement plus importantes chez l'homme que chez le rat [228], que l'efficacité catalytique du CYP1A2 est significativement supérieure chez l'homme par rapport au rat [228], et qu'à la différence de l'humain où

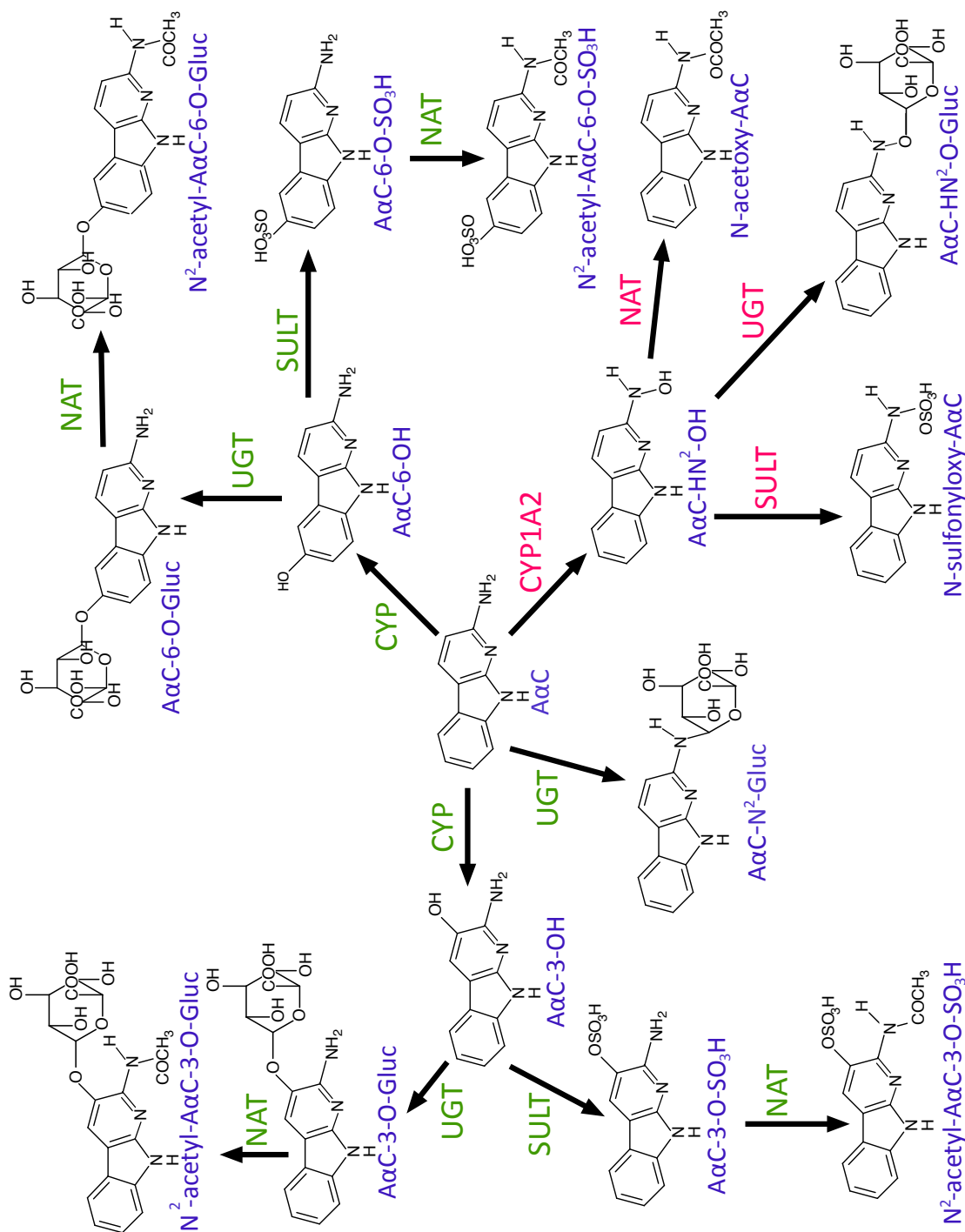


FIGURE 2.3 – Métabolisme d'AαC

AαC (9H-pyrido[2,3-b]indol-2-amine); AαC-N<sup>2</sup>-Gluc (3,4,5-trihydroxy-6-(9H-pyrido[2,3-b]indol-2-ylamino)oxane-2-carboxylic acid); AαC-3-OH (2-amino-9H-pyrido[2,3-b]indol-3-ol); AαC-3-O-SO<sub>3</sub>H (2-amino-9H-pyrido[2,3-b]indol-3-yloxidanesulfonic acid); AαC-3-O-Gluc (6-(2-amino-9H-pyrido[2,3-b]indol-3-yloxy)-3,4,5-trihydroxyoxane-2-carboxylic acid); N<sup>2</sup>-acetyl-AαC-3-O-Gluc (6-(2-acetamido-9H-pyrido[2,3-b]indol-3-yloxy)-3,4,5-trihydroxyoxane-2-carboxylic acid); N<sup>2</sup>-acetyl-AαC-6-O-Gluc (6-(2-amino-9H-pyrido[2,3-b]indol-6-yloxy)-3,4,5-trihydroxyoxane-2-carboxylic acid); AαC-6-OH (2-amino-9H-pyrido[2,3-b]indol-6-ol); AαC-6-O-Gluc (6-(2-amino-9H-pyrido[2,3-b]indol-6-yloxy)-3,4,5-trihydroxyoxane-2-carboxylic acid); AαC-6-O-SO<sub>3</sub>H (2-amino-9H-pyrido[2,3-b]indol-6-yloxidanesulfonic acid); N<sup>2</sup>-acetyl-AαC-6-O-SO<sub>3</sub>H (2-acetamido-9H-pyrido[2,3-b]indol-6-yloxidanesulfonic acid); AαC-9H<sup>2</sup>-O-Gluc (N-(9H-pyrido[2,3-b]indol-2-ylhydroxylamine)); AαC-9H<sup>2</sup>-O-Gluc (3,4,5-trihydroxy-6-[(9H-pyrido[2,3-b]indol-2-ylamino)oxy]oxane-2-carboxylic acid); AαC-9H<sup>2</sup>-O-Gluc ((9H-pyrido[2,3-b]indol-2-ylamino)oxy)sulfonic acid); N-sulfonyloxy-AαC (9H-pyrido[2,3-b]indol-2-ylamino)oxy)sulfonic acid); N-acetoxy-AαC (9H-pyrido[2,3-b]indol-2-ylamino)oxy)sulfonic acid)

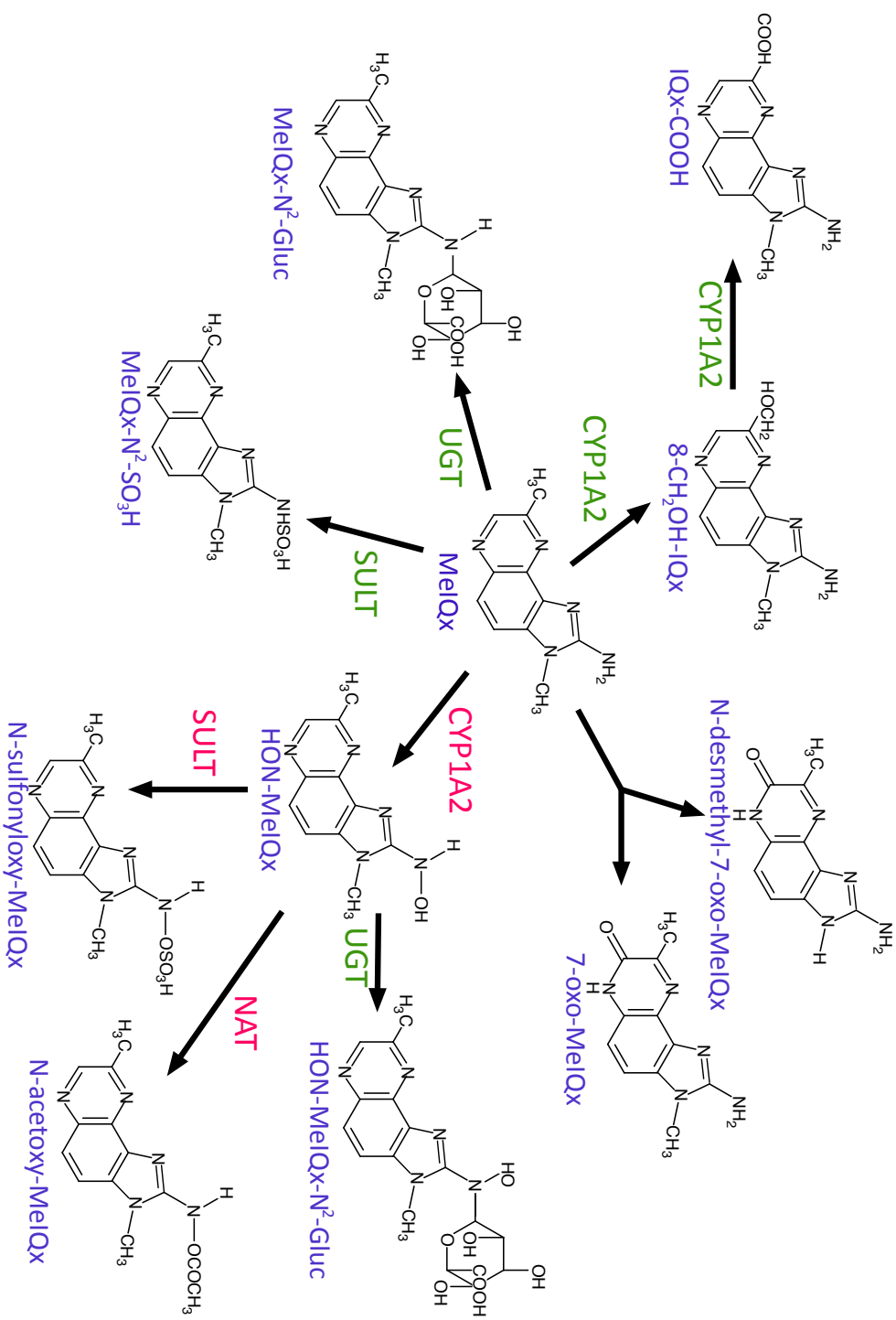


FIGURE 2.4 – Métabolisme du MeIQx

MeIQx (3,8-diméthyl-3H-imidazo[4,5-f]quinoxalin-2-amine) ; 8-CH<sub>2</sub>OH-IQx (2-amino-3-méthyl-1,3H-imidazo[4,5-f]quinoxalin-7-ol) ; MeIQx-N<sup>2</sup>-Gluc (6-(3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-7-yl)iméthanol) ; IQx-COOH (2-amino-3-méthyl-3H-imidazo[4,5-f]quinoxaline-8-carboxylique acid) ; MeIQx-N<sup>2</sup>-SO<sub>3</sub>H (6-(3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-7-yl)sulfonique acid) ; 7-oxo-MeIQx (2-amino-3,8-diméthyl-1,3H,6H,7H-imidazo[4,5-f]quinoxalin-7-one) ; N-desméthyl-7-oxo-MeIQx (2-amino-8-méthyl-1,3H,6H,7H-imidazo[4,5-f]quinoxalin-7-one) ; MeIQx-N<sup>2</sup>-SO<sub>3</sub>H (N-3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-2-yl)sulfonique acid) ; HON-MeIQx (N-3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-2-yl)hydroyxamine) ; HON-MeIQx-N<sup>2</sup>-Gluc (6-(N-3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-2-yl)amino) ; MeIQx (3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-2-yl)amino acetate) ; N-sulfonyloxy-MeIQx ((3,8-diméthyl-1,3H-imidazo[4,5-f]quinoxalin-2-yl)amino)oxysulfonique acid)

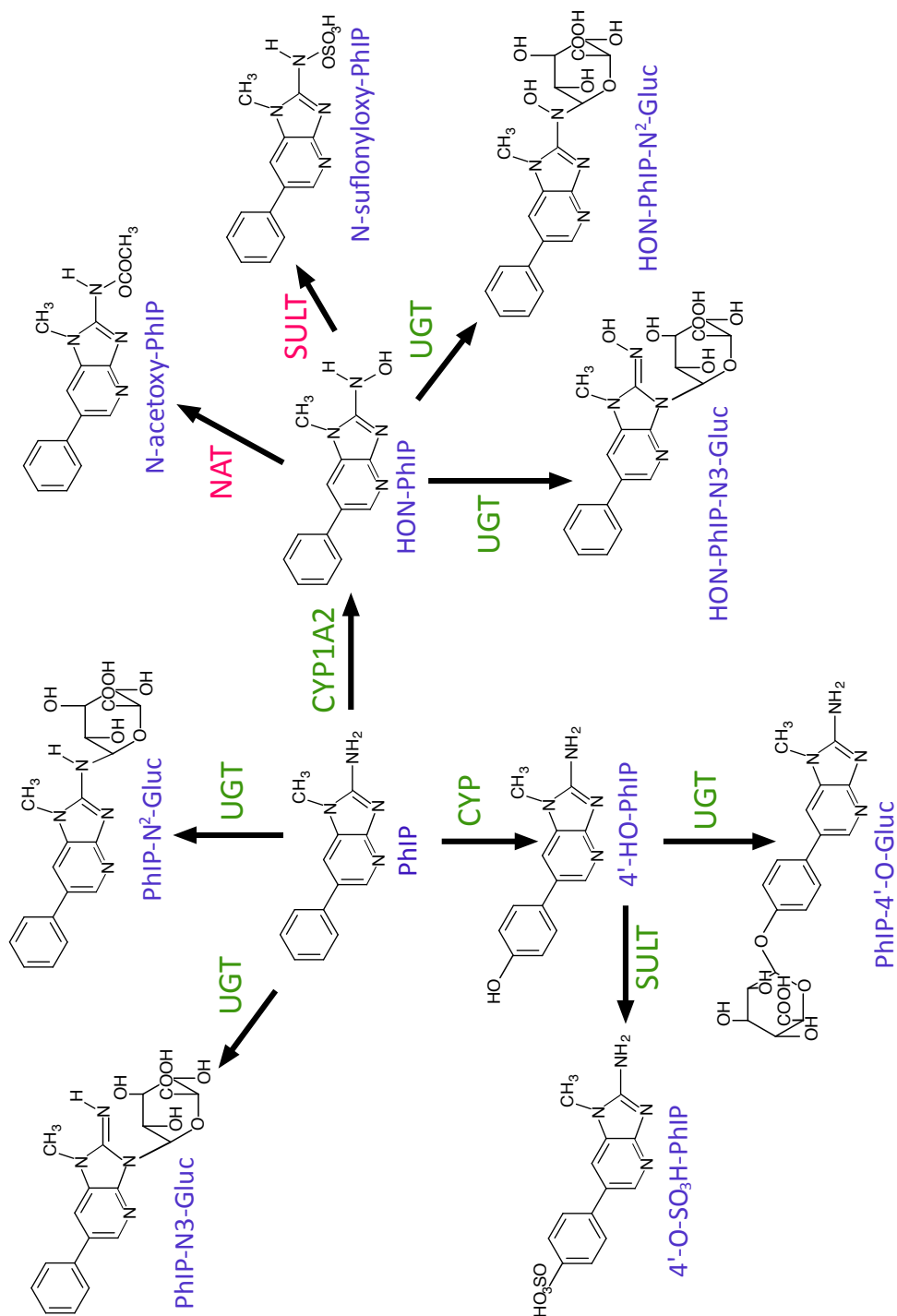


FIGURE 2.5 – Métabolisme du PhIP

PhIP (1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-2-amine); 4'-HO-PhIP (4'-2-amino-1-méthyl-1H-imidazo[4,5-b]pyridin-6-ylphénol); PhIP-4'-O-Gluc (6-(4'-2-amino-1-méthyl-1H-imidazo[4,5-b]pyridin-6-ylphénoxy)-3,4,5-trihydroxyoxane-2-carboxylic acid); 4'-O-SO<sub>3</sub>H-PhIP ((4'-2-amino-1-méthyl-1H-imidazo[4,5-b]pyridin-6-ylphényl)oxidanesulfonic acid); PhIP-N<sup>2</sup>-Gluc (3,4,5-trihydroxy-6-(1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-3-yl)oxane-2-carboxylic acid); PhIP-N<sup>2</sup>-Gluc (3,4,5-trihydroxy-6-(1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-2-ylamino)oxane-2-carboxylic acid); HON-PhIP-N<sup>2</sup>-Gluc (3,4,5-trihydroxy-6-(N-1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-3-yl)oxane-2-carboxylic acid); HON-PhIP-N<sup>3</sup>-Gluc ((1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-2-ylamino)oxy)sulfonic acid); N-acetoxy-PhIP (1-méthyl-6-phényl-1H-imidazo[4,5-b]pyridin-2-ylamino acetate)



le CYP1A2 catalyse la formation privilégiée de produits N-oxydés (bioactivation), le rat assure l'oxydation des fonctions amines (bioactivation) et des cycles aromatiques (détoxification) à des niveaux comparables [132, 228]. De plus, certaines voies métaboliques peuvent être majeures chez l'animal alors que celles-ci sont négligeables, voire absentes chez certains individus [228, 233]. L'ensemble de ces variations interspèces est une source potentielle de variations importantes de la génotoxicité entre l'homme et les modèles animaux rendant ces derniers difficilement extrapolables lors de l'étude en toxicité humaine.

L'autre élément de variabilité important concerne la variabilité interindividuelle importante chez l'homme qui est susceptible d'influencer l'importance des voies de bioactivation par rapport à celles de détoxification, et par conséquent agir sur la génotoxicité des AHA. Cette variabilité opère notamment de manière importante sur les niveaux d'expressions hépatiques d'ARNm et protéiques du CYP1A2 qui présentent une variation respective d'un facteur de 15 [57, 97, 200] pour les ARNm et de 60 [20, 228] pour les protéines. Ces variations ont été complétées par l'observation d'une variation phénotypique d'un facteur 70 lors de la prise en charge de la caféine par le CYP1A2 [31, 134, 238]. De plus, il a été mis en évidence une variation interindividuelle importante dans la formation hépatique du IQx-COOH qui varie de 16 à 41% de la quantité initiale de MeIQx [132] et illustrant une nouvelle fois les variations interindividuelles dépendantes du CYP1A2. L'ensemble des différences interindividuelles pourrait expliquer la variabilité de la cancérogenèse chez l'homme. En effet, une étude menée par Sinah et al. a suggéré l'importance des variations interindividuelles de l'activité du CYP1A2 sur les cancers associés à l'exposition des AHA [203]. En effet, une augmentation de l'activité du CYP1A2 est inversement corrélée à la quantité de MeIQx non hydroxylée. Diverses études ont essayé de proposer des pistes d'explications à cette variation interindividuelle du CYP1A2, qui serait potentiellement liée à des facteurs génétiques et environnementaux [130]. Il a été ainsi démontré en 2004 que des polymorphismes génétiques du CYP1A2 peuvent modifier l'expression et l'activité du CYP1A2, ce qui impacterait la bioactivation des AHA. Néanmoins, l'identification de génotypes par l'étude de SNPs (single nucleotide polymorphisms) affectant les variations d'expression hépatiques interindividuelle de CYP1A2 a échoué supposant l'intervention d'autres facteurs subcellulaires tels que des facteurs de régulations [103].

En conclusion, les variations interindividuelles affectant la voie métabolique catalysée majoritairement par les CYP1A2 peuvent être un facteur capital de la susceptibilité individuelle aux AHA [107, 227]. A l'heure actuelle, ces variations sont principalement expliquées par des facteurs génétiques et environnementaux qui agiraient sur la bioactivation des AHA et indirectement la formation d'adduits à l'ADN.

### 2.3.3 Le métabolisme des AHA par les enzymes de Phase II

Les enzymes de phase II participant au métabolisme des AHA et biotransformant soit directement le composé néfaste, ou soit un intermédiaire hydroxylé, regroupent les UGT, les glutathiones S-transférases (GST), les SULT, et les NAT. Ces enzymes ont presque toutes un rôle bivalent en étant impliquées à la fois dans la bioactivation et à la fois dans la détoxification.

Le métabolisme des substances de type N-OH-AHA, comme le HON-A $\alpha$ C, ou le HON-MeIQx, ou encore le HON-PhIP peuvent être N-acétylés par la famille enzymatique des NAT (Figure 2.2). Cette classe enzymatique regroupant les NAT1 exprimés principalement dans les tissus extrahépatiques et les NAT2 localisés principalement dans le foie [87, 82], forment un intermédiaire réactif de type N-acetoxy-AHA réactif qui se liera à l'ADN après clivage hétéolytique (Figure 2.2) [86, 112, 199, 230]. Les NAT ont donc un rôle majeur dans les génotoxicités des AHA. Néanmoins, le rôle des NAT, et notamment de NAT2 dans la susceptibilité au cancer reste encore incertain [84, 82]. En effet, de nombreuses études menées pour déterminer ce rôle sont contradictoires. La recherche de corrélation entre une forte activité de NAT2 et le risque de cancer colorectal est un exemple qui a divisé la communauté scientifique. En effet, il a été montré qu'il y avait une association entre ces deux facteurs dans des études [84, 131, 135], alors que celle-ci a été contredite par d'autres scientifiques [16, 84, 220]. Une hypothèse d'explication de ces résultats pourrait probablement provenir de l'intervention de facteurs génétiques comme les polymorphismes. Il a été identifié à ce jour plus de 25 polymorphismes pour les deux isoenzymes NAT affectant leur efficacité catalytique dans la prise en charge des AHA [23, 84, 83]. Parmi ces polymorphismes, on distingue trois phénotypes différents les *slow-acetylators*, les *intermediate-acetylators* et les *rapid-acetylators* [84, 82]. Certains phénotypes comme les *rapid-acetylators* ont démontré leur importance dans le risque du cancer colorectal [84]. D'autres facteurs, tels que la dépendance entre le génotype et le phénotype, les variations interindividuelles dans le métabolisme, la réparation des adduits à l'ADN, et les susceptibilités différentes aux AHA pourraient être d'autres facteurs expliquant les différences dans les résultats des études [87]. Par ailleurs, les NAT sont aussi impliqués dans la détoxification des AHA au moyen de N,O acetylation. A titre d'exemple, A $\alpha$ C, le MeIQx et le PhIP peuvent être métabolisés respectivement en N-acetoxy-A $\alpha$ C, N-acetoxy-MeIQx et N-acetoxy-PhIP, qui seront des produits de détoxification (figures 2.3, 2.4 et 2.5).

La prise en charge des composés N-Hydroxylés peut aussi être effectuée par la superclasse enzymatique des SULT qui contient deux sous-familles : les SULT1 et les SULT2 [22, 71, 162]. Néanmoins, seules les isoenzymes SULT1A1 et SULT1A2 semblent intervenir dans le métabolisme des AHA que ce soit au niveau de la sulfatation des métabolites N-OH-AHA, ou de la sulfatation directe de fonctions amines et alcools d'autres intermédiaires métaboliques [44, 159, 170, 254]. Dans le premier cas, la sulfatation du N-OH-AHA aboutit à la bioactivation de ce composé qui produira un composé chimique à l'origine d'adduits à l'ADN [199]. HON-A $\alpha$ C, HON-MeIQx et

HON-PhIP sont donc bioactivés en N-Sulfonyloxy-A $\alpha$ C (Figure 2.3), N-Sulfonyloxy-MeIQx (Figure 2.4), N-Sulfonyloxy-PhIP (Figure 2.5). Néanmoins, les SULT peuvent aussi être des acteurs de la détoxification grâce à leur action sur des composés préalablement hydroxylés sur des carbones comme les 3-OH-A $\alpha$ C, 6-OH-A $\alpha$ C (Figure 2.3) [19] et 4'-OH-PhIP (Figure 2.5) [129]. Une dernière action de détoxification a été montrée par sa capacité à sulfater dans certains cas directement les fonctions amines primaires (NH<sub>2</sub>) comme c'est le cas pour le MeIQx (Figure 2.4) [229]. A ce jour, plusieurs polymorphismes ont été identifiés pour les SULT, mais il semble que ces facteurs génétiques ne soient pas associés au développement privilégié de cancer [154, 206].

La super famille des UGT intervient principalement comme un agent de détoxification dans le métabolisme des AHA. Cette famille est subdivisée en 3 sous-familles : les UGT1A1, les UGT2A et les UGT2B [113]. Néanmoins, les UGT1A sont la classe enzymatique majoritaire dans les métabolismes des AHA [132, 146, 145], et peut transférer un groupement glucuronide soit sur une fonction amine exo ou endocyclique ou soit sur une fonction alcool exocyclique [19, 133, 132]. Les composés résultants sont ensuite éliminés dans les urines et la bile [128, 240, 76]. Ainsi, les UGT catalysent la transformation d'A $\alpha$ C en A $\alpha$ C-N<sup>2</sup>-Gluc, d'A $\alpha$ C-3-OH en A $\alpha$ C-3-O-Gluc, d'A $\alpha$ C-6-OH en A $\alpha$ C-6-O-Gluc, et de A $\alpha$ C-HN<sup>2</sup>-OH en A $\alpha$ C-HN<sup>2</sup>-O-Gluc (Figure 2.3) [19]. Elles bio-transforment aussi le MeIQx en MeIQx-N<sup>2</sup>-Gluc et le HON-MeIQx en HON-MeIQx-N<sup>2</sup>-Gluc (Figure 2.4) [132]. Enfin, les UGT agissent dans la transformation du PhIP en PhIP-N<sup>3</sup>-Gluc, et en PhIP-N<sup>2</sup>-Gluc, et du HON-PhIP en HON-PhIP-N<sup>2</sup>-Gluc et en HON-PhIP-N<sup>3</sup>-Gluc (Figure 2.5) [133]. Les UGT sont donc des enzymes fortement impliquées dans la détoxification et peut être l'enzyme majeure dans la détoxification comme c'est le cas pour MeIQx et PhIP, où les intermédiaires glucuronés peuvent représenter jusqu'à 12% de la dose pour le MeIQx [132] et jusqu'à 66% de la dose pour le PhIP [133]. Néanmoins, la glucuronidation a récemment été identifiée comme un vecteur de bioactivation chez A $\alpha$ C. En effet, l'intermédiaire A $\alpha$ C-HN<sup>2</sup>-OH, tout comme les composés N-acétoxy et N-sulfonyloxy, peut aller se fixer à l'ADN pour former un adduit [34]. De plus, des polymorphismes génétiques modifiant l'efficacité catalytique des UGT ont montré l'importance des UGT et notamment d'UGT1A1 dans la détoxification hépatique des AHA, et par conséquent de son implication indirecte dans la cancérogenèse en cas d'altération de sa fonction catalytique [75, 74].

La dernière famille participant au métabolisme des AHA lors de la phase II est la famille des GSTs. Cette famille médie la détoxification par la prise en charge du composé N-acétoxy-AHA afin de reformer le composé de départ instaurant une potentielle boucle dans le métabolisme des AHA [139]. Cette reconversion a été montrée pour le N-acétoxy-PhIP qui est retransformé en PhIP par les GST [139]. De plus, il a été montré l'importance de cette voie dans la détoxification en participant à réduire de manière significative la formation d'adduits à l'ADN [139]. Néanmoins, cette voie n'a été mise en évidence qu'uniquement sur les PhIP et l'intervention des GST dans le métabolisme des AHA n'a été démontrée que sur une seule autre AHA qui est le Tr-P-2 [195].

## 2.4 Enjeux de modélisation : un challenge informatique et biologique

Comme précédemment exposé, les AHA sont une famille complexe de xénobiotiques soulevant de nombreuses questions informatiques généralisables aux xénobiotiques. En effet, cette famille possède de nombreuses variabilités métaboliques rendant très difficile l'usage d'approches biologiques pour y répondre et laissant ainsi une part importante à l'informatique permettant à la fois de soulever de nouveaux problèmes informatiques et à la fois de générer de nouvelles connaissances biologiques. Aussi, cette famille a été suffisamment étudiée pour permettre la calibration et la validation de méthodes informatiques implémentées, mais constitue aussi un challenge en raison des nombreux inconnus inhérents aux métabolismes encore non caractérisés pour 90% des composés de cette famille et de leurs variabilités métaboliques associées. Les AHA sont donc une famille d'étude idéale dont les enjeux majeurs traités dans cette thèse sont exposés ci-après.

Un premier enjeu réside dans l'étude de la spécificité réactionnelle particulière de certains xénobiotiques. En effet, alors qu'il était considéré qu'une enzyme était spécifique d'un substrat et d'une réaction, il a été montré pour les AHA, ainsi que d'autres familles de xénobiotiques, qu'une même enzyme peut catalyser à partir du même substrat deux réactions différentes aboutissant à deux produits différents. De plus, ce phénomène appelé régiosélectivité, qui est une préférence de catalyse d'une réaction par rapport à une autre, peut changer en fonction de la concentration du substrat. Il est donc important lors d'une démarche de modélisation de soulever ce genre de problématique et de modéliser cette régiosélectivité afin d'en étudier son impact. Les approches de modélisation fines permettant l'évaluation des conséquences de tels mécanismes peu abordés dans la littérature d'un point de vue de la biologie des systèmes ont fait l'objet d'une étude approfondie dans cette thèse. Ainsi, les chapitres 5 et 6 exposeront plus précisément l'étendue de ces mécanismes présents dans le métabolisme de divers xénobiotiques et évaluera leur incidence sur la génotoxicité en prenant l'exemple du MeIQx.

Un second point concerne l'importante variabilité des paramètres de nombreuses enzymes impliquées dans le métabolisme des AHA en raison de polymorphismes génétiques et d'expression. En d'autres termes, il y a une importante variation des concentrations enzymatiques, des  $K_m$ , et des  $k_{cat}$ . Ainsi, l'utilisation de modèles paramétrisés basés sur différentes sources de la littérature peut être difficile et fausser le modèle. Il peut aussi être important d'étudier l'impact de cette diversité paramétrique pour chaque enzyme dans le cadre d'une population afin d'évaluer leurs effets sur la toxicologie par la production de métabolites réactifs.

De plus, cette variabilité s'étend aux modèles animaux qui diffèrent de l'humain. Alors que certaines approches ont été développées pour l'extrapolation des para-

mètres animaux vers l'humain [219], il peut être difficile d'utiliser ces techniques d'extrapolation en raison de la variabilité paramétrique et métabolique inter-espèce.

Enfin, les données chez l'homme reposent principalement sur trois AHA qui ont fortement attiré la communauté scientifique au détriment de 27 autres AHA. Il convient donc d'implémenter des méthodes informatiques pour prédire le métabolisme des autres AHA, de prédire leurs métabolites réactifs, et d'évaluer leurs impacts sur le cancer. Cette dernière problématique exposée a fait l'objet du développement d'une méthode générique permettant la prédiction du métabolisme des xénobiotiques et l'identification des dérivés métaboliques génotoxiques (chapitre 3). Cette dernière a été appliquée dans le chapitre 4 sur l'exemple de la famille des AHA.

Un des enjeux consiste à développer des méthodes informatiques pour élucider les problèmes ci-dessus exposés afin de pouvoir prédire le métabolisme des AHA, et étudier leur impact en santé humaine.

## 2.5 Résumé et principaux apports

L'objectif de cette thèse est d'exploiter les forces de différentes méthodes de modélisation synthétisées dans la figure 2.6 afin de proposer un travail exploratoire multi-échelle préliminaire dans le défrichage des problématiques précédemment exposées. La première partie du travail consiste en la construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions, qui sera appliquée à la famille entière des AHA. Pour répondre à ce sujet, les méthodes de l'échelle moléculaire de prédiction de réactions métaboliques et celles de prédictions des réactivités moléculaires sont combinées. Cette méthodologie permet d'annoter le graphe prédit avec des informations de génotoxicité, mais aussi de réduire significativement la combinatoire inhérente aux méthodes de prédiction de réactions métaboliques. La deuxième partie du travail s'attache à modéliser les effets de compétitions enzymatiques affectant les préférences de catalyse réactionnelle par l'exploitation de données expérimentales limitées. Cette étude a abouti au développement d'un modèle modélisant le changement dynamique de régiosélectivité en fonction de la concentration d'une substance chimique basée sur l'exploitation de la modélisation cinétique et de recherche de paramètres par fitting. Ce dernier travail a été appliqué au MeIQx qui possède cette problématique.

	Modèles	Avantages	Inconvénients
Echelle de l'organisme	PBPK	+Modélisation à l'échelle de l'organisme possible +Quantitatifs +Modélisation à l'échelle d'une population possible +Dynamiques	-Très paramétrisés  -Complexes
	Modèles topologiques	+Exploration de grands graphes +Utilisation uniquement de propriétés topologiques +Non paramétrés	-Statiques Qualitatifs
Echelle cellulaire	FBA	+Paramètres facultatifs +Analyse de grands réseaux métaboliques +Quantitatifs +Multi-compartiments possibles	-Statiques -Faible degré de précision
	Cinétiques	+Fort degré de précision +Dynamiques +Quantitatifs +Adaptés pour des mécanismes complexes	-Paramétrisés -Limités à de petits réseaux
Echelle moléculaire	SOMs	+Prédiction relativement fiable des SOMs +Diversité des approches +Prédiction à partir de la structure 2D uniquement +Adaptés à l'investigation d'une substance précise	-Largement limité au CYP  -Qualitatifs
	Réactions	+Prédiction de toutes les réactions qu'une molécule peut subir +Possible utilisation pour prédire un métabolisme dans son intégralité +Prédiction à partir de la structure 2D uniquement +Adaptés à l'investigation d'une substance précise	-Forte combinatoire  -Qualitatifs

FIGURE 2.6 – Synthèse des approches de modélisation aux différentes échelles d'étude



# Construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions

Le travail présenté dans les deux chapitres suivants fait l'objet d'une soumission imminente dans Journal of Chemoinformatics. L'objectif de ce chapitre est de présenter une méthode globale pour la construction *in silico* de graphes du métabolisme de xénobiotiques, ainsi que leur génotoxicité par la formation d'adduits à l'ADN. La méthode de prédiction des graphes métaboliques développée est basée sur la combinaison de diverses approches moléculaires permettant des prédictions combinant des sensibilités et précisions élevées. La méthode proposée reposera sur l'exploitation combinée de la prédiction de réactions biochimiques (voir chapitre 1 section 1.3.3.2) pour construire les graphes, des méthodes de prédictions des SOM (voir chapitre 1 section 1.3.3.1) pour augmenter la précision et des approches de prédiction de la toxicité (voir chapitre 1 section 1.3.4).

---

<b>3.1</b>	<b>Introduction</b> . . . . .	<b>58</b>
<b>3.2</b>	<b>Description des méthodes implémentées au sein d'un pipeline</b> . .	<b>60</b>
3.2.1	Prédiction des réactions métaboliques . . . . .	60
3.2.2	Identification de la sélectivité enzymatique et prédiction des SOM . . . . .	62
3.2.3	Prédiction de la réactivité vis-à-vis de l'ADN . . . . .	63
3.2.4	Implémentation de filtres pour limiter la combinatoire . . . . .	63
<b>3.3</b>	<b>Procédure de calibration</b> . . . . .	<b>64</b>
3.3.1	Base de données biologiques pour l'optimisation . . . . .	64
3.3.2	Optimisation des méthodes de prédictions des SOM et des SOR . . . . .	64
<b>3.4</b>	<b>Validation et analyses statistiques</b> . . . . .	<b>67</b>
<b>3.5</b>	<b>Discussion et conclusion</b> . . . . .	<b>72</b>



### 3.1 Introduction

La prédiction du métabolisme de xénobiotiques chez l'homme est capitale pour identifier et caractériser leurs voies métaboliques, ainsi qu'évaluer leurs génotoxicités, mais demeure un défi majeur suscitant un fort intérêt pour les domaines de la pharmaceutique, la cosmétique, l'alimentaire et l'environnement. La plupart des approches existantes se concentrent soit sur la prédiction de réactions métaboliques, soit sur la prédiction des sites de métabolisme (SOM).

Comme illustré dans le chapitre 1 section 1.3.3.2, la prédiction des réactions métaboliques est principalement basée sur des systèmes experts, où la prédiction des métabolites s'effectue sur l'exploitation d'un dictionnaire construit à partir de données expérimentales et contenant des règles de biotransformations (transformation d'un motif moléculaire "substrat" A en en motif moléculaire "produit" B). Ainsi, cette approche permet la prédiction de métabolites en analysant une molécule chimique, et remplaçant tous les motifs moléculaires "substrat" correspondant à une entrée dans le dictionnaire par leurs motifs "produits" correspondants. Une autre approche basée sur le Data-Mining et l'analyse statistique a été développée plus récemment [5]. Cette dernière implémentée au sein de MetaPrint2D-React utilise des empreintes moléculaires circulaires de chaque atome d'une molécule "substrat". Puis, ces empreintes moléculaires sont comparées à celles de substrats contenues dans une base de données qui stocke, à la différence des systèmes experts, un ensemble de transformations substrats/produits. En fonction de la similarité des empreintes moléculaires, les métabolites sont ensuite prédits en appliquant les règles de biotransformations correspondantes.

Malgré les très bonnes performances de ces méthodes pour prédire les réactions associées à un xénobiotique, un des principaux inconvénients de celles-ci est l'explosion de la combinatoire associée aux prédictions métaboliques et engendrant par conséquent un nombre considérable de faux positifs [29]. Dans l'optique d'améliorer la précision de ces méthodes tout en limitant la perte en sensibilité, des seuils arbitraires permettant de sélectionner un nombre défini de métabolites et des méthodes probabilistes ont été implémentés. Malgré leurs utilisations, la précision de ces méthodes prédictives reste faible lorsqu'une sensibilité élevée est nécessaire [178]. De plus, l'amélioration de cette précision est très souvent associée à une diminution significative de la sensibilité. Aussi, bien que certaines de ces solutions informatiques permettent de prédire la toxicité, elles sont dans l'incapacité à prédire la réactivité des métabolites vis-à-vis de l'ADN, ce qui est une faiblesse majeure dans la prédiction du métabolisme de certains xénobiotiques.

A l'instar des méthodes de prédiction des réactions métaboliques des xénobiotiques, des méthodes de prédiction des sites de métabolisme (SOM) ont aussi été implémentées dans le cadre du développement de médicaments afin d'optimiser les propriétés métaboliques d'une molécule chimique. Ainsi, de nombreuses approches informatiques, ainsi que de très nombreuses méthodes visant à prédire les SOM de xénobiotiques ont émergé au cours de ces dernières années (voir chapitre 1 section 1.3.3.1 et la revue [115]). En raison de la diversité des approches développées, certaines solutions hybrides prédisant les SOM pour les CYP comme IMPACTS [35], SMART-CYP [192] et Metasite [49] ont été proposées afin de combiner les forces des différentes méthodologies. Par ailleurs, quelques méthodes permettant aussi la prédiction des SOM pour les UGT ont été conçues. Les approches proposées par Xenosite [260, 51] et Way2Drug [189] ont été développées à cet effet.

Dans l'optique d'optimiser les prédictions du métabolisme des xénobiotiques, quelques méthodes visant à combiner les méthodes de prédiction de réactions métaboliques avec d'autres de prédiction de SOM ont été implémentées. Tarcsay et al. [217] ont ainsi combiné l'utilisation de MetabolExpert [52] avec le logiciel de docking GLIDE [70, 80]. Ils ont généré dans un premier temps tous les métabolites de substrats connus en utilisant MetabolExpert et ensuite amélioré la précision à l'aide du docking réduisant ainsi le nombre de faux positifs. Une approche similaire combinant l'utilisation de Metaprint2D pour prédire les réactions métaboliques et le logiciel de docking AutoDock [155] a été effectuée par Sousa et al. pour prédire le métabolisme des flavonoïdes pour CYP1A2 et CYP2C9 [205]). Au même titre que Tarcsay et al. et Sousa et al., l'étude de Piechota [178] consistait à combiner la prédiction de métabolites par Metaprint2D-React [5] avec la prédiction de SOM par SMARTCYP [192]. De telles approches hybrides visent à réduire le nombre de métabolites sur-approximés (faux positifs), mais elles demeurent limitées lors de la prédiction de réseaux métaboliques à plus large échelle. En effet, alors que les approches de docking peuvent être coûteuses en temps, en ressources informatiques et nécessitent la connaissance des isoformes impliqués dans le métabolisme des xénobiotiques, celles testées par Piechota et al. parviennent à améliorer la précision au détriment de la sensibilité.

Ce travail présente le développement d'une méthode originale et intégrative combinant les avantages de prédire le métabolisme des xénobiotiques avec une précision importante tout en discriminant les métabolites toxiques des non toxiques. L'approche proposée s'attache à combler les principales limites résultant de la prédiction des métabolites en limitant le nombre de métabolites sur-prédits (faux positifs), en améliorant l'identification des SOM par l'emploi combiné de plusieurs méthodes de prédiction tout en considérant la spécificité enzymatique pour les CYP et en identifiant les molécules responsables de la formation d'adduits à l'ADN. Cette méthode consiste à prédire dans un premier temps les voies métaboliques des xénobiotiques en octroyant une utilisation itérative de Metaprint2D-React, qui était initialement prévue pour prédire uniquement les réactions métaboliques que peut subir une molécule. Ce processus est réalisé par la transcription des transformations métaboliques en produits au format

SMILES permettant ainsi d'automatiser la prédiction du métabolisme par l'utilisation itérative de MetaPrint2D-React. Cette étape assure la production de cartes métaboliques exhaustives des xénobiotiques. Ensuite, afin de résoudre les problématiques de sur-prédiction afin d'augmenter la précision tout en conservant une sensibilité élevée, des méthodes de filtrage basées sur la prédiction des SOM ont été implémentées en combinant 6 méthodologies différentes : Metaprint2D-React [5], WhichCyp [187], Way2Drug-SOMP [189], Xenosite Metabolism 1.0 [260], Xenosite UGT 2.0 [51] et Xenosite Reactivity 2.0 [94, 93]. Enfin, afin d'identifier les métabolites réactifs vis-à-vis de l'ADN, Xenosite Reactivity 2.0 [94, 93] a été implémenté à cet effet. Cette approche intégrative générique permet la prédiction des voies métaboliques de xénobiotiques et l'identification des métabolites réactifs vis-à-vis de l'ADN.

## 3.2 Description des méthodes implémentées au sein d'un pipeline pour la construction *de novo* de graphes métaboliques

La construction *de novo* de graphes métaboliques des xénobiotiques et la prédiction des dérivés métaboliques conduisant à la formation d'adduits à l'ADN reposent sur la combinaison de six méthodes différentes. Ces six méthodologies combinées et organisées au sein d'un pipeline à des fins de prédictions sont : Metaprint2D-React [5], WhichCyp [187], Way2Drug SOMP [189], Xenosite Metabolism 1.0 [260], Xenosite UGT 2.0 [51] et Xenosite Reactivity 2.0 [94, 93]. Comme illustré dans la figure 3.1, ce pipeline peut être décomposé en deux parties principales incluant la prédiction des cartes métaboliques par utilisation itérative de Metaprint2D-React, puis leurs filtrations en prenant notamment en compte les différentes isoformes des CYP intervenant dans le métabolisme et la réactivité vis-à-vis de l'ADN. Au final, pour un xénobiotique donné en entrée, le pipeline fournit en sortie un graphe réduit des voies métaboliques prédites, ainsi que l'identification des métabolites réactifs vis-à-vis de l'ADN.

### 3.2.1 Prédiction des réactions métaboliques

- **Metaprint2D-React** [5] Metaprint2D-React est une méthode de prédiction des réactions métaboliques d'une substance chimique. La prédiction des réactions est possible avec trois configurations de flexibilité des empreintes moléculaires (Loose, Default, et Strict) et pour trois types d'organisme : l'homme, le chien et le rat. Metaprint2D-React a été amélioré dans cette méthode afin d'être implémenté de façon itérative, où chaque molécule y est entrée au format SMILES (Simplified Molecular Input Line Entry) avec la flexibilité des empreintes moléculaires paramétrée sur "default". Puis, les métabolites générés sont utilisés à leur tour comme nouvelle entrée de MetaPrint2D-React. Cette procédure est itérée de façon à obtenir deux générations de métabolites (profondeur du graphe prédit = 2), en d'autres termes la molécule initialement introduite dans MetaPrint2D-React peut subir 2 transformations consécutives. De plus, MetaPrint2D-

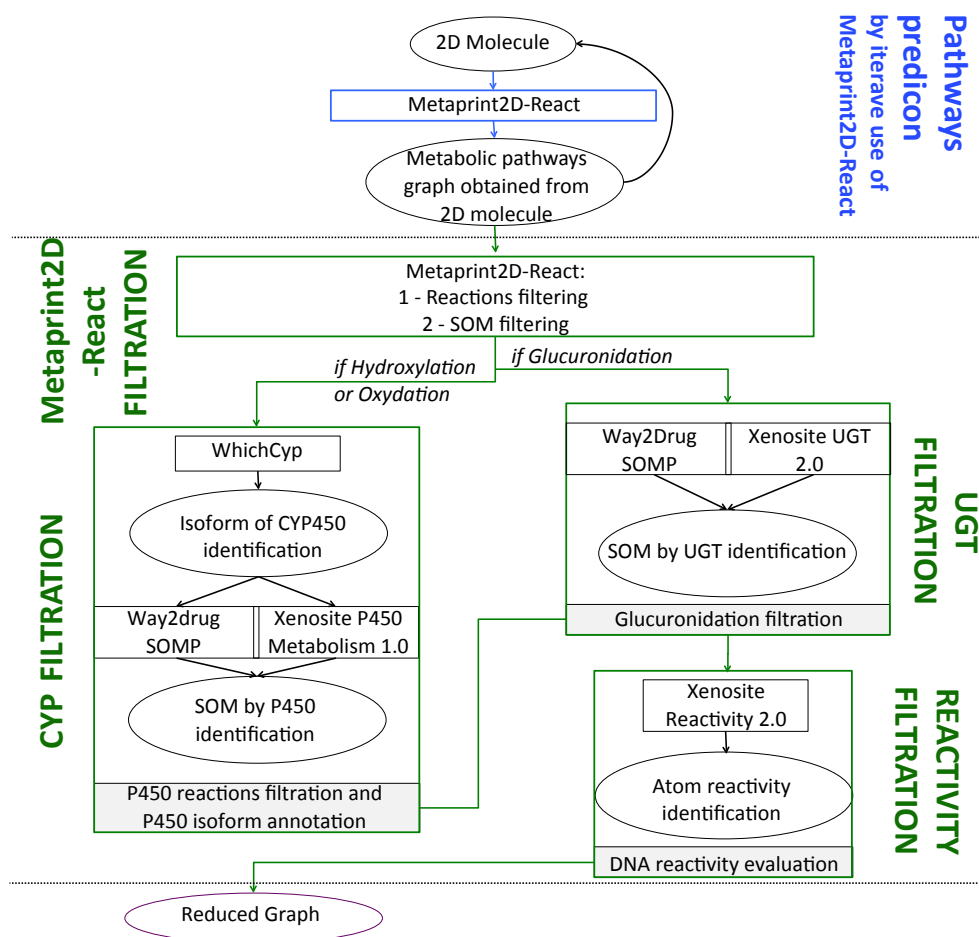


FIGURE 3.1 – Pipeline pour la prédiction des voies métaboliques des xénobiotiques et l'identification des métabolites réactifs à l'ADN

Ce pipeline consiste à prédire dans un premier temps tous les ensembles de réactions pour une molécule donnée, puis réduire les cartes métaboliques prédites par l'exploitation d'ensemble de méthodes calibrées prédisant les SOM et enfin identifier les molécules et atomes réactifs avec l'ADN. La phase de filtration inclut trois étapes distinctes : un premier filtre basé sur MetaPrint2D-React supprimant les réactions improbables et celles se produisant sur un atome ayant un score NOR en dessous du seuil NOR optimisé ; un second filtre est basé sur Way2Drug SOMP, Xenosite Metabolism 1.0 et Xenosite UGT 2.0 retenant seulement les réactions se produisant sur les atomes prédits comme des SOM par au moins l'une des deux méthodes ; un troisième filtre basé sur Xenosite Reactivity 2.0 supprimant toutes les réactions de seconde génération augmentant la réactivité (en considérant que la phase II consiste à diminuer la réactivité) et se produisant sur un groupe d'atome différent de celui de la molécule précédemment modifiée lors des réactions de premières générations. La sortie de ce pipeline est un graphe réduit contenant tous les types de réactions implémentés dans la base de données de Metaprint2d-React.

React calcule un score statistique normalisé (score NOR) évaluant la fréquence de l’empreinte moléculaire d’un SOM dans la base de données de MetaPrint2D-React. Ainsi, chaque métabolite prédit peut être associé à un score NOR variant de 0 à 1 qui correspond par extrapolation à sa probabilité de formation.

Le graphe prédit par MetaPrint2D-React est ensuite filtré en fonction de la prédiction des SOM (WhichCyp, Way2Drug SOMP, Xenosite Metabolism 1.0 et Xenosite UGT 2.0) et de la réactivité des métabolites vis-à-vis de l’ADN (Xenosite Reactivity 2.0).

### 3.2.2 Identification de la sélectivité enzymatique et prédiction des SOM

- **WhichCyp** [187] WhichCyp est une méthode utilisée pour prédire la sélectivité enzymatique des CYP 1A2, 2C9, 2C19, 2D6 et 3A4. Ce dernier prédit quelle isoenzyme est en mesure de lier/métaboliser une molécule en utilisant une simple classification oui/non. Chaque molécule est introduite en entrée de WhichCyp au format SMILES afin de prédire quelle isoforme est en mesure de métaboliser la molécule d’intérêt. L’/les isoforme(s) prédite(s) est/sont utilisée(s) en tant que référence(s) lors de l’étape d’identification des SOM ci-après présentée.
- **Way2Drug SOMP** [189] Way2Drug SOMP est une méthode de prédiction des SOM d’une molécule pour les isoformes des CYP 1A2, 2C9, 2C19, 2D6 et 3A4 et des UGT. Toutes les molécules prédites par MetaPrint2D-React comme étant potentiellement transformées par les CYP et/ou les UGT sont utilisées en entrée de cette méthode au format SMILES. Way2Drug SOMP calcule un score de probabilité variant entre 0 et 1 (une forte probabilité d’être un SOM est caractérisée par un score élevé), qui détermine si un atome du métabolite est un SOM pour l’enzyme associée. Afin de déterminer le score optimal pour l’étape de filtration, les différentes valeurs de scores possibles sont apprises par l’exploitation de la première base de données (cf. 3.3).
- **Xenosite P450 Metabolism 1.0** [260] Xenosite P450 Metabolism 1.0 est une méthode de prédiction des SOM d’une molécule pour les isoformes des CYP 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1 et 3A4. Toutes les molécules prédites par MetaPrint2D-React comme étant potentiellement transformées par les CYP sont utilisées en entrée de cette méthode au format SMILES. Cette méthode permet, en complément de Way2Drug SOMP, l’identification des SOM et la filtration des réactions catalysées par les CYP prédites par MetaPrint2D-React. De plus, à l’instar de Way2Drug SOMP, il calcule aussi un score de probabilité variant entre 0 et 1 (une forte probabilité d’être un SOM est caractérisée par un score élevé), qui détermine si l’atome du métabolite est un SOM pour l’enzyme associée. Afin de déterminer le score optimal pour l’étape de filtration, les différentes valeurs de scores possibles sont apprises par l’exploitation de la première base de données (cf. 3.3).
- **Xenosite UGT 2.0** [51] Xenosite UGT 2.0 prédit les SOM d’une molécule pour les UGT. Toutes les molécules prédites par MetaPrint2D-React comme étant

potentiellement transformées par les UGT sont utilisées en entrée de cette méthode au format SMILES. Cette méthode permet en complément de Way2Drug SOMP l'identification des SOM et la filtration des réactions catalysées par les UGT prédites par MetaPrint2D-React. De plus, à l'instar de Way2Drug SOMP, il calcule aussi un score de probabilité variant entre 0 et 1 (une forte probabilité d'être un SOM est caractérisée par un plus score élevé), qui détermine si l'atome du métabolite est un SOM pour l'enzyme associée. Afin de déterminer le score optimal pour l'étape de filtration, les différentes valeurs de scores possibles sont apprises par l'exploitation de la première base de données (cf. 3.3).

### 3.2.3 Prédiction de la réactivité vis-à-vis de l'ADN

- **Xenosite Reactivity 2.0** [94, 93] Xenosite Reactivity 2.0 prédit les sites de réactivité (SOR = atomes réactifs vis-à-vis de l'ADN) d'une substance chimique. Chaque molécule et ses métabolites dérivés d'un composé prédit comme réactif sont entrés au format SMILES. Xenosite Reactivity 2.0 permet de déterminer la réactivité d'une molécule et d'identifier ses atomes ayant la capacité de se lier à l'ADN. Le score de réactivité associé varie entre 0 et 1 (une forte probabilité d'un atome d'être réactif est caractérisée par un score élevé).

### 3.2.4 Implémentation de filtres pour limiter la combinatoire

Le pipeline présenté dans la figure 3.1 est composé de 3 étapes successives. La première étape consiste à prédire toutes les réactions que peut subir une molécule. La seconde étape permet de réduire les cartes métaboliques prédites en utilisant un ensemble de méthodes calibrées de prédiction de SOM. La dernière étape identifie les métabolites réactifs avec l'ADN. Trois filtres différents permettant la réduction des graphes prédits par MetaPrint2D-React sont implémentés :

1. Un premier filtre basé sur MetaPrint2D-React consiste à supprimer les réactions prédites qui sont improbables dans le métabolisme des AHA et celles se produisant sur un atome ayant un score NOR (recalculé en utilisant la configuration stricte) en dessous du seuil NOR optimisé.
2. Un second filtre est basé sur Way2Drug SOMP, Xenosite Metabolism 1.0 et Xenosite UGT 2.0. Ce filtre retient seulement les réactions se produisant sur les atomes prédits comme des SOM par au moins l'une des deux méthodes.
3. Un troisième filtre basé sur Xenosite Reactivity 2.0, supprime toutes les réactions de secondes générations augmentant la réactivité (en considérant que la phase II consiste à diminuer la réactivité) et se produisant sur un groupe d'atome différent de celui de la molécule précédemment modifiée lors des réactions de premières générations.

### 3.3 Procédure de calibration

La procédure de calibration repose sur l'exploitation de deux bases de données nécessaires à l'optimisation des différentes méthodes implémentées dans le pipeline.

#### 3.3.1 Base de données biologiques pour l'optimisation

La première base de données d'apprentissage nommée "base de données d'optimisation des paramètres pour la prédiction des SOM" consiste en une liste curée manuellement de 15 molécules contenant des amines aromatiques extraites de la littérature et subissant des réactions d'hydroxylation et/ou de glucuronidation (voir Annexe Table 1). Ces 15 molécules sont impliquées dans 49 réactions de 5 types différents (C-Hydroxylation, N-Hydroxylation, C-Glucuronidation, N-Glucuronidation et O-Glucuronidation). Ces types de réactions représentent celles qui sont majoritairement subies par les différentes amines aromatiques. Cette base de données est utilisée pour calibrer les paramètres des méthodes de prédiction des SOM pour CYP et UGT.

La deuxième base de données, nommée "base de données d'optimisation des paramètres pour l'identification des adduits à l'ADN" est une liste de 30 amines aromatiques curées manuellement et expérimentalement validées comme inductrices d'adduits à l'ADN (voir Annexe Table 2). Cette base de données est utilisée pour définir les différents seuils associés aux prédictions réalisées par Xenosite Reactivity 2.0 afin de classer les métabolites en fonction de leurs réactivités vis-à-vis de l'ADN.

#### 3.3.2 Optimisation des méthodes de prédictions des SOM et des SOR

Afin d'identifier le jeu de paramètres optimaux pour l'étape de filtration, toutes les combinaisons de paramètres qui maximisent tous les scores tout en retenant toutes les molécules contenues dans la première base de données sont testées (Annexe table 1). Cette dernière contient 15 molécules produisant 49 métabolites différents par deux types de réactions majeures (l'hydroxylation et la glucuronidation) retrouvées dans le métabolisme des AHA. Dans cette optique, les apprentissages du score NOR optimal pour la filtration basée sur MetaPrint2D-React, des scores associés à Way2Drug SOMP, Xenosite Metabolism 1.0 et Xenosite UGT 2.0 pour les filtrations basées sur les SOM et des scores de réactivité pour la filtration et la prédiction de la réactivité basée sur Xenosite Reactivity 2.0 (Figure 3.2) sont effectués.

Le premier apprentissage concerne l'**apprentissage du score NOR optimal de MetaPrint2D-React**. Cette étape consiste à chercher par incrémentation graduelle de 0.01 du score Nore, le score optimal retenant le maximum de métabolite de la première base de données (Annexe table 1). MetaPrint2D-React est utilisé en premier lieu pour prédire les métabolites issus des réactions catalysées par les CYP et les UGT pour toutes les molécules de la base de données. Puis, le ratio entre le nombre de métabo-

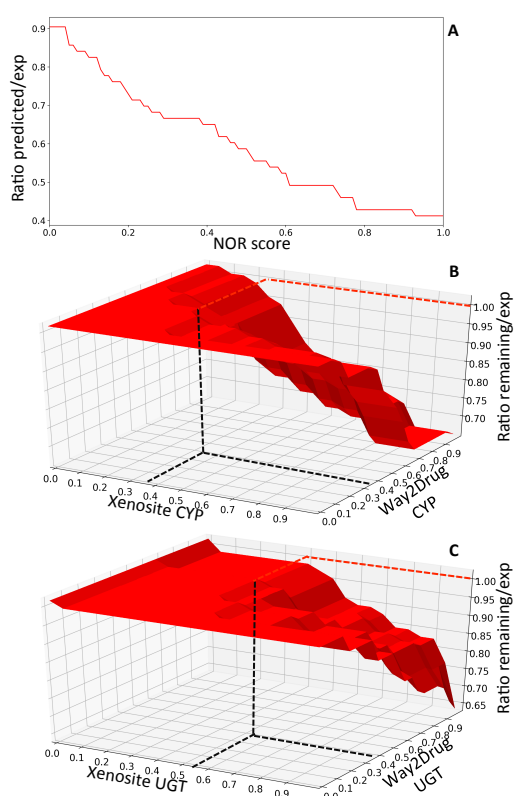


FIGURE 3.2 – Evaluation des ratios décrivant le nombre de métabolites restants après filtrations et le nombre de métabolites expérimentalement identifiés présents dans la première base de données en fonction des différents scores

A : Evaluation du ratio entre le nombre de métabolites restants après filtrations et le nombre de métabolites expérimentalement identifiés présents dans la première base de données en fonction du score NOR de Metaprint2d-React. Chaque réaction et les métabolites correspondants de la première base de données sont prédits avec la configuration "strict" de MetaPrint2D-React. Puis, le présent ratio est évalué en fonction de l'évolution du score NOR ; B : Evaluation du ratio entre le le nombre de métabolites restants après filtrations et le nombre de métabolites expérimentalement identifiés présents dans la première base de données en fonction des scores concernant les CYP de XenositeP450 metabolism 1.0 et Way2Drug SOMP ; C : Evaluation du ratio entre le le nombre de métabolites restants après filtrations et le nombre de métabolites expérimentalement identifiés présents dans la première base de données en fonction des scores XenositeUGT 2.0 et Way2Drug pour les scores des UGT. Pour les expériences B et C, les méthodes consistent à faire varier chaque paramètre et évaluer pour le jeu de paramètres correspondant le ratio précédemment décrit. Pour chaque méthode, l'objectif est de maximiser les paramètres sans affecter la sensibilité (c.-à-d. en conservant tous les métabolites présents dans la base de données). Le meilleur jeu de paramètres correspond donc au jeu où tous les métabolites contenus dans la première base de données sont conservés.



lites prédits par MetaPrint2D-React supérieur au seuil NOR défini et le nombre total de métabolites dans la première base de données est calculé. Par conséquent un ratio inférieur à un 1 (et un seuil NOR défini à 0) indique que certains métabolites de la première base de données expérimentalement identifiés n'ont pas été prédits par Metaprint2D-React (Annexe Table 1).

Comme illustré dans la Figure 3.2A, on remarque que 10% des métabolites expérimentalement identifiés contenus dans la première base de données ne sont jamais prédits par Metaprint2D-React. De plus, la filtration basée sur ce score est très sensible au score NOR. En effet, seulement un métabolite dans cette base de données possède un score de 0.04, et la moitié des métabolites de cette base de données ont un score inférieur à 0.5. Ainsi, pour remédier au biais induit par la sensibilité du score NOR, le score optimal a été défini à 0.04.

Le second apprentissage concerne l'**apprentissage des scores optimaux pour les méthodes basées sur la prédiction des SOM pour les CYP et les UGT**. Cette étape consiste à chercher les ensembles de scores optimaux qui minimise le nombre de métabolites prédits tout en maximisant le nombre de métabolites présents dans la première base de données (Annexe Table 1) pour les méthodes suivantes : Way2Drug SOMP (un score pour UGT et un autre pour CYP), Xenosite P450 Metabolism 1.0 et Xenosite UGT 2.0. Ainsi, en fonction du type de réactions (hydroxylation ou glucuronidation), toutes les combinaisons de paramètres sont testées par étapes incrémentielles de 0.01 pour chaque paramètre. De plus, le ratio entre le nombre de métabolites restant après application du jeu de paramètres associés et le nombre total de métabolites dans la première base de données est calculé pour chaque combinaison de paramètres. Un ratio de 1 signifie que tous les métabolites présents dans la première base de données sont conservés par l'étape de filtration avec le jeu de paramètres donné.

L'impact des scores de Xenosite Metabolism 1.0 et Way2Drug SOMP pour la filtration opérée pour les CYP et ceux de Xenosite UGT 2.0 et Way2Drug SOMP pour la filtration opérée par les UGT est représenté dans les figures 3.2B et 3.2C. On remarque que l'ensemble de la première base de données est intégralement prédit lors de cette seconde phase et est caractérisé sur les figures par deux plateaux. La figure 3.2B indique que le plateau illustrant la conservation des métabolites contenus dans la base de données d'apprentissage est maintenu pour toutes les combinaisons de paramètres pour Xenosite Metabolism 1.0 et Way2Drug SOMP lorsque leurs scores sont respectivement inférieurs à 0.39 et 0.35. Ces derniers correspondent donc aux scores optimaux qui seront utilisés comme référence lors de la filtration. En effet, l'augmentation de ces scores entraîne une perte de métabolites contenus dans la première base de données. De plus, on remarque que l'optimisation basée sur les CYP est plus sensible aux scores issus de Way2Drug SOMP que ceux de Xenosite Metabolism 1.0. Par ailleurs, l'optimisation des scores concernant la prédiction des UGT suit un phénomène similaire à celui observé pour les CYP (Figure 3.2C). En effet, un plateau optimal est atteint pour toutes les combinaisons de scores de Xenosite Metabolism 1.0 et Way2Drug SOMP qui

sont respectivement inférieures à 0.54 et 0.36. Ces derniers scores sont donc sélectionnés comme paramètres du pipeline pour la filtration basée sur les UGT.

Le dernier apprentissage concerne l'**apprentissage des seuils de réactivité avec l'ADN**. La deuxième base de données (Annexe Table 2) est utilisée pour classifier les probabilités de réactivité vis-à-vis de l'ADN en 5 classes (défini dans l'intervalle allant de probabilité nulle à probabilité élevée).

Les probabilités des 30 molécules contenues dans la seconde base de données sont prédites à l'aide de Xenosite Reactivity 2.0. L'analyse de ces résultats montre que 50% des métabolites de cette base de données ont une probabilité comprise entre [0,85, 1] de former des adduits à l'ADN. 20% des métabolites de la base de données d'apprentissage de la réactivité à l'ADN ont une probabilité comprise entre [0.7,0.85]. 20% des métabolites de la même base de données ont une probabilité comprise entre [0.2,0.7]. Enfin, 10% des métabolites ont une probabilité comprise entre [0.05,0.2]. Aucun des métabolites présents dans la deuxième base de données n'a une probabilité inférieure à 0.05. Fort de cette analyse, le pipeline est enrichi par une classification basée sur la probabilité d'un métabolite d'être réactif vis-à-vis de l'ADN au sein de 5 classes respectives : probabilité élevée d'être réactif avec l'ADN (score supérieur à 0.85), probabilité moyenne (score compris entre 0.7 et 0.85), faible probabilité (score compris entre 0.2 et 0.7), très faible probabilité (score compris entre 0.05 et 0.2) et probabilité nulle (score inférieur 0.05). Par extrapolation, cette dernière classe contient les composés non génotoxiques.

Ces étapes de calibration basées sur l'exploitation de deux bases de données biologiques manuellement curées permettent de calibrer chaque méthode avec les scores optimaux définis. La recherche de scores optimaux permet au pipeline de potentiellement conserver une sensibilité élevée, tout en réduisant le graphe brut produit par l'utilisation itérative de Metaprint2d-React et en prédisant les molécules qui formeront potentiellement des adduits à l'ADN.

### 3.4 Validation et analyses statistiques

La validation est basée sur l'exploitation d'une base de données contenant les dérivés métaboliques de trois molécules d'une même famille de xénobiotiques : les AHA. Cette base de données contient une liste de 30 dérivés validés expérimentalement sur deux niveaux de profondeur pour trois AHA : A $\alpha$ C, MeIQx et PhIP (voir chapitre 2 section 2.3.1 pour les métabolismes intégraux expérimentalement caractérisés par les biologistes et la table 3 en Annexe pour les informations complémentaires des métabolites). Cette dernière permet aussi d'évaluer la sensibilité et la précision de la méthode.

Afin de valider les performances de prédictions de métabolisme des xénobiotiques,

l'analyse prédictive de trois AHA (qui sont bien caractérisées en termes de métabolisme et de formation d'adduits à l'ADN dans les hépatocytes humain) sur deux niveaux de profondeur de prédiction est effectuée par l'exploitation du pipeline (voir 3.2) précédemment calibré. Les figures 3.3, 3.4, et 3.5 illustrent les résultats de la prédiction du métabolisme de ces trois AHA, ainsi que leur génotoxicité et l'identification des atomes pouvant potentiellement se lier à l'ADN. Ainsi, ces résultats sont comparés avec les résultats expérimentaux pour A $\alpha$ C, MeIQ $x$ , et PhIP, ainsi que leurs 30 dérivés (Annexe Table 3). Enfin, les performances sont évaluées par le calcul de la sensibilité (nombre de métabolites correctement prédit) et de la précision associées :

$$\text{Sensibilité}(\%) = \frac{[\text{TP}]}{\text{Total}_{\text{métabolites}}} \times 100 \mid \text{Précision}(\%) = \frac{[\text{TP}]}{[\text{TP}] + [\text{FP}]} \times 100$$

où TP représente le nombre de métabolites prédits et expérimentalement identifiés (pour une profondeur définie), FP le nombre de métabolites sur-prédits (non expérimentalement identifiés pour une profondeur donnée) et  $\text{Total}_{\text{métabolites}}$  le nombre total de métabolites expérimentalement identifiés.

Les prédictions du pipeline montrent qu'au total, 64 métabolites issus de réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation sont générés pour A $\alpha$ C (26 dérivés), MeIQ $x$  (16 dérivés), et PhIP (22 dérivés). Parmi ces métabolites prédits, 28 des 30 dérivés connus sur deux niveaux de profondeur pour les trois AHA ont été correctement prédits par le pipeline. Seuls deux métabolites expérimentalement identifiés (N-desmethyl-7-oxo-MeIQ $x$  et 7-oxo-MeIQ $x$ ) résultant de biotransformations de MeIQ $x$  sont absents des prédictions (voir chapitre 2 section 2.3.1 pour le métabolisme du MeIQ $x$  proposé par les biologistes). Ce constat peut être expliqué par l'absence d'empreintes moléculaires de ces réactions dans la base de données de MetaPrint2D-React ou par une nécessité d'améliorer les paramètres de correspondance des empreintes moléculaires.

On remarque que les composés N-sulfonyl-AHA et N-acetoxy-AHA sont tous deux prédits avec une probabilité élevée d'être réactif avec l'ADN (marqués par un encadrement rouge). Ce constat est en accord avec une précédente étude [199], qui s'attachait à montrer la réactivité des composés N-sulfonyl-AHA et N-acetoxy-AHA avec l'ADN. Ces métabolites expérimentalement identifiés et réactifs avec l'ADN sont illustrés par un cercle sur fond rouge contenant les lettres IR (Identifié et Réactif) dans les figures 3.3, 3.4 et 3.5. De plus, tous les métabolites validés expérimentalement réactifs avec l'ADN après une réaction de N-hydroxylation pour A $\alpha$ C, MeIQ $x$  et PhIP [230, 34] sont prédits avec une probabilité élevée de former des adduits à l'ADN. Aussi, le métabolite produit après une glucuronidation est indiqué par un encadré noir dans la figure 3.3 et prédit avec une probabilité moyenne d'être génotoxique. Ce dernier constat a été récemment confirmé expérimentalement [34]. A la différence des composés N-sulfonyloxy-AHA, N-acetoxy-AHA, et HON-AHA, ce métabolite a été expérimentalement mesuré [19] suggérant ainsi une stabilité supérieure aux trois autres structures chimiques. La probabilité moyenne pourrait indiquer la potentielle réacti-

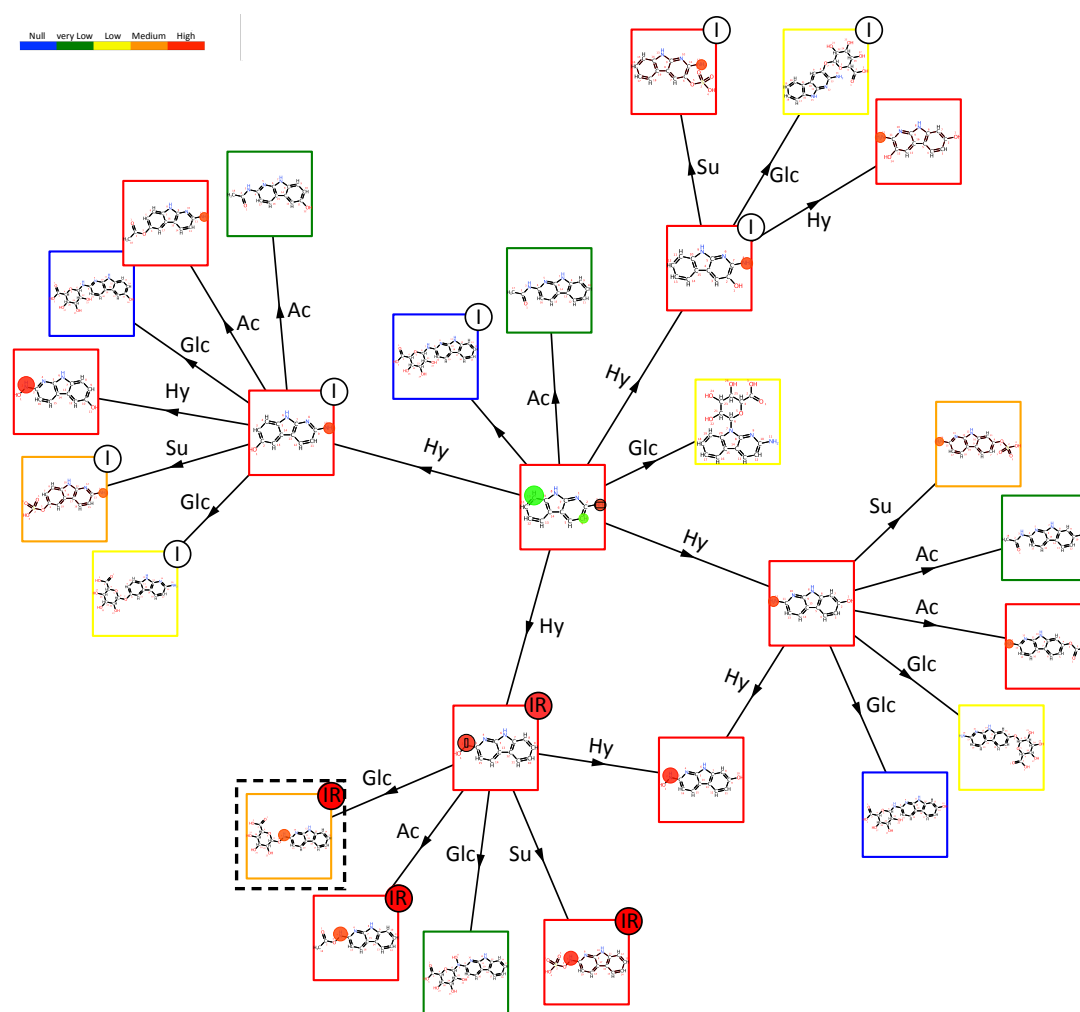


FIGURE 3.3 – **Prédiction du métabolisme d'AαC par le pipeline**

Le pipeline prédit 26 dérivés métaboliques pour AαC, ainsi que leurs génotoxicités associées. Le métabolisme prédit englobe 5 types de réactions biochimiques différentes : Hydroxylation (Hy), Glucuronidation (Glc), Acétylation (Ac), Sulfatation (Su) et Oxydation (Ox). La génotoxicité de chaque métabolite est classée en fonction de 5 niveaux de probabilité de former des adduits à l'ADN : nulle (bleue), très faible (vert), faible (jaune), moyenne (orange) et élevée (rouge). De plus, le(s) atome(s) d'un métabolite ayant la capacité de se fixer à l'ADN sont marqués par un cercle dont sa couleur représente la probabilité de chaque atome à se lier à l'ADN (Faible->Vert, Moyenne->Jaune et Orange, Elevée->Rouge). Dans ce graphe, les métabolites identifiés expérimentalement sont étiquetés par un cercle contenant la lettre I (comme Identifié) et ceux expérimentalement identifiés ET validés comme réactifs avec l'ADN sont étiquetés par un cercle contenant rouge les lettres IR (comme Identifié et Réactif). Concernant AαC, 12 métabolites sont prédits comme hautement réactifs avec l'ADN, 3 comme moyennement, 4 avec une faible probabilité, 4 avec une très faible probabilité et 3 avec une probabilité nulle.

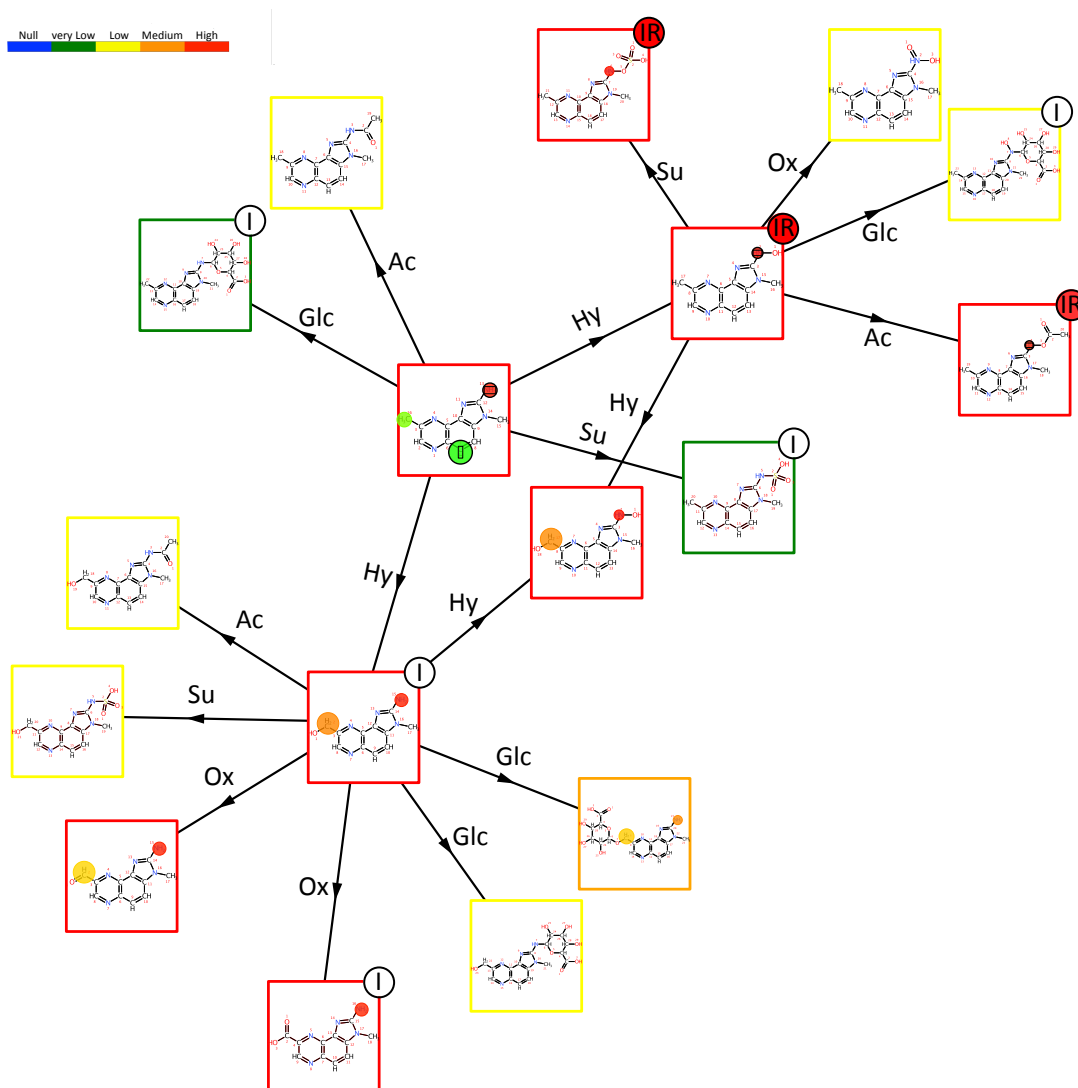


FIGURE 3.4 – Prédiction du métabolisme du MeIQx par le pipeline

Le pipeline prédit 16 dérivés métaboliques pour MeIQx, ainsi que leurs génotoxicités associées. Le métabolisme prédit englobe 5 types de réactions biochimiques différentes : Hydroxylation (Hy), Glucuronidation (Glc), Acétylation (Ac), Sulfatation (Su) et Oxydation (Ox). La génotoxicité de chaque métabolite est classée en fonction de 5 niveaux de probabilité de former des adduits à l'ADN : nulle (bleue), très faible (vert), faible (jaune), moyenne (orange) et élevée (rouge). De plus, le(s) atome(s) d'un métabolite ayant la capacité de se fixer à l'ADN sont marqués par un cercle dont sa couleur représente la probabilité de chaque atome de se lier à l'ADN (Faible->Vert, Moyenne->Jaune et Orange, Elevée->Rouge). Dans ce graphe, les métabolites identifiés expérimentalement sont étiquetés par un cercle contenant la lettre I (comme Identifié) et ceux expérimentalement identifiés ET validés comme réactifs avec l'ADN sont étiquetés par un cercle contenant rouge les lettres IR (comme Identifié et Réactif). Concernant AαC, 7 métabolites sont prédits comme hautement réactifs avec l'ADN, 1 comme moyennement, 6 avec une faible probabilité, 2 avec une très faible probabilité et 0 avec une probabilité nulle.

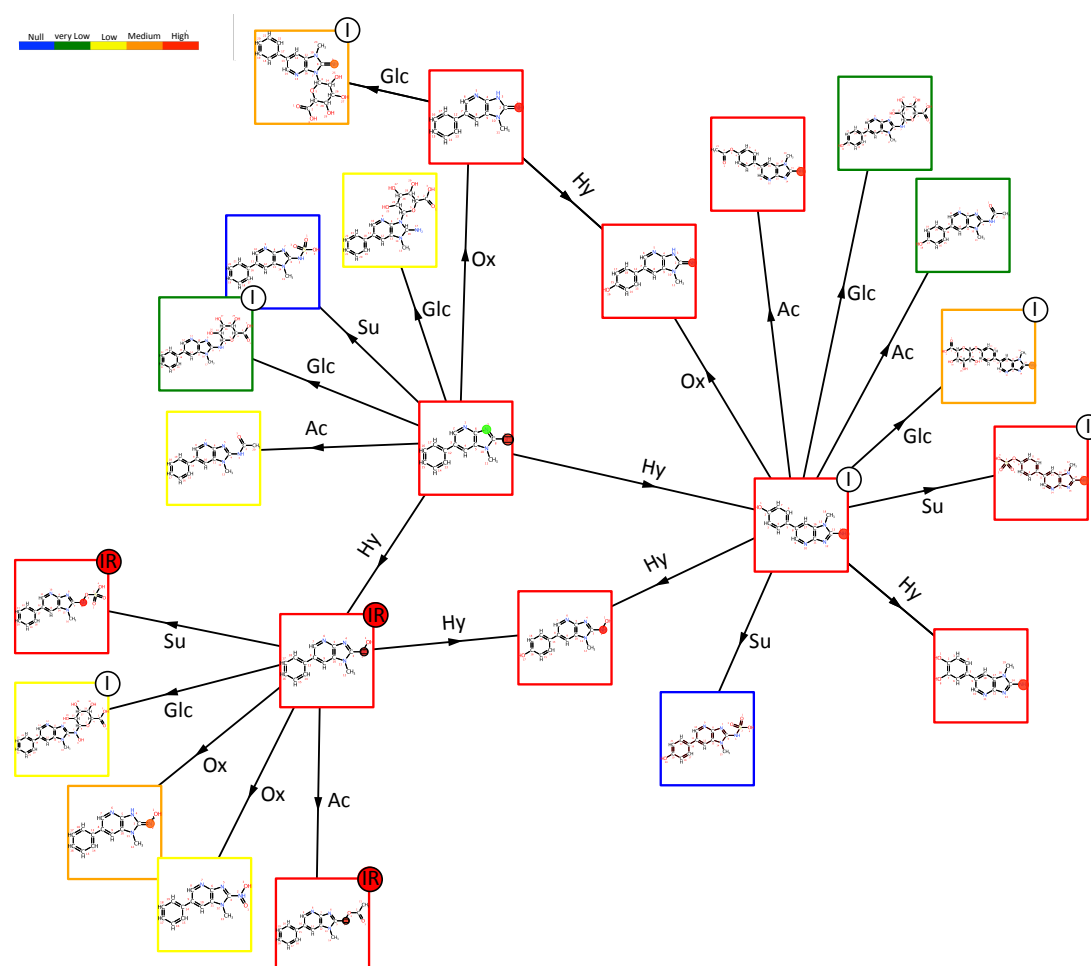


FIGURE 3.5 – **Prédiction du métabolisme du PhIP par le pipeline**

Le pipeline prédit 22 dérivés métaboliques pour PhIP, ainsi que leurs génotoxicités associées. Le métabolisme prédit englobe 5 types de réactions biochimiques différentes : Hydroxylation (Hy), Glucuronidation (Glc), Acétylation (Ac), Sulfatation (Su) et Oxydation (Ox). La génotoxicité de chaque métabolite est classée en fonction de 5 niveaux de probabilité de former des adduits à l'ADN : nulle (bleue), très faible (vert), faible (jaune), moyenne (orange) et élevée (rouge). De plus, le(s) atome(s) d'un métabolite ayant la capacité de se fixer à l'ADN sont marqués par un cercle dont sa couleur représente la probabilité de chaque atome de se lier à l'ADN (Faible->Vert, Moyenne->Jaune et Orange, Elevée->Rouge). Dans ce graphe, les métabolites identifiés expérimentalement sont étiquetés par un cercle contenant la lettre I (comme Identifié) et ceux expérimentalement identifiés ET validés comme réactifs avec l'ADN sont étiquetés par un cercle contenant rouge les lettres IR (comme Identifié et Réactif). Concernant AαC, 10 métabolites sont prédits comme hautement réactifs avec l'ADN, 3 comme moyennement, 4 avec une faible probabilité, 3 avec une très faible probabilité et 2 avec une probabilité nulle.

vité avec l'ADN d'un composé stable et mesurable.

Les résultats de la table 3.1 réfèrent des performances du pipeline par le calcul de la sensibilité et de la précision basé sur les connaissances actuelles du métabolisme d'A $\alpha$ C, MeIQx, et PhIP. On remarque en premier lieu qu'une seule itération de MetaPrint2D-React prédit seulement 11 des 13 dérivés connus de première génération pour les trois AHA. Naturellement, aucun des métabolites de seconde génération ne peut être prédit par l'utilisation itérative de MetaPrint2D-React avec une profondeur fixée à 1. Ce constat montre tout l'intérêt de l'utilisation itérative de MetaPrint2D-React pour prédire le métabolisme complet de molécule d'intérêt. Cependant, deux itérations de MetaPrint2D-React génèrent 789 métabolites parmi lesquels 28 des 30 métabolites connus sont prédits. Cette méthode permet d'atteindre une sensibilité de 93% (sur 2 générations), mais avec une précision de seulement 4%. La faiblesse de ce dernier score illustre la nécessité d'implémenter des méthodes pour limiter la sur-prédiction corrélée aux itérations multiples de MetaPrint2D-React (Figure 3.1).

L'efficacité des procédures de filtrations est basée sur l'implémentation combinée de la prédiction de la sélectivité enzymatique, de la prédiction des SOM pour les enzymes CYP et UGT, ainsi que la prédiction de la réactivité à l'ADN (Table 3.1). On remarque que la filtration permet de réduire significativement le nombre de métabolites prédits pour A $\alpha$ C, MeIQx, et PhIP de 789 à 210 (taux de réduction de 73%, précision égale à 13%). De plus, les 28 dérivés de ces trois AHA prédites par MetaPrint2D-react sont conservés lors de la filtration. A la différence de Piechota et al., la force de ce pipeline réside dans l'augmentation de la précision des résultats sans en modifier la sensibilité. En effet, en utilisant des seuils de coupures implémentés au sein de Meteor, Piechota et al. augmente la précision de 3.4 à 10.7%, mais celle-ci est effectuée au détriment de la sensibilité qui chute de 85.2 à 78.2% pour les métabolites de phase I prédits.

Les graphes réduits contiennent environ 30 types de réactions, mais du fait que seulement 5 réactions majeures ont été identifiées dans le métabolisme des AHA sur deux générations [19, 133, 132], une dernière procédure post-filtration visant à réduire une nouvelle fois ce graphe a été effectué. Ce dernier filtre conserve seulement les réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation catalysées par les enzymes CYP, UGT, NAT et SULT. Cette dernière filtration permet la réduction du nombre de métabolites prédits de 210 à 67 tout en augmentant la précision à 42%.

### 3.5 Discussion et conclusion

Ce travail offre une nouvelle perspective d'étude des xénobiotiques par la prédiction de leurs métabolismes et de leurs capacités à former des adduits à l'ADN aidé par l'utilisation d'une approche informatique intégrative. Grâce à l'implémentation réalisée, il est aisément possible de prédire les dérivés putatifs des xénobiotiques et

Etape du pipeline		Global	A $\alpha$ C	MeIQx	PhIP
<b>1 itération</b>					
Précision	Prédiction du métabolisme par 1 itération de MetaPrint2D-React	15% (11/74)	16% (4/25)	27% (4/15)	9% (3/34)
	Filtration basée sur la sélectivité des CYP, sur les SOM pour CYP et UGT, la réactivé avec l'ADN	28% (11/40)	24% (4/17)	44% (4/9)	21% (3/14)
	Sélection des réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation	58% (11/19)	57% (4/7)	80% (4/5)	43% (3/7)
Sensibilité	Prédiction du métabolisme par 1 itération de MetaPrint2D-React	85% (11/13)	100% (4/4)	67% (4/6)	100% (3/3)
	Filtration basée sur la sélectivité des CYP, sur les SOM pour CYP et UGT, la réactivé avec l'ADN	85% (11/13)	100% (4/4)	67% (4/6)	100% (3/3)
	Sélection des réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation	85% (11/13)	100% (4/4)	67% (4/6)	100% (3/3)
<b>2 itérations</b>					
Précision	Prédiction du métabolisme par 2 itérations de MetaPrint2D-React	4% (28/786)	4% (11/310)	7% (8/107)	2% (9/369)
	Filtration basée sur la sélectivité des CYP, sur les SOM pour CYP et UGT, la réactivé avec l'ADN	14% (28/207)	13% (11/86)	26% (8/31)	10% (9/90)
	Sélection des réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation	44% (28/64)	42% (11/26)	50% (8/16)	41% (9/22)
Sensibilité	Prédiction du métabolisme par 2 itérations de MetaPrint2D-React	93% (28/30)	100% (11/11)	80% (8/10)	100% (9/9)
	Filtration basée sur la sélectivité des CYP, sur les SOM pour CYP et UGT, la réactivé avec l'ADN	93% (28/30)	100% (11/11)	80% (8/10)	100% (9/9)
	Sélection des réactions d'Hydroxylation, de Glucuronidation, d'Acétylation, de Sulfatation et d'Oxydation	93% (28/30)	100% (11/11)	80% (8/10)	100% (9/9)

**TABLE 3.1 – Evaluation des performances du pipeline par l'évaluation des scores de Sensibilité et de Précision appliqués à trois AHA contrôles**

Le pipeline calibré a été appliqué à trois AHA de référence (A $\alpha$ C, MeIQx et PhIP). Les métabolites prédits pour ces trois AHA ont été comparés avec la liste des 30 dérivés métaboliques expérimentalement identifiés pour A $\alpha$ C (11 métabolites), MeIQx (8 métabolites) et PhIP (9 métabolites). Cette liste de référence est aussi utilisée pour le calcul des scores de précision et de sensibilité à chaque étape du pipeline, et appliquée pour le métabolisme global d'A $\alpha$ C, MeIQx et PhIP.



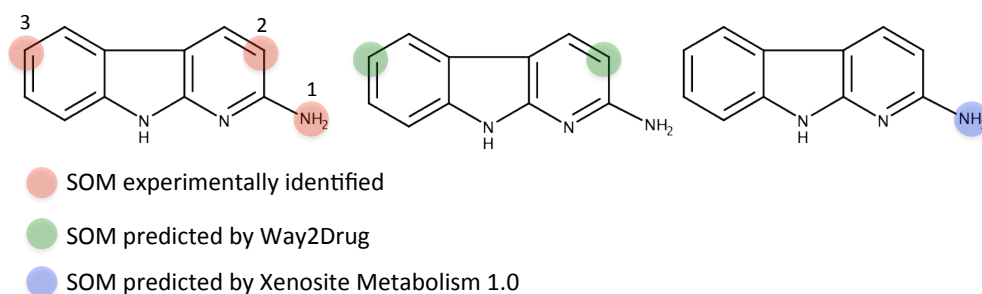
les voies métaboliques engendrant la production de métabolites réactifs avec l'ADN.

Les analyses montrent que la prédiction des enchaînements de réactions par l'utilisation itérative de MetaPrint2D-React génère un nombre important de métabolites, mais avec la caractéristique d'une sensibilité élevée comme illustrée dans la validation de notre méthode sur trois AHA bien documentées. De plus, cette approche permet aussi de détecter des intermédiaires métaboliques absents des données expérimentales en raison de leurs hautes labilités. A titre d'exemple, la prédiction du métabolisme du PhIP suggère que le PhIP-N3-Gluc ne peut être formé directement à partir du PhIP. L'analyse prédictive suppose que la formation de l'intermédiaire 1-méthyl-6-phenyl-1H,2H,3H-imidazo[4,5-b]pyridin-2-imine par oxydation d'une simple liaison en double liaison est nécessaire pour générer le PhIP-N3-Gluc. Ce constat est appuyé par l'exploration aidée par Metaprint2D-React de tous les chemins possibles formant le PhIP-N3-Gluc. L'analyse résultante indique que seulement une voie métabolique incluant cet intermédiaire est possible pour former le PhIP-N3-Gluc (Figure 3.5). Par ailleurs, du fait de la labilité des dérivés N-OH-AHA, N-acetoxo-AHA et N-sulfonyloxy-AHA, ces composés qui sont tous qualifiés expérimentalement comme réactifs avec l'ADN ont échoué à être quantifiés par les biologistes [133, 132].

Un inconvénient majeur de ce type d'approches est la sur-prédiction de métabolites, qui limite la modélisation précise de phénomènes biologiques et nécessite par conséquent l'implémentation de méthodes analytiques additionnelles. Les méthodes de docking et de prédiction des SOM font partie des méthodes principalement exposées dans la littérature pour prioriser et classifier les prédictions métaboliques tout en améliorant la précision (c.-à-d. la capacité à ne pas sur-prédire) [115]. Cependant, tandis que les approches de docking peuvent être coûteuse en temps et en ressources [217], celles basées sur l'exploitation de la prédiction des SOM peuvent souffrir d'une perte en sensibilité lorsque la précision est augmentée [178].

Pour remédier aux problèmes de sur-prédictions inhérents à la prédiction de métabolites sur plusieurs générations, des méthodes de prédiction de SOM pour les CYP et les UGT ont été implémentées en duplicate pour chaque enzyme au sein de celui-ci. Ces dernières permettent d'améliorer significativement la précision en réduisant le nombre de métabolites faux positifs, tout en conservant la sensibilité initiale. Le succès de cette méthode réside principalement dans deux points. Le premier, relève de l'optimisation des paramètres des méthodes utilisées par l'exploitation de bases de données de molécules manuellement curées et contenant des molécules structurellement proches des AHA. Le second point concerne le gain en robustesse par l'exploitation pour chaque enzyme de deux approches basées sur des méthodologies différentes permettant ainsi de réduire les biais propres à chaque méthode. A titre d'exemple, l'analyse d'AαC par Way2Drug SOMP échoue à identifier correctement la fonction amine exocyclique NH<sub>2</sub> comme un SOM pour les isoformes CYP1A2 et 2D6 (qui sont les isoformes précédemment prédites pour le métabolisme d'AαC par WhichCYP). A l'inverse, Xenosite Metabolism 1.0 identifie ce groupe NH<sub>2</sub> avec une probabilité élevée

d'être un SOM lors de la catalyse par ces deux isoformes. Les deux autres SOM sont quant à eux correctement prédits par Way2Drug, mais aussi par Xenosite Metabolism 1.0 qui les prédit avec une faible probabilité (Annexe figure 3.6). Cet exemple illustre parfaitement l'utilité et la nécessité de combiner de multiples méthodes de prédiction pour permettre une identification précise des SOM responsables de la sensibilité et la précision élevées du système.



	Way2Drug			Xenosite		
	1	2	3	1	2	3
1A2	X	0.385	0.498	0.63	0.17	0.34
2D6	X	X	0.447	0.09	0.08	0.57

X not predicted as SOM

FIGURE 3.6 – Exemple de la prédiction individuelle des SOM par Way2Drug SOMP et Xenosite Metabolism 1.0

L'évaluation de la prédiction des SOM d'AαC par Way2Drug SOMP et Xenosite Metabolism 1.0 est effectuée en prédisant les scores pour chaque isoforme de CYP préalablement prédite par WhichCyp comme métabolisant AαC. En utilisant les scores optimisés, chaque méthode utilisée individuellement est incapable de prédire l'intégralité des SOM expérimentalement identifiés pour AαC.

Parallèlement à l'étude de Piechota [178] qui se focalisait sur la prédiction des réactions de phase I à travers la prédiction de métabolites sur 4 générations, les prédictions ont été ici réalisées sur deux générations en incluant les réactions majeures de phase I et II présentes dans le métabolisme des AHA. La restriction des prédictions sur uniquement deux générations est motivée par le fait que les molécules réactives à l'ADN soient produites lors de cette seconde génération [34, 199] (voir 2.3.1). De plus, le choix

de considérer seulement les 5 réactions majeures impliquées dans le métabolisme des AHA permet l'augmentation de la fiabilité des résultats et des analyses. Enfin, à la différence de Piechota et al. qui utilisent des seuils de coupure pour augmenter la précision au détriment de la sensibilité, l'objectif était ici d'optimiser les étapes de filtration à l'aide de la connaissance biologique pour permettre l'augmentation de la précision sans en affecter la sensibilité.

L'analyse post-filtration des résultats en sortie du pipeline suggère qu'il peut exister trois fois plus de dérivés métaboliques pour A $\alpha$ C, MeIQx, et PhIP que ceux qui ont été déjà expérimentalement mesurés. Néanmoins, il convient de considérer qu'il reste probablement toujours des faux positifs parmi ces candidats prédits. En effet, le pipeline n'intègre pas encore de filtration basée sur la prédiction des SOM pour les réactions catalysées par les SULT et les UGT, car aucune méthode n'est à ce jour disponible. De plus, les métabolites résultant de ces transformations catalysées par les SULT et les NAT doivent être considérés avec parcimonie et confirmés par l'implémentation d'autres méthodes tels que le docking, le développement de nouvelles méthodes pour la prédiction des SOM, ou l'utilisation du data-mining sur d'autres bases de données associé à un score statistique à l'image de ce qui est implémenté dans Metaprint2D.

Ces analyses suggèrent aussi une hypothèse majeure dans la prédiction des métabolites, qui est la difficulté de mesurer expérimentalement certains dérivés métaboliques en raison des limites technologiques actuelles de la spectrométrie de masse et/ou de la labilité de certaines substances chimiques. En effet, les méthodes expérimentales actuellement utilisées ne permettent pas la détection de métabolites présents en très faible quantité. Aussi, l'instabilité de certains intermédiaires métaboliques empêche leurs captures par les appareils de mesure. Au regard de cette hypothèse, aucun des intermédiaires du PhIP prédit par l'approche implémentée dans ce travail n'a été à ce jour identifié expérimentalement. Une telle observation est pertinente à l'aune de l'extrême labilité et la difficulté de mesure des molécules réactives avec l'ADN telles que les intermédiaires N-Hydroxylés [226]. En revanche, les métabolites présentant une réactivité avec l'ADN plus faible sont connus pour être plus stables et sont ergo habituellement mesurés. Ce constat suggère donc que les métabolites prédits avec une probabilité plus faible d'être réactif avec l'ADN soient plus stables et plus facilement détectés. Fort de cette hypothèse, le conjugué O-glucuronidé d'A $\alpha$ C récemment identifié expérimentalement est prédit comme moins réactif avec l'ADN suggérant ainsi une plus grande stabilité que les dérivés N-acetoxy et N-sulfonyloxy.

La combinaison de l'approche développée avec une méthode de prédiction de la réactivité fournit de précieuses indications sur la nature des métabolites (intermédiaires ou finaux) et pourrait faciliter leurs identifications expérimentales. Une perspective majeure de cette approche est l'utilisation des graphes bruts et réduits contenant les dérivés métaboliques prédits afin de développer une méthode à l'image de Mass-MetaSite, MetabolitePilot ou encore IMassFrag [177, 261] pour l'identification automatique des métabolites basée sur l'exploitation de données expérimentales ac-

quises par spectrométrie de masse. En effet, ces graphes permettent de fournir de bons candidats pour l'identification de composés par spectrométrie de masse basée sur la génération *in silico* de nouvelles connaissances biologiques fournissant les formules chimiques associées à la prédiction de leurs propriétés moléculaires.

En conclusion, il est proposé dans ce présent travail une nouvelle méthode pour prédire efficacement des réseaux métaboliques, ainsi que leurs métabolites réactifs avec l'ADN. L'efficacité de cette méthode repose sur la combinaison de l'implémentation de prédiction de métabolites permettant une bonne sensibilité et l'amélioration de la précision basée sur l'implémentation de méthodes de prédiction de SOM et de SOR préalablement calibrées et couplées à des connaissances biologiques. Cette contribution ouvre de nouveaux horizons en recherche expérimentale pour une aide à une meilleure et plus subtile estimation des risques des xénobiotiques en santé humaine.

### 3.6 Résumé

La prédiction du métabolisme de xénobiotiques, ainsi que leur génotoxicité chez l'homme est un défi majeur. Actuellement, quelques solutions logiciels proposent des éléments de réponses à cette question, mais le point faible majeur de ces méthodes est l'explosion de la combinatoire associée à la prédiction des métabolites. Dans ce travail, une approche intégrative permettant de prédire automatiquement le métabolisme et la génotoxicité des xénobiotiques chez l'homme tout en limitant la combinatoire est proposée. Les forces de ce travail sont notamment la multiplicité des méthodes intégrées pour prédire avec précision sans perdre en sensibilité et la capacité à prédire pour chaque métabolite leur génotoxicité par la formation d'adduits à l'ADN, ainsi que d'identifier directement les atomes réactifs vis-à-vis de l'ADN. L'idée principale de ce travail consiste à prédire dans un premier temps le métabolisme des xénobiotiques en implémentant MetaPrint2D-React de façon itérative. Puis, dans un second temps des méthodes de prédiction de sites de métabolisme (SOM) (Way2Drug SOM, Xenosite Metabolism 1.0, Xenosite UGT 2.0) et des sites de réactivité (SOR) des métabolites vis-à-vis de l'ADN (Xenosite Reactivity 2.0) sont utilisés pour augmenter la précision de la méthode et limiter ses biais.

Pour permettre la cohésion des différentes méthodes de prédiction combinée, il a été nécessaire de calibrer et optimiser de manière non indépendante chacune de ces méthodes basées sur l'exploitation d'une base de données composée de molécules structurellement proches de la famille d'étude. L'utilisation de substances chimiques proches permet une optimisation adéquate des différentes méthodologies implémentées pour permettre d'améliorer la précision de l'approche intégrative sans impacter la précision.

L'intérêt de la calibration a été prouvé par validation de la méthode en comparant les métabolismes expérimentaux d'AαC, MeIQx et PhIP avec ceux prédits. Les résultats indiquent que 26 des 28 métabolites identifiés expérimentalement sont bien prédits et 9 des 10 métabolites réactifs avec l'ADN sont prédits avec une forte probabilité d'être réactif avec l'ADN. Grâce à l'implémentation d'un pipeline combinant une multitude de méthodologies différentes, il est maintenant possible de prédire aisément les voies métaboliques de xénobiotiques conduisant à la formation d'adduits à l'ADN.

# Chapitre 4

## Cas d'étude : Prédiction des métabolismes des AHA

Le chapitre précédent exposait une méthode générique de prédiction des graphes métaboliques des xénobiotiques. Ce chapitre présentera l'application de cette méthode sur la famille complète des xénobiotiques des AHA (voir chapitre 2) afin de prédire et évaluer leurs métabolismes et génotoxicités. Cette méthode permet la prédiction des 30 graphes métaboliques pour les 30 AHA connues. De plus, il sera présenté une interface web développée afin d'analyser les résultats de la prédiction du métabolisme et de la génotoxicité des AHA à travers la visualisation, l'exploration et la manipulation des graphes prédits.

---

<b>4.1 Prédictions des graphes métaboliques d'une famille de xénobiotique : les AHA</b> . . . . .	<b>80</b>
4.1.1 Matériel biologique d'étude . . . . .	80
4.1.2 Prédictions des graphes métaboliques d'une famille de xénobiotique : les AHA . . . . .	80
<b>4.2 Implémentation logicielle</b> . . . . .	<b>84</b>
<b>4.3 Conclusion</b> . . . . .	<b>91</b>
<b>4.4 Résumé</b> . . . . .	<b>91</b>

---

Il a été précédemment présenté une méthode de construction *de novo* de graphes métaboliques contenant les informations de génotoxicité de chaque molécule prédites. Cette dernière a été calibrée sur un ensemble de molécules structurellement proches et validées sur 30 dérivés métaboliques. La suite de ce travail consiste à appliquer cette approche sur la famille complète des AHA afin de générer de la nouvelle connaissance biologique.

## 4.1 Prédiction des graphes métaboliques d'une famille de xénobiotique : les AHA

### 4.1.1 Matériel biologique d'étude

Le support biologique d'étude utilisée à des fins de génération de nouvelles connaissances biologiques est contenu dans une base de données répertoriant les 30 AHA majeures [161, 234, 169]. Cette dernière inclut aussi A $\alpha$ C, MeIQx et PhIP qui seront prédites par le workflow (chapitre 2 section 2.3 et tables 2.2 et 2.1). Pour rappel, elles sont classées en deux catégories (dépendante de leur température de formation) : les amines aromatiques hétérocycliques pyrolytiques et les amines aromatiques hétérocycliques amino-imidazo-arènes.

### 4.1.2 Prédiction des graphes métaboliques d'une famille de xénobiotique : les AHA

La méthode de construction *de novo* de graphes métaboliques est utilisée avec une profondeur de 2 et a été appliqué pour la prédiction du métabolisme et l'évaluation des 27 autres AHA restantes. Les 30 graphes prédits associés avec les informations de génotoxicité sont résumés dans la table 4.1.

Les prédictions de la génotoxicité pour les 30 AHA ont été comparées celles des trois AHA bien documentées : A $\alpha$ C, MeIQx et PhIP. Ainsi, le niveau de génotoxicité d'une AHA a été défini comme le nombre de métabolites prédits avec une probabilité élevée d'être réactif avec l'ADN sur le nombre de métabolites prédits total dans le graphe. Comme illustrés dans la table 4.1, 64 métabolites pour A $\alpha$ C, MeIQx et PhIP dont 28 identifiés expérimentalement ont été prédits. Parmi ces 64 métabolites, 29 sont prédits comme réactifs avec l'ADN, parmi lesquels 10 ont été validés expérimentalement. De plus, le niveau de génotoxicité pour les métabolites prédits respectivement pour A $\alpha$ C, MeIQx et PhIP est de 46, 43 et 45 %. Ces résultats sont relativement semblables au niveau de génotoxicité des métabolites validés (36, 38, 33 %), ce qui suggère l'efficacité du pipeline à évaluer la génotoxicité.

L'analyse des 27 autres AHA montre qu'elles peuvent être classées en trois groupes distinctes (Table 4.1). Sept AHA possèdent un niveau de génotoxicité compris entre 48% et 54% et supérieur à ceux d'A $\alpha$ C, MeIQx et PhIP. Ces AHA regroupant IQx[4,5b], TrP1, GluP1, GluP2, TrP2, Me $\alpha$ C et Phe-P-1 pourraient être considérées comme des candidates pertinentes de futures analyses en raison de leurs fortes toxicités. Ces composés appartiennent tous, à l'exception d' IQ[4,5b], à la famille des AHA pyrolytiques et sont structurellement proches d'A $\alpha$ C, qui est connu pour être hautement carcinogène [34].

Une seconde classe de 19 AHA a une génotoxicité égale ou inférieure à celle d'A $\alpha$ C, MeIQx et PhIP. Ces AHA appartiennent principalement à la famille amino-imidazo-

arènes suggérant la plus faible génotoxicité de cette famille en comparaison aux AHA pyrolytiques.

Le dernier groupe possédant une génotoxicité nulle contient APNH, AMPNH, Harman, et NorHarman. Les graphes métaboliques prédits pour ces quatre AHA possèdent seulement 0 ou 1 métabolite. Ce faible nombre de métabolites prédits est expliqué par l'implémentation des filtres de réduction de graphes qui ont supprimé les prédictions initiales opérées par MetaPrint2D-React. Une plausible explication réside dans le fait, qu'à l'inverse des autres AHA, ni Harman et ni NorHarman ne possède une fonction amine exocyclique. En effet, comme indiqué dans le schéma général de biotransformations des AHA, les métabolites réactifs vis-à-vis de l'ADN proviennent de la modification de leurs fonctions exocycliques. De plus, les recherches bibliographiques indiquent qu'aucune carcinogénicité directe n'a été expérimentalement identifiée [225]). Les scores de réactivité avec l'ADN par Xenosite Reactivity 2.0 soulignent aussi les faibles réactivités d'APNH et d'AMPNH, qui ont respectivement des scores de 0.66 et 0.70. Ces scores sont largement inférieurs à ceux des autres AHA, qui sont tous supérieurs à 0.85. Cette prédiction suggère qu'APNH et AMPNH pourraient avoir un potentiel génotoxique faible et être moins inducteurs d'adduits à l'ADN que les 26 autres AHA.

Pour évaluer l'impact des 5 réactions majeures des AHA (Hydroxylation (Hy), Glucuronidation (Glc), Acétylation (Ac), Sulfatation (Su) et Oxydation (Ox)) sur la formation des métabolites réactifs avec l'ADN, la réactivité des métabolites produits après une ou deux réaction(s) est exposée dans la figure 4.1. Ainsi, les métabolites résultant de la combinaison de réactions d'hydroxylation (Hy) et d'acétylation (Ac) ou d'hydroxylation (Hy) et de sulfatation (Su) ont une forte probabilité d'être réactifs avec l'ADN. Ce constat est en accord avec les expérimentations biologiques de [34, 199] qui prouvent que ces enchaînements de réactions sont sources d'intermédiaires métaboliques labiles et hautement réactifs avec l'ADN. Ces prédictions sont aussi pertinentes dans l'identification du métabolite  $A\alpha C-HN^2-O-Gluc$  comme un conjugué d' $A\alpha C$  glucuroné réactif avec l'ADN [34].

Il convient aussi de souligner que les enchaînements de réactions combinant les schémas réactionnels hydroxylation et hydroxylation (Hy-Hy) pour 12 AHA ou hydroxylation et oxydation (Hy-Ox) pour 20 AHA forment des métabolites avec une probabilité élevée d'être réactifs avec l'ADN. De tels enchaînements de réactions pourraient être de nouvelles voies de bioactivation à l'origine d'adduits à l'ADN et pourraient être utilisées comme des biomarqueurs communs aux AHA.

Ces analyses suggèrent aussi que la majorité des conjugués résultant de la glucuronidation (Glc) ont une faible ou très faible probabilité d'être réactifs avec l'ADN (Figure 4.1). Malgré l'identification récente de la glucuronidation dans la bioactivation chez  $A\alpha C$  [34], ce constat souligne que la glucuronidation reste majoritairement associée au processus de détoxification [109]. Ainsi, les enchaînements de réactions



AHA	Métabolites prédits			Métabolites validés		
	Nombre de métabolites réactifs avec l'ADN	Nombre total de métabolites	Niveau de génotoxicité	Nombre de métabolites réactifs avec l'ADN	Nombre total de métabolites	Niveau de génotoxicité
Pas de Génotoxicité détectée						
AMPNH	0	1	0%			
APNH	0	1	0%			
Harman	0	0	0%			
NorHarman	0	0	0%			
Génotoxicité égale ou inférieure aux 3 AHA de références						
1,5,6-TMIP	3	8	37%			
1,6-DMIP	3	8	37%			
4,7,8-TriMeIQx	3	8	37%			
6,7-DiMeIQx	3	8	37%			
7,8-DiMeIQx	3	8	37%			
IgQx	3	8	37%			
IQ	3	8	37%			
IQx	3	8	37%			
MeIQ	3	8	37%			
3,5,6-TMIP	4	10	40%			
4'-OH-PhIP	17	39	43%			
4,8-DiMeIQx	7	16	43%			
7,9-DiMeIQx	7	16	43%			
7-MeIQx	7	16	43%			
IFP	4	10	40%			
MeIQx	7	16	43%	3	8	38%
PhIP	10	22	45%	3	9	33%
A $\alpha$ C	12	26	46%	4	11	36%
4-CH2OH-8-MeIQx	7	15	46%			
Génotoxicité élevée						
IQ[4,5-b]	12	25	48%			
TrP1	4	8	50%			
GluP1	3	6	50%			
GluP2	3	6	50%			
TrP2	6	11	54%			
MeA $\alpha$ C	12	22	54%			
PheP1	5	9	55%			

TABLE 4.1 – Classification des AHA en fonction de leur niveau de génotoxicité

Pour chaque graphe métabolique filtré, le ratio entre le nombre de métabolites réactif avec l'ADN et le nombre de métabolites total dans un graphe donné est calculé. Les AHA sont classifiées en fonction de leur niveau de génotoxicité et comparées aux trois AHA de référence. De plus, les génotoxicités de ces trois AHA de références sont aussi rapprochées avec celles expérimentalement validées pour A $\alpha$ C, MeIQx et PhIP. Ces dernières sont mises en valeur dans le tableau par un fond gris. Ces trois AHA comptent 10 métabolites réactifs avec l'ADN parmi les 28 expérimentalement identifiés et 29 des 64 prédits sont aussi identifiés comme réactifs avec l'ADN.

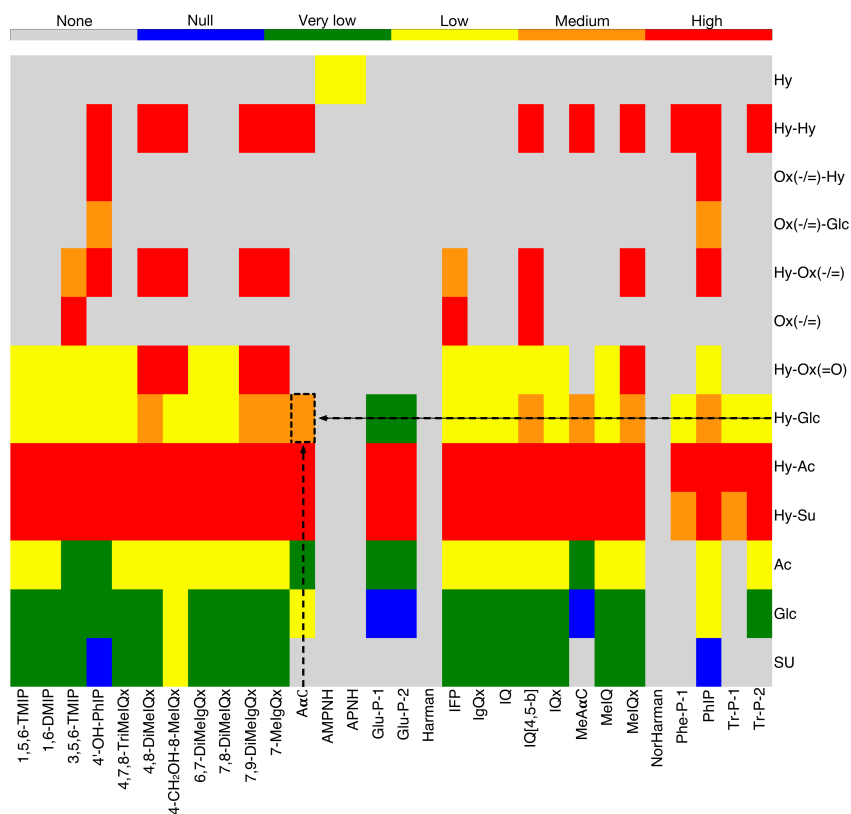


FIGURE 4.1 – Réactivité avec l'ADN des métabolites résultant de la combinaison des réactions d'Hydroxylation (Hy), de Glucuronidation (Glc), d'Acétylation (Ac), de Sulfatation (Su) et d'Oxydation (Ox) appliquée au métabolisme des 30 AHA

Cette heatmap est une représentation globale de la génotoxicité de chaque métabolite produit à la suite d'une combinaison de réactions impliquant les réactions d'Hydroxylation (Hy), de Glucuronidation (Glc), d'Acétylation (Ac), de Sulfatation (Su) et d'Oxydation (Ox) appliquée au métabolisme des 30 AHA. Il convient de noter que deux types d'oxydation sont ici représentés : l'oxydation d'une simple liaison en une double liaison (-/=) et l'addition d'un atome d'Oxygène lié par une double liaison (=O). L'axe horizontal référence les 30 AHA pour lesquelles un graphe de dérivés métaboliques a été prédit par le pipeline. L'axe vertical illustre les enchaînements d'une ou deux réactions produisant un métabolite et étant commun à au moins deux AHA. Pour chaque enchaînement de réactions apparaissant dans le graphe métabolique d'une AHA donnée, l'ensemble des métabolites dérivant d'une combinaison de réactions, ainsi que leurs réactivités avec l'ADN sont intégrées dans la heatmap. Les couleurs indiquent les probabilités de réactivité avec l'ADN de chaque métabolite produit par un enchaînement de réactions. Ces probabilités sont classées en 6 niveaux différents : None (gris), nulle (bleu), très basse (vert), faible (jaune), moyenne (orange), élevée (Rouge). Le niveau "None" indique qu'aucun métabolite n'est produit par l'enchaînement de réactions associé. Le niveau nul signifie que le métabolite associé est considéré par extrapolation comme un produit de détoxification. Enfin, l'encadré noir illustre la voie métabolique produisant le conjugué glucuroné d'AcC réactif avec l'ADN et expérimentalement validé [34].

hydroxylation et glucuronidation conduisent à la formation de métabolites principalement impliqués dans la détoxification, même s'ils peuvent parfois être bioactivés. En d'autres termes, la combinaison des réactions hydroxylation-glucuronidation peut être dans certains cas impliquée à la fois dans des processus de bioactivation et détoxification. Fort de ce constat, il pourrait être intéressant d'approfondir les recherches sur les métabolites glucuronidés résultant du 4,8-DiMeIQx, 7,9-DiMeIqQx, 7-MeIqQx, IQ[4,5-b], Me $\alpha$ C, MeIQx et PhIP.

## 4.2 Implémentation logicielle pour explorer les graphes prédits

Pour permettre la libre exploration des graphes prédits filtrés, l'interface web EPI-GRAPH (Explore Predicted GRAPH) basée sur d3.js [27] a été développée pour visualiser, manipuler et parcourir les voies métaboliques des 30 AHA. Cet outil permet la modification dynamique des graphes en fonction des réactions choisies et/ou d'activation d'options comme l'affichage des annotations des réactions ou encore de la réactivité. De plus, grâce à cette interface, chaque métabolite prédit peut-être indépendamment étudié en cliquant sur le métabolite d'intérêt. Cette action affiche dans un espace dédié, le métabolite, son identifiant SMILES associé et la possibilité de générer toutes les biotransformations possibles associées à chaque atome en cliquant sur l'atome et la réaction désirée. Enfin, le graphe produit par l'utilisateur peut-être exporté aux formats JSON (contenant aussi toutes les informations d'analyses de chaque software), ou en image au format PDF, PNG, ou svg. Cet outil est disponible en ligne à l'adresse suivante : <http://epigraph.genouest.org/>

L'interface web de visualisation et parcours de graphes est composée de trois parties : une partie de contrôle (figure 4.2), une de visualisation et d'exploration du graphe (figure 4.3), et une dernière affichant les détails d'un nœud (figure 4.4).

L'interface de contrôle (figure 4.2) permet de modifier dynamiquement le graphe réduit en fonction de différents critères. Le panneau "Reaction(s) filtering" permet à l'utilisateur de sélectionner le type de réactions métaboliques afin de filtrer le graphe selon les critères désirés. Les options "Display/hide reaction names" et "Display/hide ID numbers" affichent respectivement les noms des types de réactions sur les arrêtes et les numéros d'identification de chaque nœud. "Depth" permet de définir la profondeur du graphe. L'option "root connected" basée sur l'exploitation de l'algorithme de profondeur en largeur BFS (Breadth First Search) permet d'afficher uniquement les molécules qui sont connectées de manière directe ou indirecte à la molécule d'origine en recherchant les différents chemins possibles permettant leur production à partir de la molécule d'origine. Cette option peut être utile lors de la sélection de plusieurs types de réactions qui éclatent le graphe, et par conséquent tous les nœuds ne sont pas forcément connectés à la molécule d'origine. L'option "Only reactive pathways" permet comme son nom l'indique de fournir le graphe contenant uniquement les molécules

réactives avec l'ADN. L'avant-dernière option "Color reactivity" permet à l'utilisateur d'afficher la réactivité de chaque molécule selon les 5 niveaux précédemment définis. Enfin, le bouton "Reset" régénère le graphe d'origine avec les options par défaut. Les autres options servent à l'export du graphe avec la possibilité de choisir le nom du fichier d'export et la dimension de l'image associée.

La deuxième partie (figure 4.3) de l'outil offre à l'utilisateur la possibilité d'explorer le graphe généré à partir des différentes options sélectionnées. Ce graphe est généré avec d3.js et le placement des nœuds est automatiquement réalisé grâce à l'utilisation de l'algorithme "force". De plus, chaque nœud peut être déplacé, fixé/défixé à souhait par l'utilisateur. Enfin, une dernière fonction permet de zoomer dans le graphe en utilisant la molette de la souris.

La dernière partie de l'interface web (figure 4.4) permet à l'utilisateur de s'intéresser au détail d'un nœud. Le parcours de chaque atome, en utilisant le pointeur de la souris, permet d'afficher les réactions pouvant opérer sur celui-ci. De plus, le produit peut-être généré en cliquant sur la réaction désirée. Les chaînes SMILES correspondantes à chaque molécule sont aussi données.

Grâce à l'utilisation de cet outil, il est parfaitement possible de générer aisément les graphes présentés précédemment. A titre d'exemple, la figure 4.5 représente la génération d'un graphe pour A $\alpha$ C affichant tous les types de réactions avec une profondeur égale à 1. Les informations de réactivité de chaque molécule ainsi que les réactions les ayant générées sont aussi incluses dans ce graphe qui peut être exporté aux formats images ou en JSON afin d'être manipulé.

L'utilisation de cet outil pour étudier le MeIQx en ne sélectionnant que les réactions majeures identifiées par les biologistes (Hydroxylation, Glucuronidation, Acétylation, Sulfatation et Oxydation), indique que 16 métabolites sont prédits (voir chapitre 3 figure 3.4). En comparaison avec le graphe expérimentalement validé, les composés N-desmethyl-7-oxo-MeIQx et 7-oxo-MeIQx ne sont pas prédits (voir chapitre 2 section 2.3.1). Bien que les métabolites majeurs aient été identifiés par Langouët et al. [132], des inconnus qui sont probablement des dérivés issus de voies de biotransformations mineures et en très faibles quantités persistent. Ainsi, 8 métabolites en considérant les réactions majeures et 23 en tenant compte des possibles réactions exotiques peuvent être de bons candidats pour l'identification des composés restants (figure 4.6). La force de cette interface réside notamment dans la sélection de réactions que le biochimiste peut privilégier afin de réduire l'espace de recherche de candidats potentiels. A titre d'exemple, la présente figure indique la possibilité de méthylations engendrant la production d'autres métabolites, qui en raison de leurs présences dans le métabolisme de certains xénobiotiques pourraient être aussi considérées lors de l'identification des inconnus demeurant encore dans le métabolisme du MeIQx.



FIGURE 4.2 – Interface de contrôle pour la génération de graphe pour une AHA sélectionnée par l'utilisateur

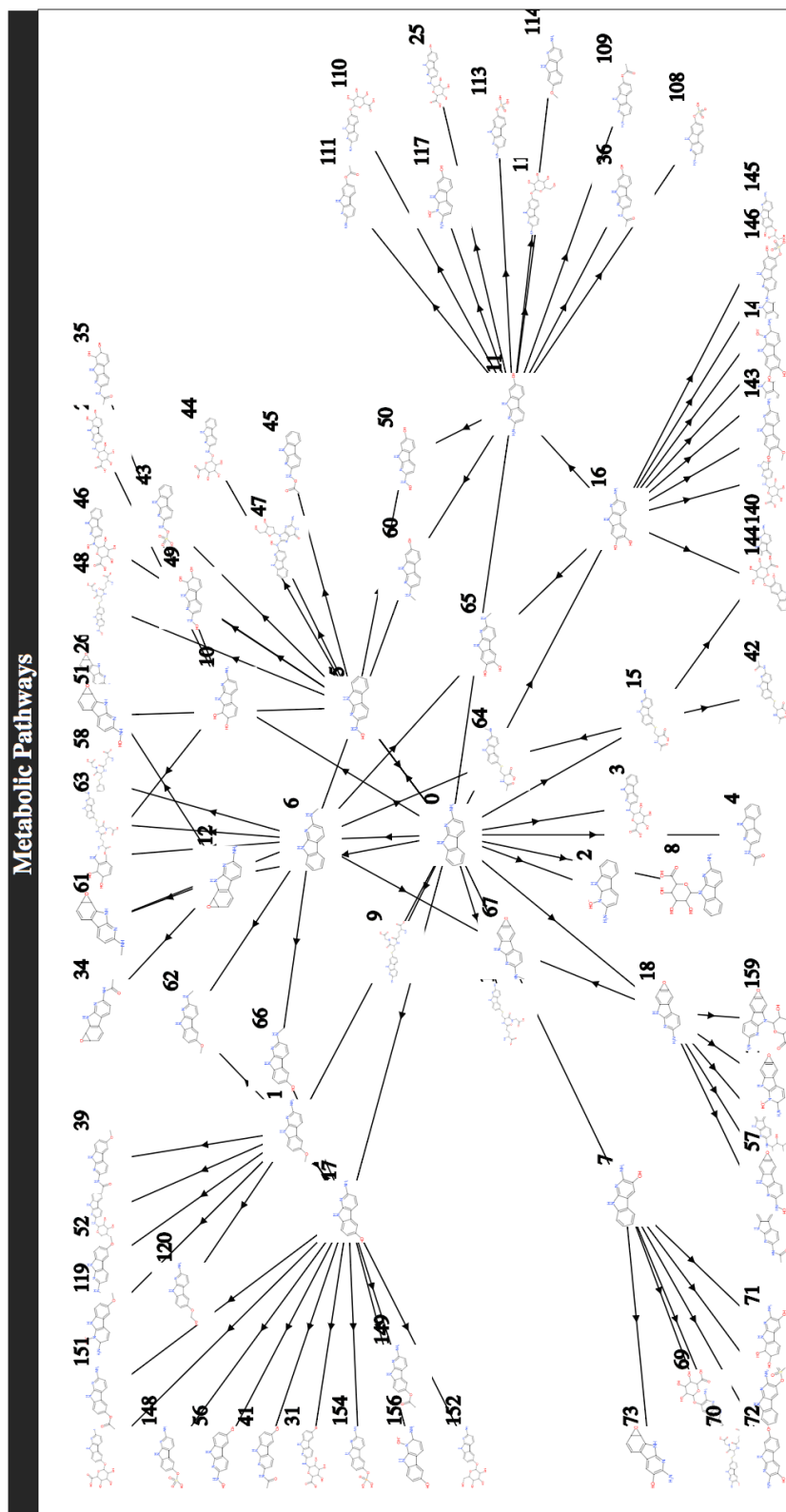


FIGURE 4.3 – Interface d'exploration du graphe généré

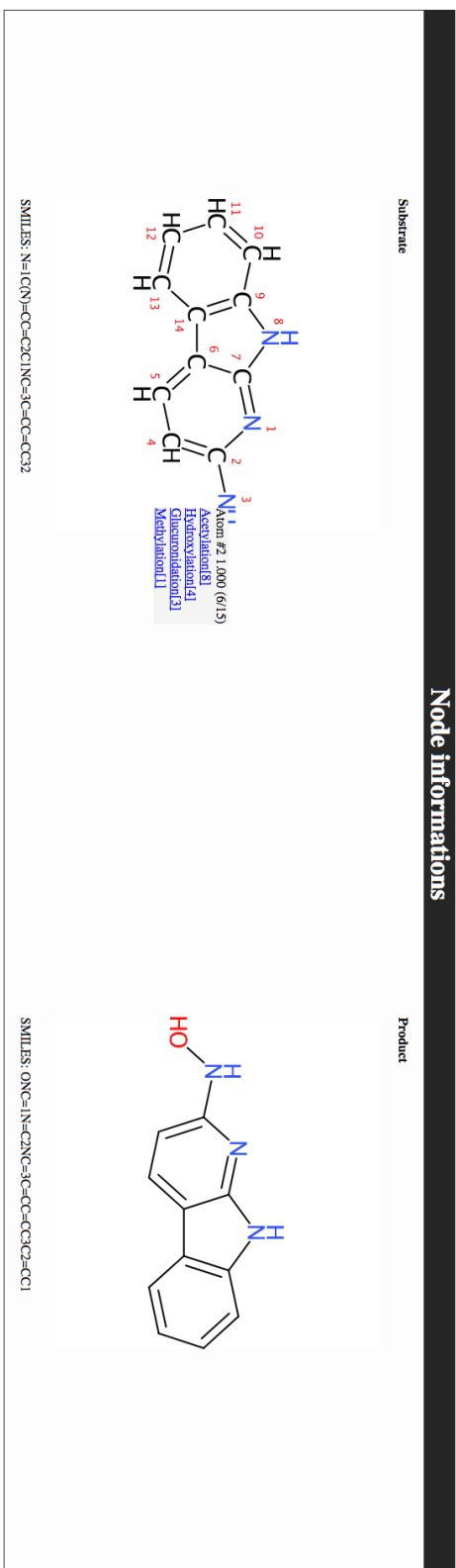


FIGURE 4.4 – Interface d’exploration en détail d’un nœud

Controls panel

**Options**

Reaction(s) filtering  
 Acetylation  
 Acyl/Steratation  
 Epoxidation  
 Epoxidation/hydrolysis

Display/Hide reaction names

Display/Hide ID numbers

Root connected

Only reactive pathways

Depth  sl

Color Reactivity

Reset

**Export**

Filename:

Width

Height

Size

submit Size

Metabolic Pathways

FIGURE 4.5 – Exemple de génération d'un graphe pour AcC avec une profondeur de 1 et les information de réactivité et des types de réactions



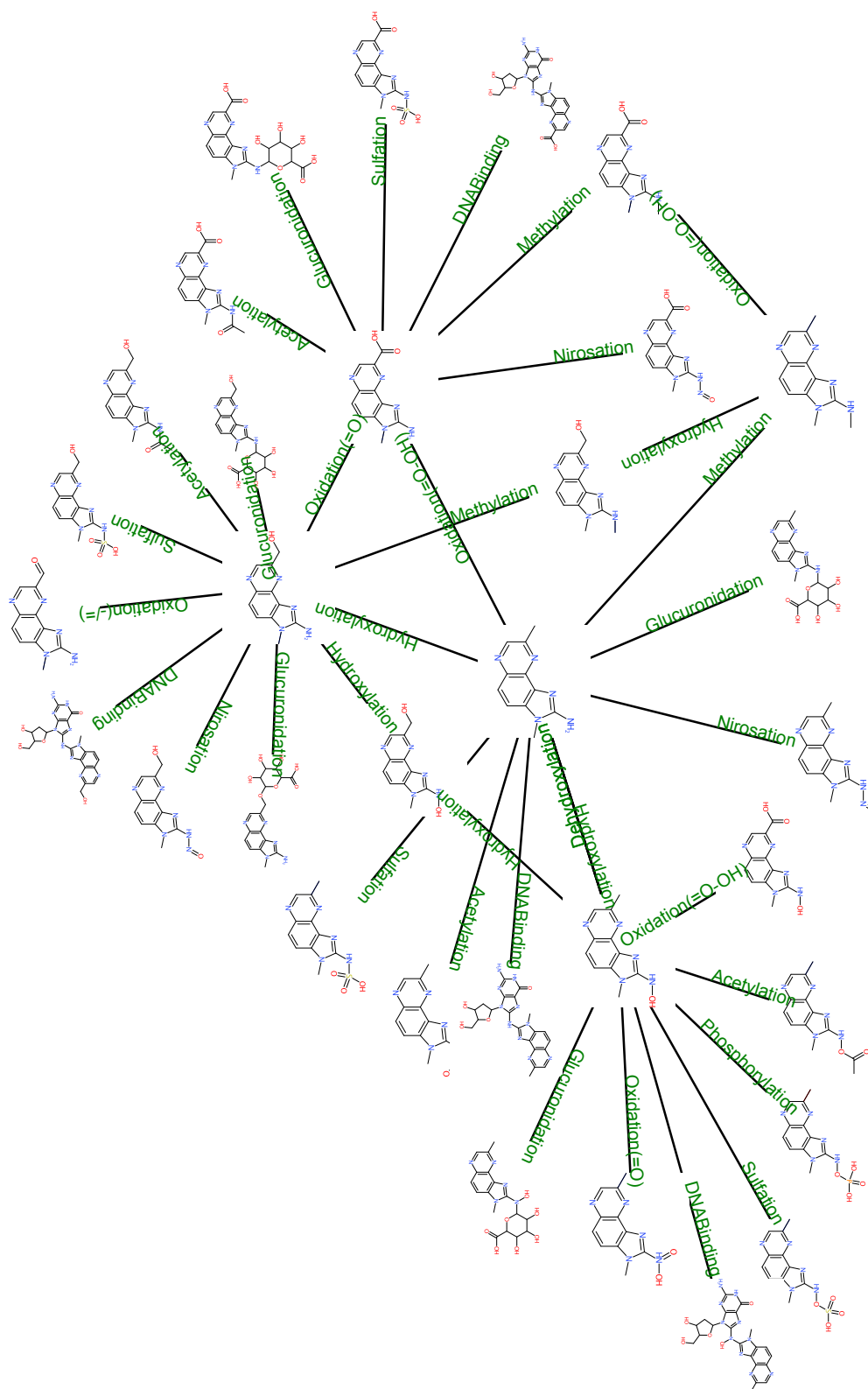


FIGURE 4.6 – Graphe métabolique du MelQx prédit avec deux niveaux de profondeur (extrait de l'interface web)

### 4.3 Conclusion

Ce travail offre une nouvelle perspective d'étude des AHA par la prédiction de leurs métabolismes et de leurs capacités à former des adduits à l'ADN aidé par l'utilisation d'une approche informatique intégrative. Grâce à l'implémentation réalisée, il est aisément possible de prédire les dérivés putatifs de 30 AHA et les voies métaboliques engendrant la production de métabolites réactifs avec l'ADN. De plus, la comparaison des prédictions pour les trois AHA de référence offrent de nouveaux candidats pour l'identification des composés qui demeureraient inconnus suite aux expérimentations biologiques. Cette approche permet aussi la classification de la famille des AHA au sein de trois classes de génotoxicité associant chaque AHA à un risque de toxicité comme potentiellement faible, moyen ou élevé. Cette dernière montre que 23 des 27 AHA peu documentées sont sujettes à une génotoxicité significative et 6 d'entre elles mériteraient une attention particulière. Enfin, elle offre aussi l'identification de nouveaux biomarqueurs des voies de bioactivation par l'identification des réactions finales d'oxydation et d'hydroxylation conduisant à un métabolite réactif avec l'ADN pour une majorité d'AHA.

### 4.4 Résumé

La méthode de construction *de novo* de graphes métaboliques préalablement développée dans le chapitre précédent a été appliquée dans un cadre biologique d'intérêt majeur en recherche cancérologique. Ainsi, la famille des amines hétérocycliques aromatiques (AHA) dont une trentaine ont été actuellement identifiées, mais seulement trois d'entre elles ont été entièrement caractérisées sur un plan métabolique et génotoxique dans les hépatocytes humains ( A $\alpha$ C (2-amino-9H-pyrido [2,3-b] indole), MeIQx (2-amino-3,8- Diméthylimidazo [4,5-f] quinoxaline) et PhIP (2-amino-1-méthyl-6-phényl-imidazo [4,5-b] pyridine)) a été choisie pour générer de la nouvelle connaissance biologique. Pour rappel, ces dernières sont des contaminants environnementaux et alimentaires provoquant des instabilités génomiques par la formation d'adduits à l'ADN et sont classées comme cancérogènes probables ou possibles par l'Agence internationale de recherche sur le cancer (IARC).

Il a été démontré par l'analyse comparative de cette famille que 26 des 30 AHA possèdent une génotoxicité significative et 6 d'entre elles mériteraient une attention particulière. De plus, cette analyse a aussi identifié deux enchaînements de réactions induisant un métabolite réactif avec l'ADN commun à de nombreuses AHA. Enfin, la comparaison des voies métaboliques des différentes AHA permet l'identification de nouveaux biomarqueurs de bioactivation et de détoxification hautement partagés par ces contaminants.



# Développement d'un modèle dynamique pour étudier le métabolisme du MeIQ<sub>x</sub>

Une grande partie des éléments présentés dans ce chapitre, ainsi que le suivant ont fait l'objet d'une publication acceptée dans PeerJ [54]. Ce chapitre présente le développement d'un modèle générique modélisant des effets de compétitions enzymatiques au sein de réactions métaboliques. Ces mécanismes de compétitions enzymatiques affectant l'équilibre des voies métaboliques sont implémentés par le développement d'une méthode justifiant l'emploi de modèles numériques sous-déterminés consistant en la sélection d'un ensemble de jeux de paramètres estimés décrivant la dynamique d'un système.

---

<b>5.1</b>	<b>Introduction</b> . . . . .	<b>94</b>
<b>5.2</b>	<b>Un modèle numérique des compétitions enzymatiques du métabolisme</b> . . . . .	<b>95</b>
5.2.1	Connaissances biologiques . . . . .	95
5.2.2	Données expérimentales . . . . .	95
5.2.3	Modélisation des compétitions enzymatiques . . . . .	99
<b>5.3</b>	<b>Evaluation des dynamiques enzymatiques</b> . . . . .	<b>101</b>
5.3.1	Ajustement de modèles, estimation des paramètres et filtration <i>a posteriori</i> . . . . .	102
5.3.2	La régiosélectivité est dépendante de la concentration de MeIQ <sub>x</sub>	104
5.3.3	73 ensembles de paramètres expliquent les données . . . . .	106
5.3.4	Validation du modèle . . . . .	107
<b>5.4</b>	<b>Discussion et conclusion</b> . . . . .	<b>107</b>
<b>5.5</b>	<b>Résumé</b> . . . . .	<b>113</b>

---

## 5.1 Introduction

La modélisation dynamique d'un système biologique à l'aide d'équations différentielles ordinaires (ODE) est souvent simplifiée par l'utilisation d'une équation de Michaelis-Menten en condition saturante ou non. Néanmoins, la biologie est régie par des mécanismes complexes rendant certains mécanismes classiques de modélisation de biotransformations obsolètes. En effet, il existe pléthore de molécules comme l'ibuprofène [40, 252], l'ifosfamide [188], ou encore le propranolol [235] qui peuvent être catalysées par la même isoforme enzymatique et synthétiser plusieurs produits à partir du même substrat. De plus, ces molécules possèdent pour certaines des cinétiques peu conventionnelles comme cela a été découvert pour le bisphénol A [181], le propranolol [149, 150] ou encore l'ibuprofène [40] qui possèdent toutes une cinétique biphasique. Il a aussi été constaté par détermination graphique que seule la considération de deux  $K_m$  différents illustrant une cinétique biphasique pouvait ajuster les données [40]. L'ensemble de ces observations exposant les dynamiques enzymatiques illustrent la nécessité de développer un modèle simple modélisant la regioselectivité enzymatique, c'est-à-dire la capacité d'une enzyme à favoriser la catalyse d'une réaction par rapport à une autre, tout en prenant en considération l'effet biphasique des cinétiques qui est souvent associé.

Il est parfois difficile de développer un modèle mathématique simple modélisant un phénomène biologique complexe en raison de la difficulté des expérimentateurs biologistes à pouvoir produire les données nécessaires. Par conséquent, les données associées à ce genre de mécanisme sont souvent limitées rendant la tâche difficile. Pour permettre l'exploitation de ces données restreintes, ce travail s'est attaché à proposer une méthode modélisant des processus de compétition enzymatiques affectant l'équilibre des voies métabolique et facilitant le développement d'approches de toxicologie systémique lorsque les données expérimentales disponibles sont très limitées. Cette approche est appliquée dans le cadre de l'étude du métabolisme du MeIQx chez l'homme, qui est une molécule qui possède la particularité d'être catalysée par la même isoforme enzymatique (CYP1A2) pour former deux produits différents (un toxique et l'autre non toxique). De plus, il a été démontré que la préférence de catalyse de ces produits change en fonction de la concentration du MeIQx. Elle a aussi la caractéristique d'avoir des enzymes partagées au sein des voies métaboliques catalysées affectant la dynamique enzymatique. Ces constatations illustrent la complexité des processus de régiosélectivités et l'intérêt de les modéliser notamment dans un cadre de toxicologie pour aider à la compréhension des processus médiant la toxicité.

Les données acquises sur le MeIQx possèdent l'avantage de mesurer les métabolites produits à trois concentrations initiales de MeIQx. Néanmoins, ces données restent limitées par le faible nombre d'expérimentations et le nombre important de paramètres dans le système. Étant donné l'insuffisance des jeux de données par rapport au nombre de paramètres du modèle pour les estimer précisément, la méthodologie classique consistant à sélectionner le jeu de paramètres ajustant au mieux les données

expérimentales a été écartée. Ainsi, il a donc été choisi de sélectionner en fonction de certains critères tous les jeux de paramètres ajustant bien les données expérimentales. À cet effet, 260 modèles paramétrés répondant aux critères d'exigences ont d'abord été identifiés, puis la procédure de filtrage a réduit la famille à 73 modèles numériques qui expliquent toutes les données de manière équivalente. Cette famille de modèles rend possible la prédiction de la régiosélectivité pour toute concentration initiale de MeIQx et à tout temps  $t$ . De plus, les prédictions des 73 modèles à faible dose sont compatibles avec les observations expérimentales de Gu et al. et Stillwell et al. [76, 208]. Ces arguments justifient donc l'étude d'une famille de plusieurs modèles au lieu d'une seule et la fiabilité de la méthode développée.

## 5.2 Un modèle numérique des compétitions enzymatiques du métabolisme

### 5.2.1 Connaissances biologiques

L'ensemble des biotransformations du MeIQx mises en évidence dans la littérature ont été présentées dans le chapitre 2 sous-section 2.3.1 et sont récapitulées dans la figure 5.1, qui expose aussi la génotoxicité des différents métabolites par un système de couleurs (toxique : rouge et orange ; non toxique : vert). Pour rappel, la formation de l'acide 2-amino-3-méthylimidazo[4,5-f]quinoxaline-8-carboxylique (IQx-8-COOH) est une voie de détoxification majeure du MeIQx catalysée par le CYP1A2. Ce composé résulte d'une double oxydation successive du MeIQx impliquant la formation de l'intermédiaire : 2-Amino-8-(hydroxyméthyl)-3-méthylimidazo[4,5-f]quinoxaline (8-CH<sub>2</sub>-OH-IQx) [233, 76]. Le IQx-COOH prédomine pour des concentrations de MeIQx faibles et approchant l'exposition humaine alors qu'il est mineur à des concentrations plus élevées. Par ailleurs, le MeIQx peut être bioactivé en N-hydroxylamine (HONH-MeIQx) qui peut à son tour subir des estérifications catalysées par les enzymes N-acétyltransférases (NAT) et sulfotransférases (SULT). L'estérification par ces 2 enzymes aboutit à la formation de composés toxiques, qui se transforment spontanément en ions nitrenium réactifs vis-à-vis de l'ADN [234]. Ces composés génotoxiques pouvant former des ions nitrenium sont notés "Composé-Potentiellement-Génotoxique". Les métabolites réactifs HONH-MeIQx peuvent également être conjugués par les UGT formant des composés N<sup>2</sup>-glucurono-conjugués (HON-MeIQx-N<sup>2</sup>-Gl) impliqués dans la détoxification. La détoxification est également décrite par la formation du MeIQx-N<sup>2</sup>-SO<sub>3</sub>H et du MeIQx-N<sup>2</sup>-Gl. N-desméthyl-7-oxo-MeIQx et 7-oxo-MeIQx sont quant à eux des produits de détoxification probablement formés par une réaction catalysée par la xanthine déshydrogénase.

### 5.2.2 Données expérimentales

Les données expérimentales sont issues de la publication de Langouët et al. [132], où des hépatocytes humains ont été isolés à partir de biopsies hépatiques humaines.

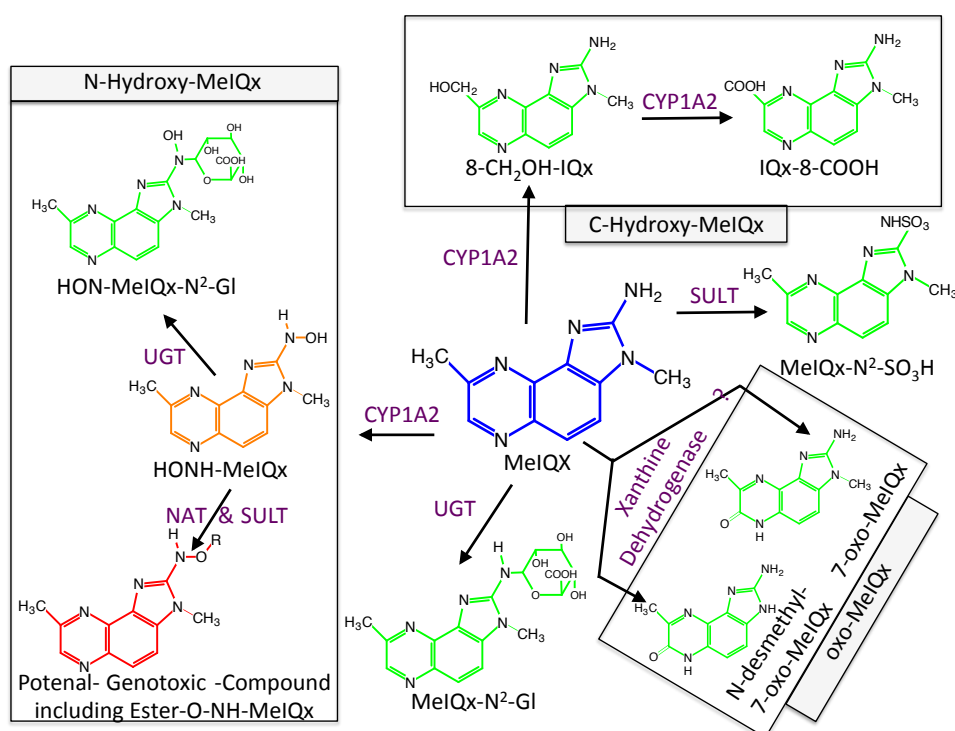


FIGURE 5.1 – Métabolisme du MeIQx (adaptée de [132])

Le métabolisme du MeIQx peut être décomposé en métabolites impliqués dans les voies de détoxification (marqués en vert) et en dérivés métaboliques impliqués dans les voies de bioactivation (marqués en rouge et orange)

Nom complet des molécule chimique (IUPAC) : MeIQx, 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline; HONH-MeIQx, 2-(hydroxyamino)-3,8-dimethylimidazo[4,5-f]quinoxaline; MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, N<sup>2</sup>-(3,8-dimethylimidazo[4,5-f]quinoxalin-2-yl)sulfamic acid; MeIQx-N<sup>2</sup>-Gl, N<sup>2</sup>-(β-1-glucosiduronyl)-2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline; 8-CH<sub>2</sub>OH-IQx, 2-amino-8-(hydroxymethyl)-3-methylimidazo[4,5-f]quinoxaline; IQx-8-COOH, 2-amino-3-methyl-imidazo[4,5-f]quinoxaline-8-carboxylic acid; HON-MeIQx-N<sup>2</sup>-Gl, N<sup>2</sup>-(β-1-glucosiduronyl)-2-(hydroxyamino)-3,8-dimethylimidazo[4,5-f]quinoxaline; 7-oxo-MeIQx, 2-amino-3,8-dimethyl-6-hydro-7H-imidazo[4,5-f]quinoxalin-7-one; N-desmethyl-7-oxo-MeIQx, 2-amino-6-hydro-8-methyl-7H-imidazo[4,5-f]quinoxalin-7-one;

Ils ont été ensuite traités avec trois concentrations différentes de MeIQx (1, 10 et 50  $\mu\text{M}$ ). Les milieux de culture ont été récupérés à 24h et la concentration molaire des 10 métabolites suivants a été estimé : MeIQx, MeIQx-N<sup>2</sup>-Gl, MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, N-desmethyl-7-oxo-MeIQx, 7-oxo-MeIQx, IQx-8-COOH, 8-CH<sub>2</sub>-OH-IQx, HON-MeIQx-N<sup>2</sup>-Gl, HONH-MeIQx et Composé-Potentiellement-Génotoxiques (Figure 5.2).

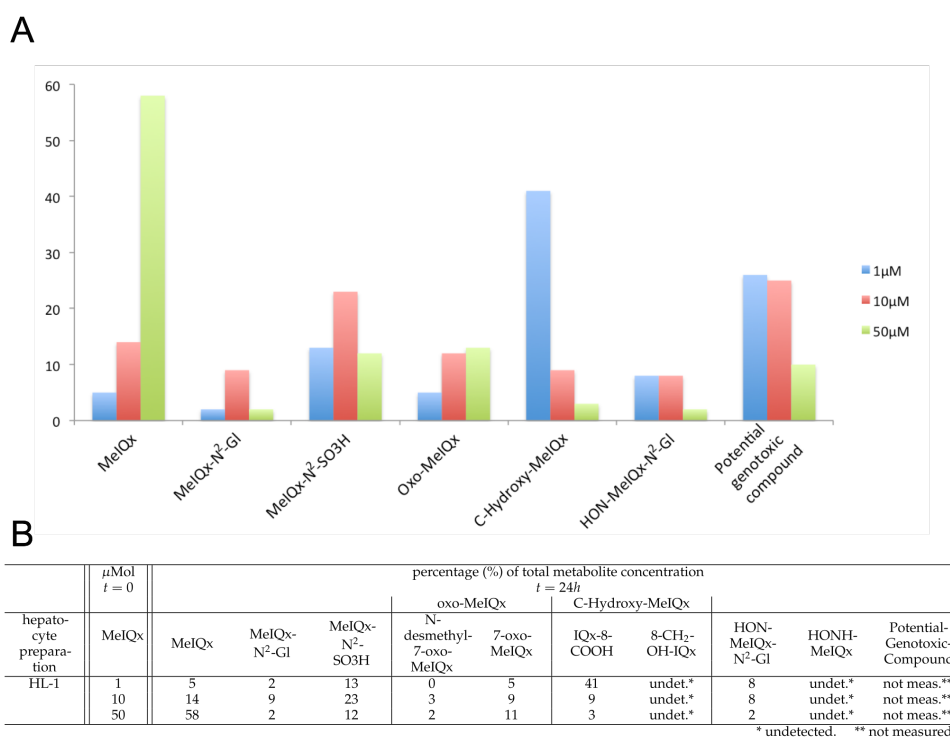


FIGURE 5.2 – Distribution des métabolites du MeIQx en fonction de sa concentration initiale dans les hépatocytes humains (adaptée de [132])

Les données illustrées par un histogramme (A) et complétées par une table (B) sont exprimées en pourcentage de la concentration initiale de MeIQx.

Afin de diminuer le nombre de variables dans le système, plusieurs réductions basées sur les hypothèses suivantes ont été opérées :

1. On suppose que 8-CH<sub>2</sub>-OH-IQx a un effet transitoire car il n'a pas été détecté à  $t = 24h$ . Par conséquent, la voie qui hydroxylait le MeIQx en 8-CH<sub>2</sub>-OH-IQx et IQx-8-COOH a été fusionnée en une réaction générique qui produit un composé générique appelé "C-Hydroxy-MeIQx". Ce composé générique rassemble tous les composés hydroxylés sur le carbone C8.
2. On suppose aussi que la formation de N-desmethyl-7-oxo-MeIQx et de 7-oxo-MeIQx est médiée par la même enzyme, probablement la xanthine déshydrogénase. Un nouveau composé générique appelé "oxo-MeIQx" a été introduit dans



le modèle. Ce dernier rassemble tous les composés avec un groupe chimique oxo. Ce métabolite générique oxo-MeIQx est produit à partir d'une seule réaction générique transformant le substrat MeIQx.

3. On suppose que le HONH-MeIQx a un effet transitoire parce qu'il n'a pas été détecté à  $t = 24h$ .
4. Afin d'établir un système fermé respectant la loi de conservation de la masse, on suppose que tout le MeIQx présent à  $t = 0$  a soit été transformé et réparti entre les différents métabolites du système, ou soit n'a pas été encore métabolisé. De ce fait, une dernière variable nommée Composé-Potentiellement-Génotoxique est introduite dans le système. Elle est estimée par la différence entre la concentration initiale de MeIQx et les concentrations métaboliques mesurées. Cette variable correspond à l'Ester-O-NH-MeIQx et tout autre composé potentiellement non identifié. Il convient de noter que dans ce système biologique, l'Ester-O-NH-MeIQx est le principal métabolite non mesuré et que tous les autres composés potentiellement non identifiés seraient par conséquent en quantité infimes.

L'intégration de ces hypothèses a abouti à l'élaboration d'un modèle réduit (figure 5.3), où trois nouveaux métabolites génériques sont ainsi estimés :

$$\begin{aligned}
 [\text{C-Hydroxy-MeIQx}] &= [\text{IQx-8-COOH}] + [8\text{-CH}_2\text{-OH-IQx}] \\
 [\text{oxo-MeIQx}] &= [\text{N-desmethyl-7-oxo-MeIQx}] \\
 &\quad + [7\text{-oxo-MeIQx}] \\
 [\text{Composé-Potentiellement-Génotoxique}] &= [\text{MeIQx}_{t=0}] - [\text{MeIQx}_{t=24}] \\
 &\quad + [\text{MeIQx-N}^2\text{-Gl}] + [\text{MeIQx-N}^2\text{-SO}_3\text{H}] \\
 &\quad + [\text{oxo-MeIQx}] + [\text{C-Hydroxy-MeIQx}] \\
 &\quad + [\text{HON-MeIQx-N}^2\text{-Gl}]
 \end{aligned}$$

Une estimation exhaustive est obtenue pour les concentrations des différents composés impliqués dans le modèle numérique au temps  $t = 24h$  : MeIQx, C-Hydroxy-MeIQx, MeIQx-N<sup>2</sup>-Gl, MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, oxo-MeIQx, HONH-MeIQx, HON-MeIQx-N<sup>2</sup>-Gl et Ester-N-OH-MeIQx. Les quantités pour chaque composé sont représentées dans la figure 5.2 sauf pour HONH-MeIQx pour lequel on postule une concentration nulle en raison de son effet transitoire.

L'enjeu de cette étude est de comparer les voies de bioactivation avec celles de détoxification en fonction de la concentration de MeIQx. À cet effet, les deux variables suivantes modélisant la bioactivation et la détoxification ont été définies comme sorties du système.

$$\begin{aligned}
[\text{Détoxification}] &= [\text{MeIQx-N}^2\text{-Gl}] + [\text{MeIQx-N}^2\text{-SO}_3\text{H}] + [\text{oxo-MeIQx}] \\
&\quad + [\text{C-Hydroxy-MeIQx}] + [\text{HON-MeIQx-N}^2\text{-Gl}] \\
[\text{Bioactivation}] &= [\text{Composé-Potentiellement-Génotoxique}]
\end{aligned}$$

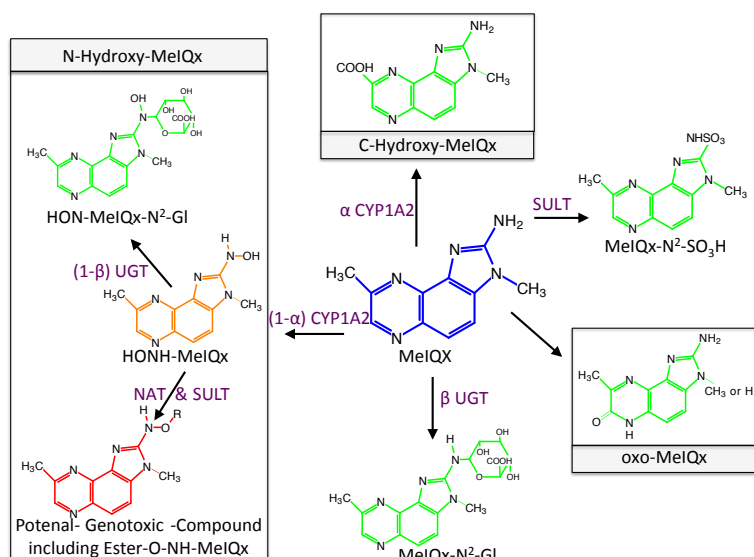
Ces deux variables permettent aussi d'implémenter un ratio de toxicité *Bioactivation/Détoxification* pour évaluer l'importance de chaque voie métabolique en fonction de la concentration de MeIQx.

### 5.2.3 Modélisation des compétitions enzymatiques

Pour modéliser ces effets de compétition, un modèle numérique d'équations différentielles (ODE) décrivant le métabolisme du MeIQx a été construit à partir des données et des hypothèses précédemment présentées. Le modèle d'ODE contient 8 variables : MeIQx, C-Hydroxy-MeIQx, MeIQx-N<sup>2</sup>-Gl, MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, oxo-MeIQx, HONH-MeIQx, HON-MeIQx-N<sup>2</sup>-Gl et le Composé-Potentiellement-Génotoxique. Les trois réactions enzymatiques qui ne présentent aucune dépendance vis-à-vis d'autres réactions, c'est-à-dire les réactions formant l'Oxo-MeIQx, le MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, et le Composé-Potentiellement-Génotoxique, ont été modélisé à l'aide d'une dynamique de Michaelis-Menten. Chaque fonction de Michaelis-Menten a été paramétrée par une constante de Michaelis  $K_m$  et une vitesse maximale  $V_m$ . Pour ces trois réactions, le modèle présente six paramètres :  $Vm_{\text{Composé-Potentiellement-Génotoxique}}$ ,  $Vm_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}$ ,  $Vm_{\text{oxo-MeIQx}}$ ,  $Km_{\text{Composé-Potentiellement-Génotoxique}}$ ,  $Km_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}$ ,  $Km_{\text{oxo-MeIQx}}$ .

Une caractéristique principale de ce modèle est l'implication des enzymes CYP1A2 et UGT au sein de deux réactions différentes conduisant à des compétitions aux ressources enzymatiques disponibles pour chaque réaction. Tout d'abord, le CYP1A2 métabolise le même substrat MeIQx en deux produits :  $P1 = \text{C-Hydroxy-MeIQx}$  et  $P2 = \text{HONH-MeIQx}$ . L'enzyme UGT métabolise quant à elle à la fois les substrats  $S1 = \text{MeIQx}$  et  $S2 = \text{HONH-MeIQx}$  en produits  $P1 = \text{MeIQx-N}^2\text{-Gl}$  et  $P2 = \text{HON-MeIQx-N}^2\text{-Gl}$ . Dans les deux cas, une enzyme  $E$  catalyse deux réactions  $R1$  et  $R2$ . Dans ce contexte, deux paramètres  $\alpha$  (pour la compétition CYP1A2) et  $\beta$  (pour la compétition UGT) ont été introduit pour tenir compte de la proportion de la concentration enzymatique totale utilisée dans chacune des réactions. Ainsi, le paramètre  $\alpha$  [CYP1A2] et  $(1 - \alpha)$  [CYP1A2] désigne la quantité de CYP1A2 impliquée dans la catalyse respective du MeIQx en C-Hydroxy-MeIQx et HONH-MeIQx. En outre, le paramètre  $\beta$  [UGT] et  $(1 - \beta)$  [UGT] reflète la quantité d'UGT impliquée dans la catalyse respective du MeIQx en MeIQx-N<sup>2</sup>-Gl et de HONH-MeIQx en HON-MeIQx-N<sup>2</sup>-Gl (Figure 5.4).

En plus des coefficients de distribution  $\alpha$  et  $\beta$  modélisant les compétitions pour les ressources enzymatiques disponibles, dix autres paramètres ont été introduits pour ces quatre réactions : quatre constantes catalytiques ( $k_1, k_2, k_3, k_4$ ), quatre constantes de Michaelis-Menten ( $Km_1, Km_2, Km_3, Km_4$ ), et la concentration totale de CYP1A2



$$\begin{aligned} \frac{d[\text{MeIQx}]}{dt} &= -\frac{\alpha k_1 [\text{CYP1A2}] [\text{MeIQx}]}{K_{m_1} + [\text{MeIQx}]} - \frac{(1-\alpha) k_2 [\text{CYP1A2}] [\text{MeIQx}]}{K_{m_2} + [\text{MeIQx}]} \\ &\quad - \frac{\beta k_3 [\text{UGT}] [\text{MeIQx}]}{K_{m_3} + [\text{MeIQx}]} - \frac{V_{m_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}} [\text{MeIQx}]}{K_{m_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}} + [\text{MeIQx}]} \\ &\quad - \frac{V_{m_{\text{oxo-MeIQx}}} [\text{MeIQx}]}{K_{m_{\text{oxo-MeIQx}}} + [\text{MeIQx}]} \\ \frac{d[\text{C-Hydroxy-MeIQx}]}{dt} &= \frac{\alpha k_1 [\text{CYP1A2}] [\text{MeIQx}]}{K_{m_1} + [\text{MeIQx}]} \\ \frac{d[\text{MeIQx-N}^2\text{-Gl}]}{dt} &= \frac{\beta k_3 [\text{UGT}] [\text{MeIQx}]}{K_{m_3} + [\text{MeIQx}]} \\ \frac{d[\text{MeIQx-N}^2\text{-SO}_3\text{H}]}{dt} &= \frac{V_{m_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}} [\text{MeIQx}]}{K_{m_{\text{MeIQx-N}^2\text{-SO}_3\text{H}}} + [\text{MeIQx}]} \\ \frac{d[\text{oxo-MeIQx}]}{dt} &= \frac{V_{m_{\text{oxo-MeIQx}}} [\text{MeIQx}]}{K_{m_{\text{oxo-MeIQx}}} + [\text{MeIQx}]} \\ \frac{d[\text{HONH-MeIQx}]}{dt} &= \frac{(1-\alpha) k_2 [\text{CYP1A2}] [\text{MeIQx}]}{K_{m_2} + [\text{MeIQx}]} - \frac{(1-\beta) k_4 [\text{UGT}] [\text{HONH-MeIQx}]}{K_{m_4} + [\text{HONH-MeIQx}]} \\ &\quad - \frac{V_{m_{\text{Compose-Potentielle-Genotoxique}}} [\text{HONH-MeIQx}]}{K_{m_{\text{Compose-Potentielle-Genotoxique}}} + [\text{HONH-MeIQx}]} \\ \frac{d[\text{HONH-MeIQx-N}^2\text{-Gl}]}{dt} &= \frac{(1-\beta) k_4 [\text{UGT}] [\text{HONH-MeIQx}]}{K_{m_4} + [\text{HONH-MeIQx}]} \\ \frac{d[\text{Composé-Potentielle-Genotoxique}]}{dt} &= \frac{V_{m_{\text{Compose-Potentielle-Genotoxique}}} [\text{HONH-MeIQx}]}{K_{m_{\text{Compose-Potentielle-Genotoxique}}} + [\text{HONH-MeIQx}]} \end{aligned}$$

FIGURE 5.3 – Modèle réduit du métabolisme du MeIQx

Le modèle possède 8 variables et 16 paramètres. Deux coefficients de distribution  $0 \leq \alpha \leq 1$  et  $0 \leq \beta \leq 1$  modélise les quantités d'enzymes CYP1A2 et UGT impliquées dans les paires de réactions qu'elles catalysent. Ces coefficients de distribution peuvent varier en fonctions des conditions expérimentales.

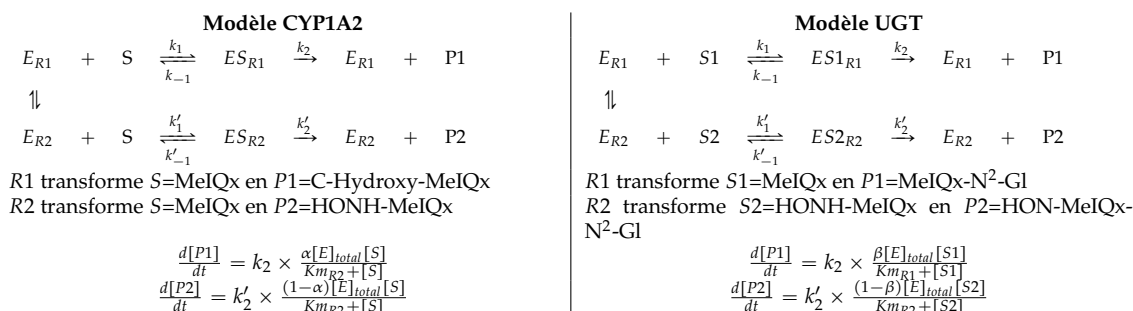


FIGURE 5.4 – Un modèle générique des compétitions enzymatiques au sein de réactions catalysées par la même enzyme

Le premier modèle modélise la transformation du même substrat (MeIQx) en deux produits différents par la même enzyme (CYP1A2). Le coefficient de distribution  $\alpha$  représente la répartition de l'enzyme CYP1A2 au sein des deux réactions qu'elle catalyse. L'enzyme UGT est quant à elle impliquée dans la catalyse de deux réactions en compétition l'une avec et l'autre et faisant intervenir deux substrats différents. La répartition enzymatique d'UGT au sein de ces deux réactions est modélisée par le coefficient de distribution  $\beta$ .

et UGT. Pour conclure, le modèle présenté dans la figure 5.3 rapporte la dynamique des variables de hauteur selon 16 paramètres fixes (constantes de Michaelis, vitesses maximales, CYP1A2 et UGT) et 2 paramètres supplémentaires ( $\alpha$ , et  $\beta$ ) qui peuvent varier en fonction des conditions expérimentales.

### 5.3 Evaluation des dynamiques enzymatiques

Pour évaluer la dynamique potentielle des distributions des enzymes CYP1A2 et UGT entre les deux voies métaboliques qu'elles catalysent, deux modèles différents ont été implémentés : un modèle de Michaelis-Menten classique testant l'hypothèse de saturation des voies métaboliques et un modèle dynamique dans lequel les enzymes se répartissent dynamiquement entre les deux voies qu'elles catalysent en fonction de la concentration de substrat. La différence principale entre ces deux hypothèses réside dans la dépendance vis-à-vis de la concentration du substrat des coefficients  $\alpha$  et  $\beta$  modélisant les distributions des enzymes CYP1A2 et UGT au sein des voies qu'elles catalysent.

- *L'hypothèse de saturation* suppose que les coefficients de distribution enzymatiques  $\alpha$  et  $\beta$  sont fixes et ne dépendent pas de la concentration du MeIQx. En d'autres termes, les quantités de CYP1A2 et d'UGT partagées au sein des voies qu'elles catalysent est constante pour toutes concentrations de MeIQx.
- *L'hypothèse de dose-dépendance* suggère que les coefficients de distribution enzymatiques  $\alpha$  et  $\beta$  soient dépendants de la concentration du MeIQx, tel que  $\alpha = \varphi(a_1, b_1, \theta_1, n_1, [MeIQx])$  et  $\beta = \varphi(a_2, b_2, \theta_2, n_2, [MeIQx])$ . Cette dépendance dynamique en fonction de la concentration du substrat est modélisée par l'utilisation d'une fonction sigmoïde  $\varphi(a, b, \theta, n, S) = a \frac{\theta^n}{\theta^n + S^n} + b$ , où  $a$  et

$b$  ( $0 \leq a, b \leq 1$ ) sont les bornes de la sigmoïde ;  $n = 5$  décrit la pente de la sigmoïde ;  $\theta$  est le point d'inflexion de cette courbe.

### 5.3.1 Ajustement de modèles, estimation des paramètres et filtration *a posteriori*

Les paramètres estimés par fitting ont été sélectionnés de manière à ce que toutes les prédictions du modèle ajustent de manière optimale l'ensemble des données présentées dans la figure 5.2.

A cet effet, on introduit dans le modèle les deux points de temps  $t_1 = 0$  et  $t_2 = 24h$  qui correspondent au temps d'acquisition des données par les expérimentateurs. Puis, on ajoute aussi les concentrations initiales des 3 expériences biologiques  $d_1 = 1\mu M$ ,  $d_2 = 10\mu M$  et  $d_3 = 50\mu M$ . On note également les concentrations des 8 composés du modèle réduit  $C_1, \dots, C_8$  présentés dans la figure 5.2, c'est-à-dire MeIQx, C-Hydroxy-MeIQx, MeIQx-N<sup>2</sup>-Gl, MeIQx-N<sup>2</sup>-SO<sub>3</sub>H, oxo-MeIQx, HONH-MeIQx, HON-MeIQx-N<sup>2</sup>-Gl et le Composé-Potentiellement-Génotoxique. Ainsi, on pose  $y_{k,l,i}^{obs}$  comme la concentration expérimentalement mesurée pour chaque composé  $C_k$  au temps  $t_l$  dans les hépatocytes, pour une concentration initiale de MeIQx  $d_i$ . Au final, l'intégration de ces données donne un total de 48 points de données expérimentales.

Tous les modèles décrivent les variations des 8 variables  $C_1, \dots, C_8$  en fonction du temps. Chaque modèle est paramétré par un ensemble de paramètres  $\theta$  incluant (i) les 16 paramètres fixes du modèle et (ii) les paramètres associés aux coefficients de distribution  $\alpha$  et  $\beta$ . La prédiction de la concentration d'un composé  $C_k$  au temps  $t_l$  dans les hépatocytes à une concentration initiale de MeIQx  $d_i$  est qualifiée par  $y_k(\theta, t_l, d_i)$ .

La toolbox PottersWheel [143] pour Matlab a été utilisé pour échantillonner tous les ensembles de paramètres possibles  $\theta$ . Pour ce faire, un ensemble de paramètres est tout d'abord aléatoirement sélectionné. Puis, l'algorithme "trust-region" implémenté dans Matlab modifie de manière itérative ce jeu de paramètres afin de minimiser la distance entre les données expérimentales et les données théoriques en utilisant la méthode des moindres carrés [183], où  $\sigma_{k,l,i}$  représentent l'erreur interne liée à l'estimation des paramètres.

$$\chi^2(\theta) = \sum_{k=1}^8 \sum_{l=1}^2 \sum_{i=1}^3 \left( \frac{y_{k,l,i}^{obs} - y_k(\theta, t_l, d_i)}{\sigma_{k,l,i}} \right)^2$$

Lorsque l'algorithme "trust-region" a atteint un minimum local, la toolbox PottersWheel calcule une  $p$ -value qui permet d'évaluer la qualité de l'ensemble des paramètres prédits. L'ensemble des paramètres a été considéré comme pertinent si la  $p$ -value était égale à 1 (cas le plus strict).

La valeur ajoutée de PottersWheel consiste à itérer cette procédure afin d'éviter les biais introduits par les minimums locaux et d'explorer l'espace complet d'un ensemble de paramètres. À cette fin, l'ancien ensemble de paramètres identifiés par PottersWheel est perturbé pour permettre l'exploration de nouvelles régions de l'espace de paramètres. L'algorithme "trust-region" a de nouveau été appliqué de façon itérative pour atteindre un autre minimum local correspondant à un autre ensemble de paramètres. Ce dernier est considéré comme pertinent lorsque sa  $p$ -value associée est égale à 1. Cette procédure a été itérée 1000 fois pour les hypothèses de saturation et de dose-dépendance. Pour améliorer les performances de l'algorithme, l'espace de recherche des paramètres est transposé dans un espace logarithmique. Toutes les bornes des paramètres utilisées pour exécuter l'ajustement de données par PottersWheel sont détaillées dans le tableau en annexe 5.1.

Nom du Paramètre	initial value	type	minValue	maxValue
k1	0.1	global	1e-01	10000
k2	0.1	global	1e-01	10000
k3	0.1	global	1e-15	10000
k4	0.1	global	1e-15	10000
$Vm_{[MeIQx-N^2-SO_3H]}$	0.1	global	1e-15	1e-6
$Vm_{oxo-MeIQx}$	0.1	global	1e-15	1e-6
$Vm_{Potential-Genotoxic-Compound}$	0.1	global	1e-15	1e-6
$Km_{IQx-8-COOH}$	0.1	global	1e-07	0.0001
$Km_{HONH-MeIQx}$	0.1	global	1e-07	0.0001
$Km_{MeIQx-N^2-GI}$	0.1	global	1e-7	1e-3
$Km_{[HONH-MeIQx-N^2-GI]}$	0.1	global	1e-7	1e-3
$Km_{[MeIQx-N^2-SO_3H]}$	0.1	global	1e-7	1e-3
$Km_{oxo-MeIQx}$	0.1	global	1e-7	1e-3
$Km_{Potential-Genotoxic-Compound}$	0.1	global	1e-7	1e-3
CYP	0.1	global	1e-18	1e-8
UGT	0.1	global	1e-18	1e-8
a1	0.1,	global	0	1
b1	0.1,	global	0	1
$\theta 1$	0.1	global	1e-9	1e-3
n1	5	fix	5	5
a2	0.1	global	0	1
b2	0.1	global	0	1
$\theta 2$	0.1	global	1e-9	1e-3
n2	5	fix	5	5

TABLE 5.1 – Espace de recherche des différents paramètres

Pour utiliser l'algorithme "trust region" lors de l'ajustement des données un espace de paramètres pour le modèle ODE doit être défini. Les intervalles de recherche de chaque paramètre sont spécifiés dans ce tableau.

Par conséquent, cette méthode permet de sélectionner une famille d'ensemble de paramètres qui minimise la distance entre les 48 points de données expérimentales  $y_k(\theta, t_l, d_i)$  et les points de données expérimentales prédits  $y_k(\theta, t_l, d_i)$  pour un jeu de paramètres  $\theta$ .

Dans un second temps, tous les ensembles de paramètres ayant une  $p$ -value égale

à 1 sont filtrés afin d'éliminer les faux positifs. Cette filtration est opérée par resimulation de tous ces jeux de paramètres sélectionnés en utilisant le même solveur "radau5 que PottersWheel configuré telle que : Absolute Tolerance= $1 \times 10^{-10}$  et Relative Tolerance= $1 \times 10^{-6}$ . Seuls les modèles avec les caractéristiques suivantes sont retenus : (i) la formation de toutes les espèces suit une dynamique monotone (ii) afin de limiter la variabilité des modèles sur chaque point de données, les espèces les plus variables sont contraintes (C-Hydroxy-MeIQx et [Detoxification]) pour une concentration initiale  $d_1 = 1\mu M$ , tel que :

$$\left| \frac{y_{\text{C-Hydroxy-MeIQx}}(\theta, t_2, d_1) - y_{\text{C-Hydroxy-MeIQx},2,1}^{obs}}{y_{\text{C-Hydroxy-MeIQx},2,1}^{obs}} \right| \leq 0.10,$$

$$-0.05 \leq \frac{[\text{Detoxification}](\theta, t_2, d_1) - [\text{Detoxification}]_{2,1}^{obs}}{[\text{Detoxification}]_{2,1}^{obs}} \leq 0.2;$$

où  $[\text{Detoxification}]_{2,1}^{obs}$  correspond à la valeur de la détoxification au temps  $t_2 = 24h$  pour une concentration initiale de  $d_1$ .

Pour finir, la voie de détoxification, la formation du "Composé-Potentiellement-Génotoxique" et par extension la bioactivation de MeIQx sont estimées par simulation des familles de modèles paramétrés (ayant répondu aux exigences ci-dessus pour des concentrations initiales de MeIQx de 0.05, 0.1, 0.5, 1, 2, 5, 10, 15, 30, 50, 75 et 100  $\mu M$  pendant 120 heures pour évaluer l'équilibre entre les voies de bioactivation et de détoxification.

### 5.3.2 La régiosélectivité est dépendante de la concentration de MeIQx

Afin d'étudier le métabolisme du MeIQx chez l'homme, un système d'équations différentielles ordinaires (ODE) a été construit en fonction des données issues d'hépatocytes humains HL-1 rapportés dans Langouët et al. [132]. Deux hypothèses ont été introduites pour modéliser le processus de compétition enzymatique en fonction de la concentration initiale du MeIQx et testées en recherchant les ensembles de paramètres qui expliquent le mieux les données.

La table 5.2 expose les scores d'ajustement optimaux associés à chaque métabolite, ainsi que le score local correspondant à la somme des scores d'ajustement optimaux des métabolites pour une concertation donnée. Enfin, le score global représente la somme des scores d'ajustement optimaux associés à chaque métabolite pour les 3 concentrations initiales de MeIQx. Chaque score d'ajustement optimal d'un composé donné est considéré comme valide lorsque la prédiction optimale du modèle à  $t_1 = 24$  varie dans l'intervalle d'erreur admissible prédit par PottersWheel. Pour une concentration initiale de MeIQx, une hypothèse est considérée comme valide lorsque tous les scores individuels sont valides ( $\text{Chi}^2 < 1$ ) et que la  $p$ -value *a posteriori* calculée est égale à 1.

	$\mu$ Mol	Meilleur score $\text{Chi}^2$ par séquence de 1000 fits										Score $\text{Chi}^2$ global	Best p-value	
		MetQx	MeIQx- N <sup>2</sup> -GI	MeIQx- N <sup>2</sup> -SO <sub>3</sub>	oxo- MeIQx	C- Hydroxy- MeIQx	HON- MeIQx- N <sup>2</sup> -GI	HONH- MeIQx	Potential- Genotoxic- Compound	Score $\text{Chi}^2$ local				
$\alpha$ et $\beta$ fixes	1	✓ (0.01)	✓ (0.82)	✓ (0.00)	✓ (0.00)	✓ (0.00)	✓ (0.01)	✓ (0.00)	✓ (0.01)	✓ (0.854)	✓ (0.01)	✓ (0.854)	<b>Hypothèse de saturation 1.</b> La saturation des voies métaboliques explique-t-elle les données ? Les coefficients de distribution de CYP1A2 ( $\alpha$ ) et UGT ( $\beta$ ) sont considérés comme global, c'est à dire qu'ils sont indépendants de la concentration initiale de MeIQx et ont la même valeur pour les 3 doses initiales de MeIQx	<b>X 0.893335</b>
	10	✓ (0.01)	X (12.98)	✓ (0.00)	X (1.88)	✓ (0.23)	✓ (0.00)	✓ (0.00)	✓ (0.39)	X (15.50)	X (4.610)	X 20.9622		
	50	✓ (0.04)	✓ (0.35)	✓ (0.00)	X (2.99)	✓ (0.02)	✓ (0.00)	✓ (0.00)	X (1.21)	X (4.610)	X (4.610)	X 20.9622		
$\alpha$ et $\beta$ dynamiques	1	✓ (0.84)	✓ (0.00)	✓ (0.01)	✓ (0.05)	✓ (0.02)	✓ (0.03)	✓ (0.00)	✓ (0.04)	✓ (0.987)	✓ (0.04)	✓ (0.987)	<b>Hypothèse de dose-dépendance 2.</b> Peut-on modéliser la dynamique avec une fonction directement dépendante de la concentration du substrat ? L'ajustement des données est effectué en considérant que les coefficients de distribution $\alpha$ et $\beta$ dépendent de la concentration de MeIQx. $\alpha$ et $\beta$ évoluent dynamiquement avec la concentration de MeIQx en fonction du temps. Ces paramètres sont modélisés par une fonction sigmoïde dépendante de la concentration de MeIQx.	<b>✓ 1</b>
	10	✓ (0.58)	✓ (0.03)	✓ (0.35)	✓ (0.30)	✓ (0.02)	✓ (0.07)	✓ (0.00)	✓ (0.01)	✓ (1.34)	✓ (0.01)	✓ (1.34)		
	50	✓ (0.04)	✓ (0.00)	✓ (0.00)	✓ (0.00)	✓ (0.00)	✓ (0.01)	✓ (0.00)	✓ (0.19)	X (0.248)	X (0.248)	X 2.57838		

TABLE 5.2 – Meilleurs scores d’ajustement des données obtenus avec les ODE exposées dans la figure 5.3 pour les hypothèse de saturation et de dose-dépendance

Ce tableau expose les meilleurs scores d’ajustement des données pour chaque composé métabolique. La  $p$ -value est un score calculé à posteriori par la toolbox PottersWheel afin d’évaluer la qualité des ajustements de données pour un ensemble de paramètres estimé. Les scores  $\text{Chi}^2$  locaux et globaux représentent plus précisément la qualité de l’ajustement en indiquant la minimisation par rapport aux données expérimentales (plus ce score se rapproche de 0, meilleur est l’ajustement). Un score  $\text{Chi}^2$  local est considéré comme valide (✓) lorsqu’il est inférieur à un et que sa valeur est localisée entre les barres d’erreurs calculées par PottersWheel au temps  $t = 24h$ . Un score  $\text{Chi}^2$  global est valide lorsque tous les scores locaux sont validés. Si ces critères ne sont pas respectés, alors le score associé est considéré comme invalide X et l’hypothèse associée est rejetée.



Comme le montre la table 5.2, l'*hypothèse de saturation* (ligne 1) ne peut pas être expliquée avec les familles de paramètres identifiées par la procédure d'ajustement des données. En effet, tous les modèles présentent une  $p$ -value inférieure à 1, ce qui est trop faible pour considérer cette hypothèse comme pertinente. De plus, 4 scores  $\text{Chi}^2$  sont supérieurs à 1 et en dehors de l'intervalle d'erreur admissible. On remarque que MeIQx-N<sup>2</sup>-Gl et C-Hydroxy-MeIQx sont particulièrement éloignés des données biologiques et affectent par conséquent la pertinence du modèle à expliquer les données expérimentales. En considérant les coefficients de distribution comme des paramètres fixes, et donc que la quantité d'enzymes impliquées dans chaque voie métabolique restent fixe, les données expérimentales ne peuvent pas être expliquées. L'*hypothèse de saturation* ne peut pas être conservée.

En revanche, les ensembles de paramètres prédits pour l'*hypothèse de dose-dépendance* permettent à la fois de reproduire et d'expliquer les données biologiques pour tous les métabolites mesurés dans les hépatocytes humains (table 5.2). La procédure de recherche de paramètres a abouti à une famille de 260 ensembles de paramètres (également appelés "fit") qui expliquent le mieux les données biologiques et ont tous une  $p$ -value égale à 1. Ces résultats suggèrent que la distribution des enzymes CYP1A2 et UGT entre leurs voies métaboliques qu'elles catalysent dépendent de la concentration de MeIQx et soulignent l'importance de l'aspect dynamique du système biologique. En outre, nos données démontrent que la modélisation d'une telle dépendance par une fonction sigmoïde est appropriée pour décrire la dynamique de processus biologiques.

### 5.3.3 73 ensembles de paramètres expliquent les données

La filtration des 260 ensembles de paramètres en fonction de critères biologiques définie dans la section 5.3.1 a abouti à une famille de 73 ensembles de paramètres qui expliquent tous de manière équivalente les données expérimentales.

On remarque que tous ces modèles possèdent des constantes d'affinité proches et que les réactions catalysées par l'enzyme CYP1A2 ont des  $K_m$  variant entre 2.21 et  $7\mu\text{M}$  pour  $K_{m_{\text{IQxCOOH}}}$  (C-Hydroxy-MeIQx) et entre 2.45 et  $4.12\mu\text{M}$  pour  $K_{m_{\text{HONH-MeIQx}}}$  (N-Hydroxy-MeIQx) (voir annexe figure 5.5). Cet intervalle de grandeur est en accord avec les paramètres observés chez l'humain [258, 98]. Par ailleurs, il faut noter que ces paramètres d'affinité pour les deux réactions catalysées par l'enzyme CYP1A2 sont proches les uns des autres. Cela suggère a priori que l'affinité pour la bioactivation et la détoxification catalysées par le CYP1A2 n'est pas un critère expliquant qu'un sentier métabolique peut être privilégié par rapport à un autre. En d'autres termes, le MeIQx peut tout autant se transformer en produit de bioactivation qu'en produit de détoxification. Ces réactions consistent à transférer un atome d'oxygène de l'Hème CYP1A2 vers le site de métabolisme du MeIQx. Ces réactions de transfert se produisent dans le site catalytique du CYP1A2 où réside l'Hème. Cela suggère que les affinités sont liées à la conformation privilégiée du MeIQx dans le site catalytique du CYP1A2. A

l'aide du Docking, les différentes conformations que peut prendre le MeIQx dans le site catalytique du CYP1A2 sont prédites. L'analyse des résultats de cette expérimentation montre que les conformations privilégiant le positionnement du carbone (C) ou de l'azote (N) près de l'Hème ont des scores quasi équivalents (Figure 5.6). Ce constat peut fournir une explication plausible quant aux valeurs des constantes d'affinité ( $K_m$ ) prédites et de l'effet catalyseur de l'enzyme CYP1A2 pour le MeIQx.

### 5.3.4 Validation du modèle

Pour valider le modèle, les 73 modèles sélectionnés ont été simulés et les profils métaboliques des deux composés majeurs de détoxification (C-Hydroxy-MeIQx et HONH-MeIQx-N<sup>2</sup>-Gl) ont été comparés avec les données issues de la littérature. Ces deux métabolites sont étudiés sur laps de temps réduit et à de faibles concentrations correspondant aux expositions humaines utilisées par Gu et al. et Stillwell et al. [76, 208]

Pour des concentrations initiales de MeIQx comprises entre 0.05 et 0.5  $\mu\text{M}$ , la formation du composé C-Hydroxy-MeIQx atteint un plateau (état stationnaire) avant 7 heures de traitement (Figure 5.7). On remarque que pour toute dose initiale inférieure à 0.5  $\mu\text{M}$  à 7h, la quantité prédite de C-Hydroxy-MeIQx varie entre 30% et 70% de la dose initiale. Cette observation est conforme aux données acquises par Gu et al. 5.1 précisant que le C-Hydroxy-MeIQx représentait de 35.8% à 73.1% de la dose de MeIQx mesurée dans l'urine 10h après ingestion de 920ng de MeIQx par des patients volontaires (voir table 5.1 en annexe).

Pour des concentrations initiales de MeIQx variant entre 0.5  $\mu\text{M}$  et 1.5  $\mu\text{M}$ , les quantités de HONH-MeIQx-N<sup>2</sup>-Gl prédites atteignent un plateau entre 7h et 12h de traitement. On remarque que la quantité prédite de HONH-MeIQx-N<sup>2</sup>-Gl varie entre 6 % et 16 % dans cet intervalle de temps. Ce constat est conforme aux données de Stillwell et al. déclarant que le HONH-MeIQx-N<sup>2</sup>-Gl représente de 2.2 à 17.1% de la dose de MeIQx mesurée dans l'urine à 12h pour une exposition comprise entre 1500 et 3000ng (correspondant à une exposition allant de 0.7 à 1.4  $\mu\text{M}$ ) (Figure 5.8).

Ces deux comparaisons avec des données biologiques réelles acquises sur humain fournissent donc une validation indirecte de l'approche de modélisation proposée. Elles mettent aussi en évidence le haut niveau de variabilité de la réponse du système expérimentalement mesurée, qui est ici correctement reflétée par le choix d'une famille de modèles ODE au lieu d'un seul.

## 5.4 Discussion et conclusion

Afin d'étudier de modéliser les comportements enzymatiques peu conventionnels de compétitions enzymatiques souvent caractérisées par une cinétique biphasique, un

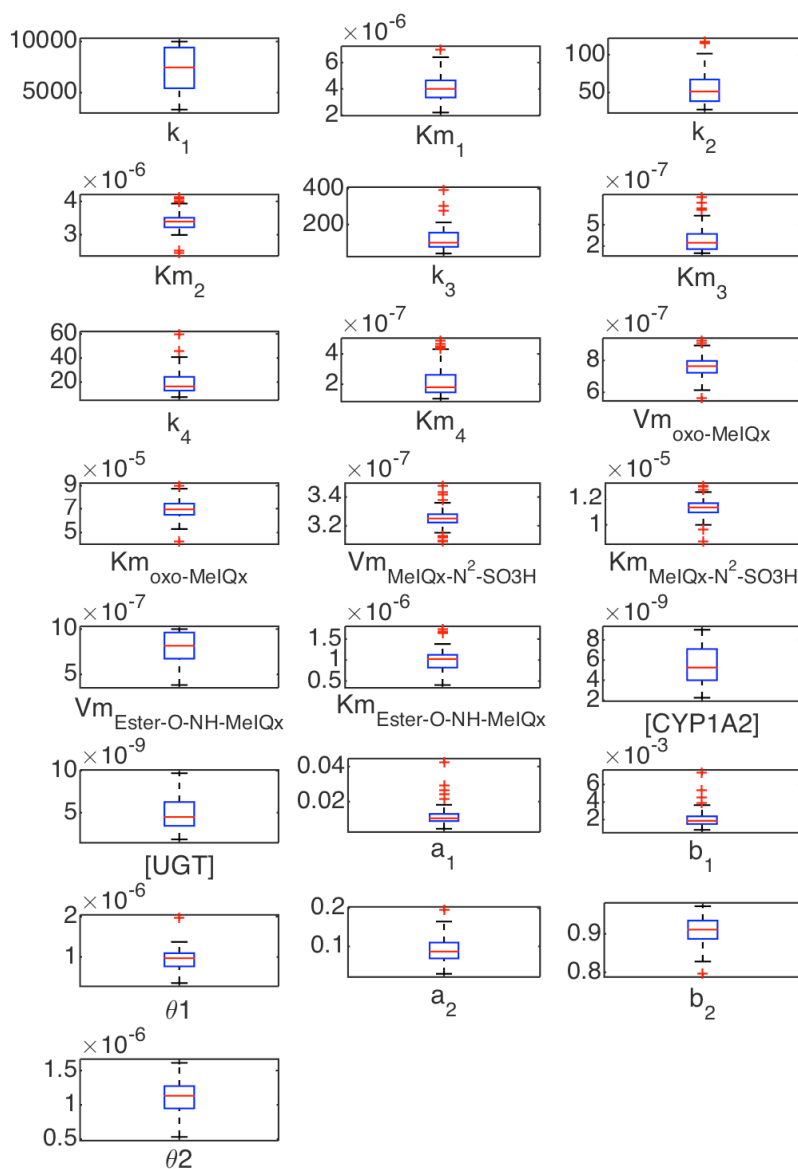


FIGURE 5.5 – Variation des paramètres estimés

$n_1$  et  $n_2$  sont les puissances des fonctions sigmoïdes associées respectivement aux coefficients de distribution de CYP1A2 et UGT et sont fixés à 5. Les  $Km$ ,  $\theta$  et les concentrations sont exprimés en Molaire,  $k$  en  $h^{-1}$  et  $Vm$  en  $M.h^{-1}$

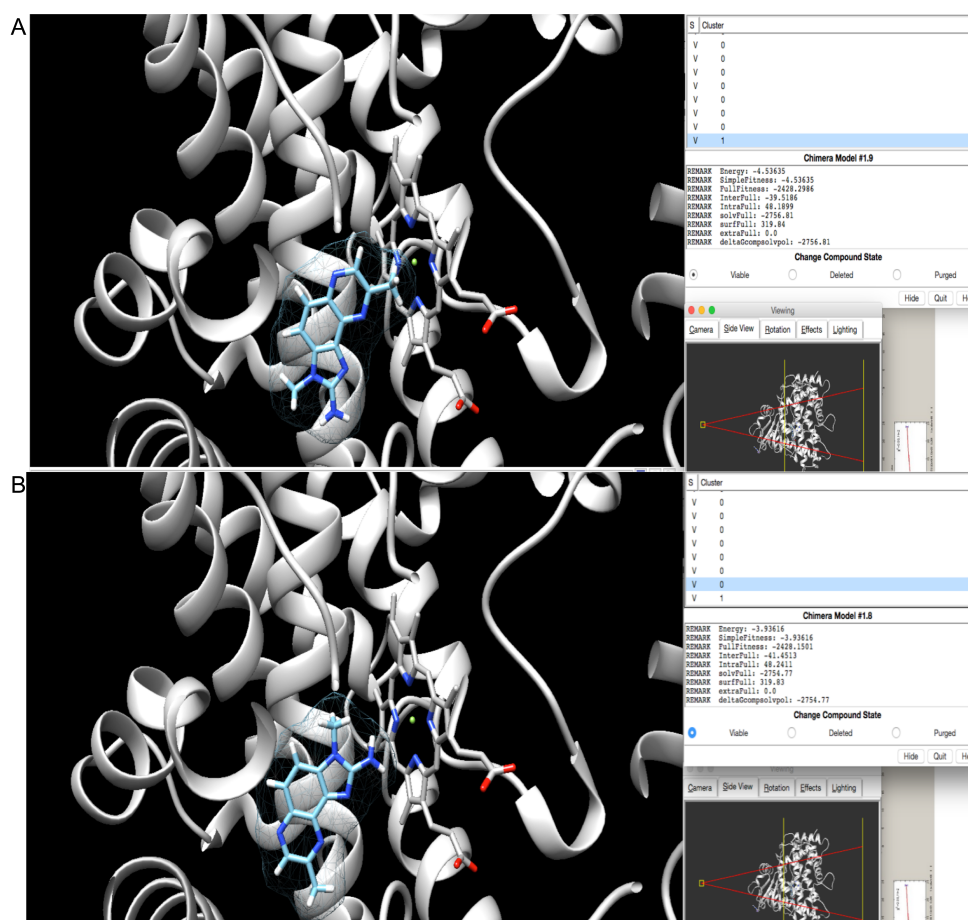


FIGURE 5.6 – Docking du MeIQx avec le CYP1A2 par l'utilisation du logiciel Swiss-Dock

La conformation calculée par SwissDock ayant le groupe CH<sub>3</sub> le plus proche de l'Hème est associée à une énergie de -4.53653 kcal/mol (A), alors que la seconde conformation possédant le groupe NH<sub>2</sub> au plus proche de l'Hème présente une énergie estimée à -3.93616 kcal/mol (B). Ces données indiquent que ces deux conformations sont proches et donc équiprobables même si la première conformation calculée avec le groupe CH<sub>3</sub> proche de l'Hème semble être légèrement privilégiée.

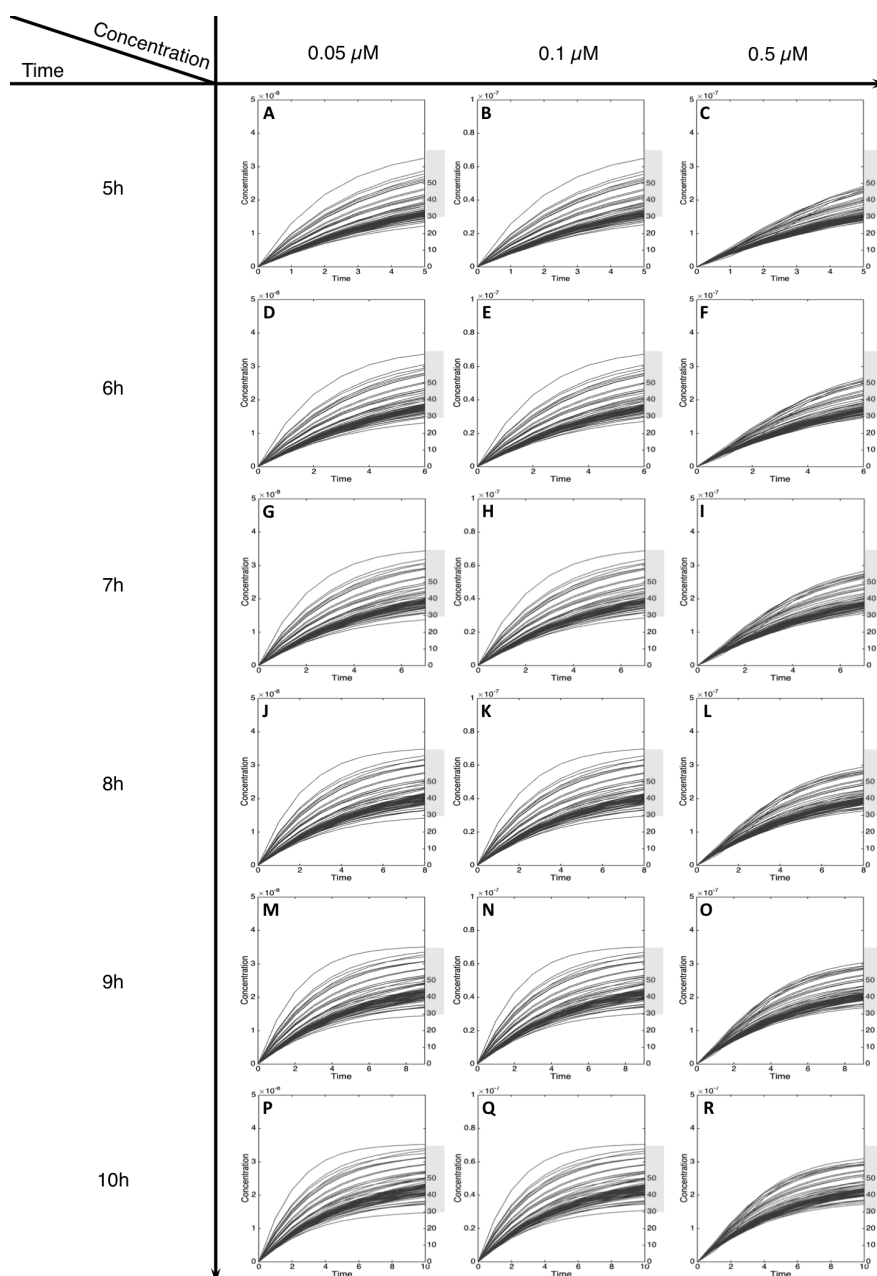


FIGURE 5.7 – Métabolisme du C-Hydroxy-MeIQx pour des concentrations initiales de MeIQx de 0.05, 0.1, 0.5  $\mu\text{M}$  après 5, 6, 7, 8, 9, 10h d'exposition

Le rectangle gris représente la zone comprise entre 30 et 70% du métabolisme du MeIQx. Les concentrations sont exprimées en Molaire sur l'axe des ordonnées et le temps est exprimé en heure sur l'axe des abscisses.

A-D-G-J-M-P : Formation de C-Hydroxy-MeIQx pour une concentration initiale de MeIQx de 0.05 $\mu\text{M}$ ; B-E-H-K-N-Q : Formation de C-Hydroxy-MeIQx pour une concentration initiale de MeIQx de 0.1 $\mu\text{M}$ ; C-F-I-L-O-R : Formation de C-Hydroxy-MeIQx pour une concentration initiale de MeIQx de 0.5 $\mu\text{M}$

	Relation d'équivalence		Patient											
	Total		S-001	S-003	S-005	S-008	S-009	S-012	S-013	S-014	S-015	S-020		
% Dose	100		56	64.8	68.7	59.4	54.1	66.2	65.4	35.8	71.9	73.1		
Mass ( $\mu\text{g}$ )	106.62		0.5152	0.5962	0.6320	0.5465	0.4977	0.6090	0.6017	0.3294	0.6615	0.6725		
Concentration molaire ( $\mu\text{M}$ )	50	0.4314	0.2416	0.2796	0.2964	0.2563	0.2334	0.2856	0.2822	0.1545	0.3102	0.3154		

TABLE 5.1 – **Equivalence des données de Gu's data en concentration molaire**

Les données prédites par les modèles sont comparées avec celles de Gu et al. [76]. Ils se sont attachés à mesurer la quantité restante de MeIQx, 8-CH<sub>2</sub>OH-IQx et IQx-8-COOH dans les urines humaines après 10h d'ingestion de viande de boeuf cuite contenant du MeIQx. Pour comparer les données prédites par les 73 modèles pour la voie métabolique C-hydroxy-MeIQx, les quantités mesurées par Gu de 8-CH<sub>2</sub>OH-IQx et IQx-8-COOH ont été additionnées. Ces dernières ont été converties en concentrations molaires par l'emploi de la relation liant la masse (m) avec la concentration molaire (C), la masse molaire (M) et le volume (V) :

$$m = C \times M \times V$$

En guise de références, les données suivantes ont été utilisées : la masse molaire du MeIQx (M) égale à 213.23854 g/mol est extraite de PubChem (Compound record, CID : 62275 [66]), une masse moyenne (m) de 920 ng de MeIQx est ingérée pour les patients de dans l'étude de Gu, un volume (V) de 10 mL correspond au volume moyen utilisé lors des expérimentations réalisées par Langouët et al., et une concentration (C) de 50 $\mu\text{M}$  est utilisée.

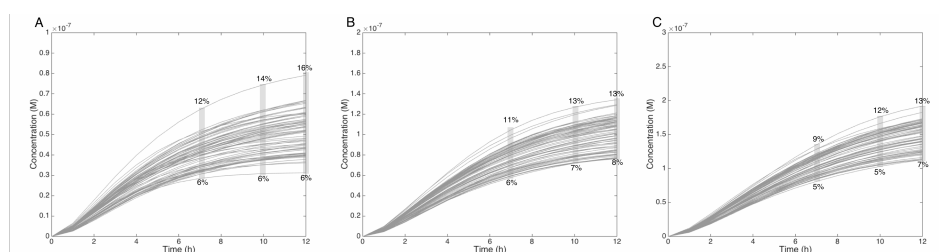


FIGURE 5.8 – Métabolisme du HON-MeIQx-N<sup>2</sup>-Gl pour des concentrations initiales de 0.5 (A), 1 (B), 1.5 (C)  $\mu\text{M}$  après 12h d'exposition

Le rectangle gris représente l'intervalle de variation du HON-MeIQx-N<sup>2</sup>-Gl pour les 73 modèles et est exprimé en pourcentage. Stillwell et al. ont étudié le pourcentage de MeIQx converti en HON-MeIQx-N<sup>2</sup>-Gl dans les urines après 12 pour des doses de MeIQx variant entre 1500 et 3000 ng (correspondant à des concentrations de 0.7 à 1.4 $\mu\text{M}$  en appliquant la même méthode basée sur une relation d'équivalence et proposée lors de l'exploitation des données de Gu en annexe). Ce pourcentage varie entre 2.2 à 17.1% du MeIQx ingéré. En simulant la formation du HON-MeIQx-N<sup>2</sup>-Glc pour des concentrations de MeIQx égales à 0.5, 1 et 1.5 $\mu\text{M}$ , l'intervalle maximum de conversion du MeIQx en N-OH-MeIQx-N<sup>2</sup>-Glc est compris entre 6 et 12% après 7h d'exposition et entre 6 et 16% après 12h d'exposition. Les concentrations sont exprimées en Molaire sur l'axe des ordonnées et le temps est exprimé en heure sur l'axe des abscisses.

modèle mathématique a été développée basée sur l'exploitation de données métaboliques restreintes et acquises dans des hépatocytes humains pour la molécule MeIQx.

La principale originalité de cette approche de modélisation repose sur le fait que contrairement aux approches classiques qui sélectionnent le jeu de paramètres qui ajuste au mieux les données, il a été ici choisi d'utiliser un ensemble de jeux de paramètres qui expliquent les données. Comme indiqué dans [237, 28, 79], le nombre de jeux de données et de points temporels sont insuffisants par rapport au nombre de paramètres pour les estimer avec précision. Par conséquent, pour gagner en robustesse et pour limiter le biais dans le processus de sélection de paramètres [125], tous les modèles paramétrés avec une *p*value égale à 1 dans Potterswheel [143] ont été sélectionnés. Cette *p*value est un score statistique mis en place dans Potterswheel pour déterminer si un modèle doit être rejeté ou non [143].

De plus, le modèle (consistant en une famille de 73 ODE paramétrées) peut prédire l'équilibre entre la bioactivation et la désintoxication de MeIQx dans les hépatocytes humains à une large gamme de concentrations. Il est intéressant de noter que la prédiction du métabolisme du MeIQx est plus robuste pour des concentrations élevées plutôt qu'à de faibles concentrations, où les courbes illustrant la variation du ratio bioactivation/détoxification présentent une dispersion plus large. La dispersion à faible concentration s'explique par l'importante variabilité des scores d'ajustements locaux

au sein de tous les modèles paramétrés. Soulignons aussi qu'une telle dispersion est cohérente avec les résultats obtenus dans la littérature par Gu et al et Stillwell et al. [76, 208] qui se sont attachés à illustrer la variabilité du métabolisme du C-hydroxy-MeIQx et du HON-MeIQx-N<sup>2</sup>-GI à de faibles concentrations.

En conclusion, cette étude a prouvé qu'il était possible à la fois de modéliser simplement un mécanisme complexe de compétition enzymatique régissant l'équilibre des voies métaboliques et à la fois d'implémenter un modèle dynamique exploitant un nombre limité de données biologiques. L'implémentation d'un modèle sous-déterminé est possible par une sélection d'une famille de modèles décrivant la dynamique d'un système biologique. La considération d'une famille contenant un ensemble de paramètres permet ainsi de limiter les biais inhérents à la faible quantité de données.

## 5.5 Résumé

La modélisation dynamique de phénomènes biologiques complexes peut se heurter à deux problèmes majeurs : la quantité de données biologiques limitées inhérentes à la complexité biologique rendant difficile leur acquisition et la difficulté à implémenter un modèle simple modélisant ces processus complexes.

Ce travail a donc consisté à l'implémentation d'une méthode permettant d'exploiter des données biologiques restreintes à travers un modèle sous-déterminé. Contrairement aux approches classiques qui sélectionnent le jeu de paramètres ajustant au mieux les données expérimentales, la méthodologie employée consiste à étalonner le modèle en considérant un ensemble de jeux de paramètres expliquant les données. Les avantages de prendre une famille de paramètres sont la réduction des biais possibles liés aux données limitées et la faculté à pouvoir modéliser la dynamique d'un système biologique.

Cette étude a aussi permis de mettre en place un modèle mathématique simple permettant de modéliser les compétitions enzymatiques régulées par une sigmoïde. Ce dernier permet aussi de prendre en compte la dynamique d'un système biologique affectant l'équilibre des voies métaboliques par l'emploi d'une sigmoïde qui permet de le réguler en fonction de la concentration d'une substance chimique. Ainsi, ce modèle pourrait être aussi aisément utilisé pour la modélisation des nombreuses cinétiques biphasiques constatées dans la littérature et ouvre donc à de nouvelles perspectives pour appréhender un mécanisme complexe.





# Chapitre 6

## Cas d'étude : Application au métabolisme du MeIQx

Ce chapitre présentera l'exploitation de la méthodologie précédemment décrite dans le cadre de l'évaluation de la toxicité d'un contaminant alimentaire : le MeIQx. Le modèle sera utilisé pour évaluer l'équilibre entre les voies de bioactivation conduisant à la génération de métabolites toxiques et celles de détoxification. Par ailleurs, l'objectif est aussi de comprendre l'effet des mécanismes de compétition sur la toxicité et d'identifier les acteurs majeurs du système affectant l'équilibre des voies métaboliques médiant la toxicité.

---

<b>6.1</b>	<b>Introduction</b>	<b>115</b>
<b>6.2</b>	<b>Influence de la biphasicité sur la réponse biologique</b>	<b>116</b>
6.2.1	Effet de biphasicité sur l'équilibre des voies métaboliques	116
6.2.2	La régulation des voies métaboliques opère à faible concentration de MeIQx	117
6.2.3	Preuve de l'effet biphasique et évaluation de la génotoxicité	120
6.2.4	Influence de la régulation du CYP1A2 sur l'équilibre des voies métaboliques	120
<b>6.3</b>	<b>Discussion et conclusion</b>	<b>122</b>
<b>6.4</b>	<b>Résumé</b>	<b>126</b>

---

### 6.1 Introduction

Les amines aromatiques hétérocycliques (AHA) sont des contaminants environnementaux et alimentaires classés comme cancérogènes humains probables ou possibles [2]. Parmi les AHA, le MeIQx (2-Amino-3,8-diméthylimidazo [4,5-f] quinoxaline) est l'une des plus abondantes dans les aliments cuits [161]. Le métabolisme de MeIQx a été étudié dans les hépatocytes humains et il a été clairement démontré que le MeIQx

est activement transformé en métabolites soit non toxiques (détoxification), ou soit en substances potentiellement toxiques (réactifs vis-à-vis de l'ADN) [132]. Néanmoins, les études existantes sur les AHA utilisant des modèles biologiques humains pour étudier leurs biotransformations sont très limitées en raison de leurs difficultés à mettre en place [234, 133]. Les quelques expérimentations existantes sur les cellules humaines ont mis en évidence une variabilité très importante de la transformation du MeIQx en composés toxiques et non toxiques parmi les patients traités [76]. Cette limitation motive fortement la nécessité d'approches prédictives pour éclairer les processus biologiques complexes [209].

La méthodologie développée dans le chapitre 5 a été employée pour prédire l'équilibre entre les voies de bioactivation (c'est-à-dire la voie toxique de formation d'adduits à l'ADN) et celle de détoxification du MeIQx en fonction de sa concentration initiale. A cet effet, la famille de 73 modèles numériques sélectionnée et rendant possible la prédiction de la bioactivation et de la détoxification pour toute concentration initiale de MeIQx et à tout temps  $t$  a été utilisée pour simuler l'équilibre des voies métaboliques affectant la toxicité.

Curieusement, cette approche appliquée au MeIQx suggère que la variabilité résultante s'explique par des processus de compétitions enzymatiques. Ces phénomènes de compétitions enzymatiques sont à l'origine d'une évolution non linéaire de l'équilibre entre les voies de bioactivation et de détoxification qui présente un maximum de bioactivation par rapport à la détoxification pour une concentration correspondant à  $10\mu\text{M}$  de MeIQx dans le modèle de cellules hépatocytaires. Cette évolution biphasique est similaire à un comportement déjà observé dans l'exposition au bisphénol A, le propranolol, l'ibuprofène et d'autres composés toxicologiques. Cela suggère que la compétition enzymatique est le contrôleur clé de la réponse non monotone et très variable du système. De plus, les investigations approfondies de ce mécanisme sur l'équilibre entre la bioactivation et la détoxification montrent clairement que son effet est à l'origine de la régulation de la bioactivation et que les études impliquant ce genre de mécanismes mériteraient d'être approfondies.

## 6.2 Influence de la biphasicité sur la réponse biologique

### 6.2.1 Effet de biphasicité sur l'équilibre des voies métaboliques

Comme illustré dans la figure 5.3, le métabolisme global de MeIQx peut être décomposé en voie de détoxification et en voie de bioactivation. La voie de détoxification inclut le C-hydroxy-MeIQx ainsi que tous les conjugués de phase II (représentés verts de la figure 6.1). En revanche, la bioactivation correspond à la formation des composés potentiellement génotoxiques, c'est-à-dire l'Ester-O-NH-MeIQx ainsi que tous les métabolites potentiellement non identifiés (représentés en rouge dans la Figure 6.1). Afin d'évaluer la bioactivation de MeIQx en fonction de sa concentration initiale, les 73 modèles mathématiques précédemment sélectionnés et associés à leurs paramètres

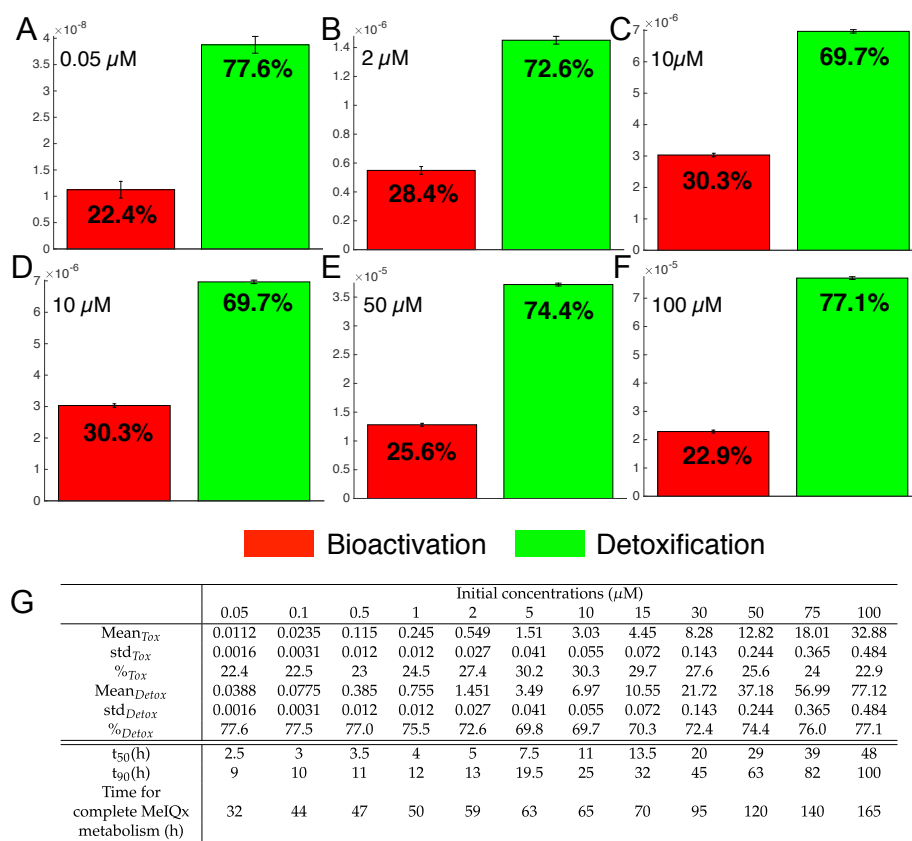
estimés ont été utilisés pour simuler les voies de détoxification et de bioactivation du MeIQx sur une échelle de temps variant de 0 à 180h. Ce dernier temps correspond au moment où tous les modèles ont atteint un état stable.

La distribution de MeIQx à l'état d'équilibre en produits de bioactivation et de désintoxication exprimés en pourcentage (Figure 6.1A) ou en concentrations (Figure 6.1B,  $\mu\text{M}$ ) est illustrée dans la figure 6.1. Les résultats montrent clairement que la voie de détoxification prédomine toujours malgré le changement de l'équilibre entre les voies de bioactivation et détoxification en fonction de la concentration du MeIQx. Plus précisément, le pourcentage du MeIQx impliqué au sein de la voie bioactivation augmente sur une gamme de concentrations de MeIQx située entre 0.05  $\mu\text{M}$  et 10  $\mu\text{M}$ . En revanche, pour des concentrations supérieures à 10  $\mu\text{M}$ , l'équilibre des voies de bioactivation et détoxification diminue progressivement en faveur de la détoxification.

De plus, comme illustré dans la figure 6.1B, la durée requise pour transformer 50% du MeIQx correspond de 10% à 25% du temps nécessaire pour dégrader complètement le MeIQx. Par exemple, pour une dose initiale de 0,05  $\mu\text{M}$  de MeIQx, le métabolisme complet prend 32h, alors que seulement 2.5h sont nécessaires pour métaboliser 50% de la dose de MeIQx et 9 h pour métaboliser 90% de la dose de MeIQx (Figure 6.1B).

## 6.2.2 La régulation des voies métaboliques opère à faible concentration de MeIQx

Dans un but d'approfondir la régulation biphasique régulant l'équilibre entre la bioactivation et la détoxification, les variations des coefficients de distribution du CYP1A2  $\alpha$  (Figure 6.2A) et des coefficients de distribution d'UGT  $\beta$  (Figure 6.2B) en fonction à la concentration du MeIQx ont été dessinées. Pour rappel, tous les 73 modèles numériques paramétrés qui expliquent au mieux le métabolisme du MeIQx ont été construits en introduisant un coefficient de distribution dépendant du MeIQx pour CYP1A2 et UGT modélisés chacun à l'aide une fonction sigmoïde. L'évolution de cette dépendance en fonction de la concentration du MeIQx est représentée dans la figure 6.2. Les simulations suggèrent que le coefficient de distribution du CYP1A2 varie entre 0.005 pour des concentrations élevées de MeIQx à 0.05 pour de concentrations faibles de MeIQx. En raison de ces faibles valeurs modulant la formation de C-Hydroxy-MeIQx ( $\alpha$  pondère la formation du C-Hydroxy-MeIQx alors que  $1 - \alpha$  est introduit pour la bioactivation), cela suggère que la distribution du CYP1A2 favorise toujours la voie de bioactivation. Néanmoins, pour des concentrations de MeIQx inférieures à 0.5  $\mu\text{M}$ , la quantité relative de CYP1A2 impliquée dans la détoxification est multipliée par un facteur 10. Une interprétation possible peut être cette variation de la quantité de CYP1A2 impliquée dans la voie de détoxification, même si celle-ci reste faible, elle suffit au système pour favoriser la voie de détoxification bien que celle-ci soit cinétiquement défavorisée. Ceci est conforme aux mesures expérimentales des activités



**FIGURE 6.1 – Biotransformation du MeIQx en produits de bioactivation (rouge) et en produits de détoxification (vert) pour des concentrations initiales différentes de MeIQx**

73 modèles ont été simulés en utilisant 12 concentrations initiales différentes (0.05, 0.1, 0.5, 1, 2, 5, 10, 15, 30, 50, 75, 100  $\mu\text{M}$ ) après 180h (temps nécessaire pour atteindre l'état d'équilibre pour toutes les concentrations initiales) d'exposition au MeIQx. Chaque histogramme illustre l'équilibre entre les produits de bioactivation et ceux de détoxification qui favorise progressivement la bioactivation pour des concentrations de MeIQx comprises entre 0.05 et 10  $\mu\text{M}$  (A-B-C), alors que la détoxification reprend le dessus pour des concentrations comprises entre 10 et 50  $\mu\text{M}$  (D-E-F). Le poids de chaque produit dans cet équilibre est exprimé en pourcentage. Le tableau expose les concentrations en  $\mu\text{M}$  des produits de bioactivation et de détoxification pour les 73 modèles simulés (G). Les scores correspondants sont calculés pour les 73 modèles : la moyenne, l'écart type (std) et le pourcentage de MeIQx transformé soit en produits de bioactivation ou soit en produits de détoxification. De plus,  $t_{50}$  expose le temps nécessaire pour que 50% du MeIQx soit métabolisé et  $t_{90}$ , le temps nécessaire pour que 90% du MeIQx soit consommé. Le temps nécessaire pour que le MeIQx soit intégralement transformé est aussi indiqué.

enzymatiques par Turesky et al. [231] démontrant que la cinétique de la voie de détoxification est plus lente que celle de bioactivation.

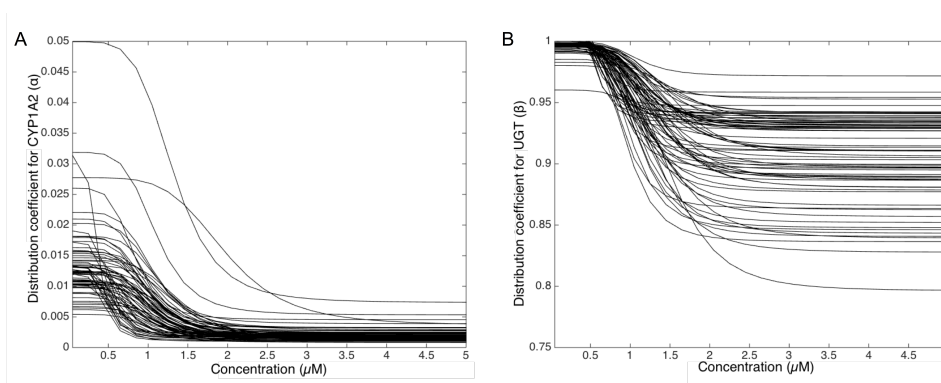


FIGURE 6.2 – Variation du coefficient de distribution  $\alpha$  du CYP1A2 (A) et du coefficient de distribution  $\beta$  d'UGT (B) modélisés par une fonction sigmoïde dépendante de la concentration du MeIQx

Les 73 modèles numériques paramétrisés qui expliquent le mieux les données de transformation du MeIQx en différents métabolites ont été construits en introduisant deux coefficients de distribution pour CYP et UGT qui sont dépendant de la concentration de MeIQx et modélisés par une fonction sigmoïde. L'évolution de cette dépendance en fonction de la concentration du MeIQx pour les deux enzymes est dessinée.

Ces simulations suggèrent aussi par l'analyse du coefficient de distribution d'UGT que la transformation directe du MeIQx en composés glucuronidés est fortement favorisée par rapport à la transformation des composés intermédiaires HONH-MeIQx en métabolites glucuronidés. Cependant, on remarque un effet levier de cette régulation à forte concentration de MeIQx qui se produit au point d'inflexion situé aux alentours de 1,25  $\mu\text{M}$ . Ce changement suggère que le système se mobilise pour transformer les métabolites intermédiaires toxiques HONH-MeIQx lorsque leur concentration est trop élevée et d'assurer au maximum la détoxification globale du MeIQx.

Pour résumer cette régulation des voies métaboliques, le scénario suivant est suggéré. Tout d'abord, pour une dose initiale de MeIQx supérieure à 2  $\mu\text{M}$ , le CYP1A2 est principalement impliqué dans la catalyse de la voie de bioactivation tandis que l'UGT est impliqué dans la transformation des composés intermédiaires HONH-MeIQx. Puis, lorsque la concentration du MeIQx atteint un seuil aux alentours de 1,25  $\mu\text{M}$ , l'UGT est redistribué pour favoriser une transformation directe de MeIQx en composés glucuroniques. Enfin, lorsque la concentration de MeIQx atteint environ 0,75  $\mu\text{M}$ , le CYP1A2 favorise la catalyse de la voie de détoxification qui est néanmoins cinétiquement défavorisée. Bien que les seuils (points d'inflexion) soient clairement dépendants du patient, il convient de noter que ces phénomènes se produisent dans une gamme de

concentration de MeIQx correspondant à une exposition réelle potentielle et influe directement sur le ratio bioactivation/détoxification où un "creux" peut être observé aux alentours des points d'inflexion des sigmoïdes (figure 6.3A).

### 6.2.3 Preuve de l'effet biphasique et évaluation de la génotoxicité

Pour évaluer l'effet de la biphasicité sur la génotoxicité, l'évolution du ratio entre la bioactivation et la détoxification en fonction du temps et de la concentration de MeIQx a été illustrée dans la figure 6.3. La figure 6.3A illustre l'évolution du ratio bioactivation/détoxification pour des concentrations de MeIQx comprises entre 0.05 et 100  $\mu\text{M}$  sur une plage de temps de 180h. Cette figure est complétée par 4 coupes illustrant l'évolution du ratio bioactivation/détoxification à 6h, 24h, 72h et 120h (figure 6.3B). Tout d'abord, pour n'importe quel point de temps, le rapport entre les voies de bioactivation et de détoxification n'est pas linéaire, mais dépend principalement de la concentration avec une concentration maximale autour de 10  $\mu\text{M}$ . Cette observation confirme celle de la figure 6.1. Par ailleurs, bien que les profils de courbes et leurs valeurs associées soient assez similaires, des variations subtiles peuvent être observées. Par exemple, on peut constater que la forme du pic est plus aiguë à 6h qu'à 120h (figure 6.3B). Ces simulations illustrent aussi que la génotoxicité est fonction aussi bien de la concentration de MeIQx que du temps de métabolisme (figure 6.3A). En effet on remarque qu'outre le pic à 10 $\mu\text{M}$ , ce ratio est corrélé positivement au temps. En s'attachant à la plage de concentration où agissent les sigmoïdes régulant la distribution des enzymes CYP1A2 et UGT (c'est à dire aux alentours de 1 $\mu\text{M}$ ), on remarque un "creux" dans l'évolution du ratio sur l'échelle des concentrations. Ce constat montre l'adaptation du système au changement de préférences des voies métaboliques qui influe directement sur le ratio et la génotoxicité. Enfin, en raison de la variabilité des prédictions très faibles pour des concentrations supérieures à 100  $\mu\text{M}$ , on peut suggérer que la famille des 73 modèles d'ODE suit des comportements similaires à ceux illustrés dans la présente figure.

### 6.2.4 Influence de la régulation du CYP1A2 sur l'équilibre des voies métaboliques

A l'aide des fonctions implémentées au sein de PottersWheel, une analyse de sensibilité 2D pour étudier l'influence de chaque paramètre sur le système a été réalisée. Comme le montre la figure 6.4, cette analyse illustre clairement que la concentration totale de CYP1A2 est un acteur majeur de la formation des produits de bioactivation. Afin d'illustrer et approfondir l'influence de ce paramètre, la valeur du paramètre de la concentration totale de CYP1A2 a été modifiée dans les 73 modèles (multiplication par 10 et division par 10). Les modèles perturbés ont été simulés afin de calculer l'équilibre entre les voies de bioactivation et de détoxification. Les résultats de ces simulations sont représentés dans la figure 6.5. Lorsque la concentration du CYP1A2 est diminuée d'un facteur 10, la voie de détoxification prédomine très majoritairement, ré-

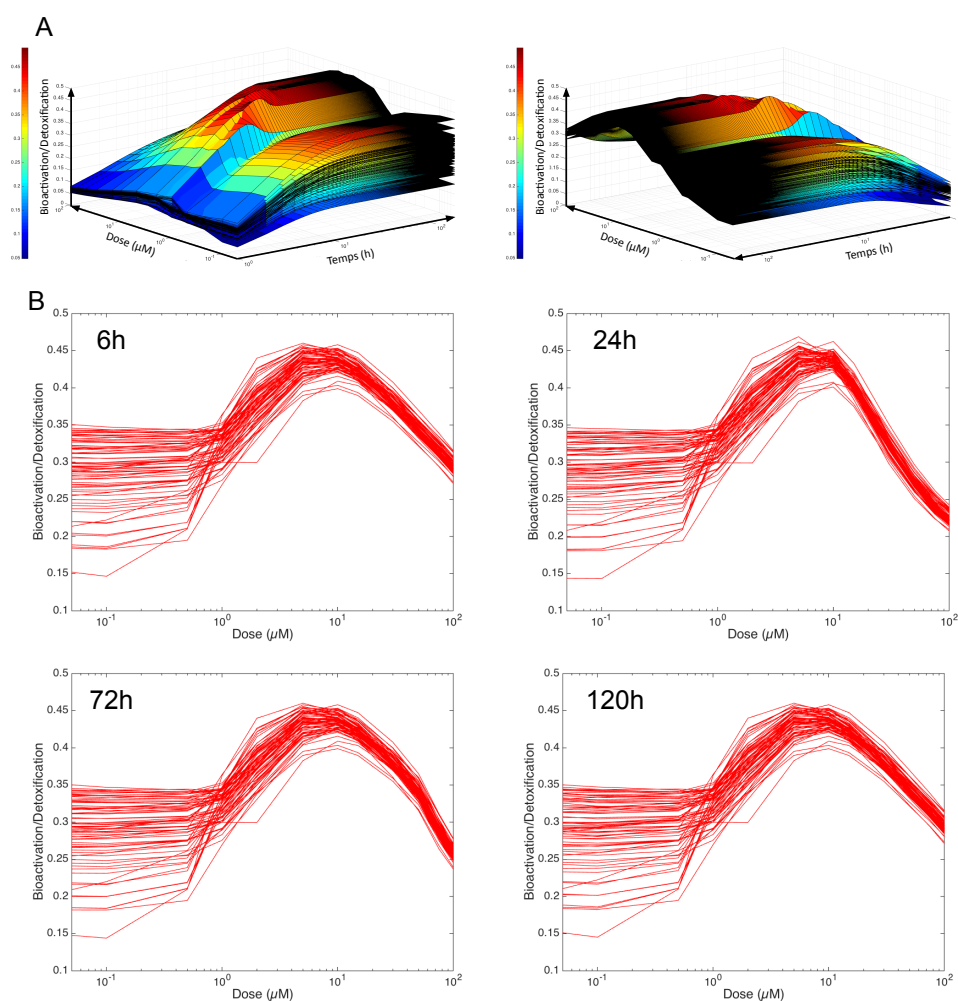


FIGURE 6.3 – Evolution du ratio Bioactivation/Detoxification en fonction de la concentration de MeIQx

Le ratio est ici exposé pour les 73 modèles simulés avec 12 concentrations initiales différentes de MeIQx (0.05, 0.1, 0.5, 1, 2, 5, 10, 15, 30, 50, 75, 100  $\mu\text{M}$ ). La partie A illustre l'évolution du ratio (axe z) en fonction du temps en heure (axe x) et de la concentration en  $\mu\text{M}$  (axe y). Les axes x et y sont exprimés en logarithme afin de mieux évaluer l'effet du ratio. La partie B de la figure représente des coupes à 180° des figures 3D après 6h, 24h, 72h et 120h d'exposition au MeIQx.



duisant l'importance de la voie de bioactivation qui représente dans ce scénario moins de 10% du métabolisme total. Cependant, le métabolisme du MeIQx complet n'est pas atteint à 180h pour une concentration de MeIQx égale à  $100\mu\text{M}$ . Au contraire, l'augmentation de la concentration du CYP1A2 d'un facteur 10 provoque une perturbation de l'équilibre entre les voies de bioactivation et celles de détoxification en faveur de la bioactivation pour des concentrations élevées de MeIQx (supérieures à  $10\mu\text{M}$ ). À faible concentration, la détoxification prédomine encore malgré l'augmentation de l'importance de la voie de bioactivation. Cela suggère que le CYP1A2 est un acteur clé dans la régulation de l'équilibre entre la bioactivation et la détoxification, et que sa concentration et son mode de fonctionnement sont essentiels pour moduler cet équilibre.

### 6.3 Discussion et conclusion

Afin d'étudier le comportement d'un contaminant alimentaire qu'est le MeIQx dans les hépatocytes, le modèle mathématique développé dans le chapitre 5 a été utilisé pour évaluer l'équilibre entre la bioactivation et la détoxification en fonction de la concentration de MeIQx et des temps d'exposition.

Ce travail montre que l'évolution du ratio de bioactivation/détoxification suit une tendance non monotone, où la toxicité du MeIQx est à son maximum pour les concentrations aux alentours de  $10\mu\text{M}$ . La variation abrupte de ce ratio suggère que des mécanismes de régulation des voies de bioactivation ou de détoxification pourraient avoir lieu dans les hépatocytes. Une première hypothèse plausible pourrait être expliquée par la saturation dans un premier temps de la voie de détoxification qui entraîne une redistribution du CYP1A2 vers la voie de bioactivation. Cette hypothèse a été testée en effectuant une procédure d'inférence de paramètres, en supposant la constance des paramètres  $\alpha$  et  $\beta$  qui représentent la distribution de CYP1A2 et UGT. Comme illustré dans le tableau 5.2, les ajustements de paramètres les plus adaptés lors du test de cette hypothèse sont incapables de bien reproduire données biologiques. Cela suggère que la distribution du CYP1A2 entre les deux voies métaboliques n'a pas seulement résulté d'effets de saturation. Comme l'ajustement du modèle nécessite une distribution dynamique des coefficients ( $\alpha$  et  $\beta$ ) pour CYP1A2 et UGT, on peut supposer que des mécanismes complexes peuvent moduler les activités enzymatiques et affecter la régiosélectivité de CYP1A2 pour MeIQx. Les cinétiques biphasiques ne sont pas nouvelles et ont déjà par exemple été expérimentalement découvertes dans le cas du métabolisme bisphénol A dans des hépatocytes de rat où la cinétique d'UGT suivait une évolution biphasique impliquant deux valeurs  $V_m$  distinctes [181]. Au regard de cette observation expérimentale, l'introduction d'une fonction sigmoïde dans l'équation de Michaelis Menten semble être une approche appropriée pour modéliser la dynamique du CYP1A2 et d'UGT, ainsi que leur dépendance vis-à-vis de la concentration du substrat.

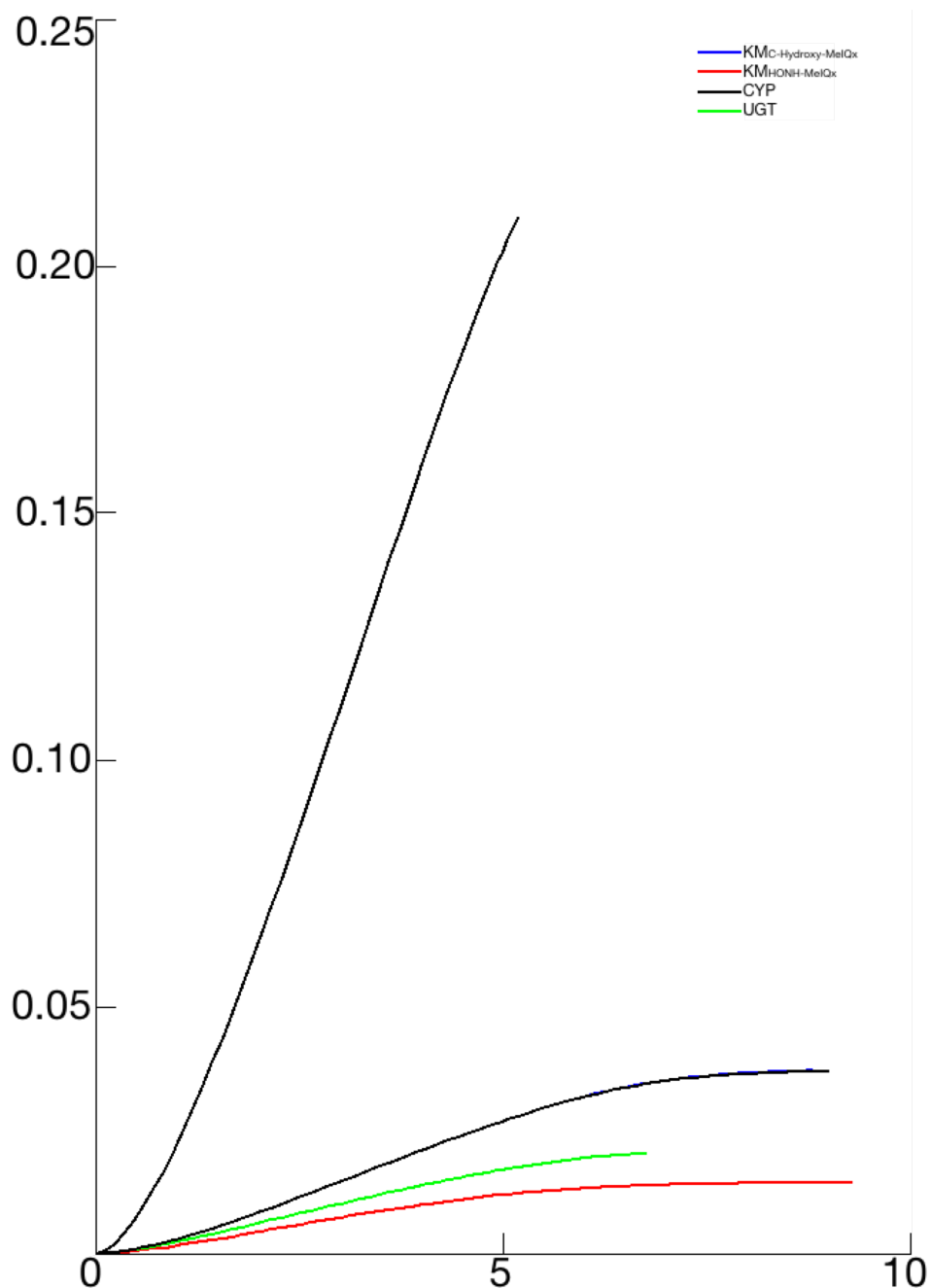
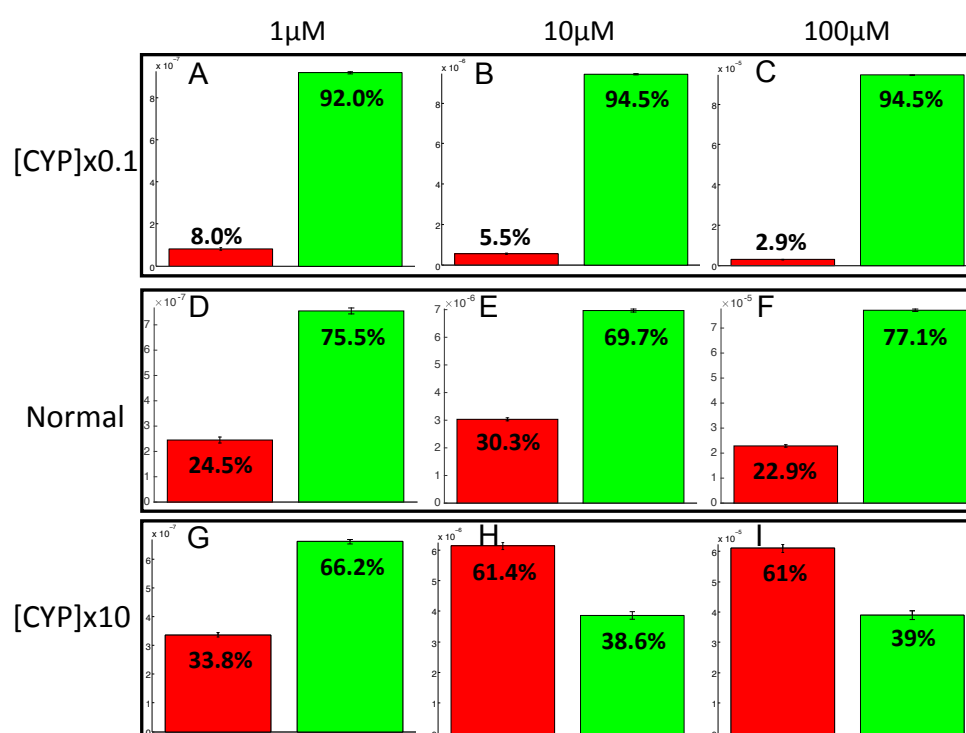


FIGURE 6.4 – Analyse sensible 2D réalisée à l'aide de PottersWheel

Tous les paramètres ont été perturbés par un facteur 10 lors de l'analyse sensible 2D afin d'étudier l'effet sur la bioactivation par la formation des Composés-Potentiellement-Génotoxiques. Cette figure expose seulement les paramètres les plus sensibles et compare l'influence de CYP et UGT sur la réponse du système.



**FIGURE 6.5 – Evaluation de la formation des produits de bioactivation et de détoxification en fonction des changements de concentrations totales de CYP1A2**

Les 73 modèles ont été simulés pour trois doses de MeIQx (1 (A, D, E), 10 (B, E, H) et 100 (C, F, I)  $\mu\text{M}$ ) pendant 180h d'exposition au MeIQx. Les simulations sont réalisées dans trois conditions expérimentales différentes : une concentration normale de CYP1A2 où les concentrations totales de CYP1A2 correspondent à celles des 73 modèles sélectionnés (D, E, F) ; une condition de faible concentration de CYP1A2 où les concentrations de CYP1A2 pour les 73 modèles sont diminuées d'un facteur 10 (A, B, C) ; et une condition de forte concentration de CYP1A2 où les concentrations de CYP1A2 pour les 73 modèles sont augmentées d'un facteur 10 (G, H, I). Après 180h, les moyennes des produits de bioactivation et détoxification sont calculées pour les 73 modèles.

En guise de confirmation du rôle du CYP1A2 dans l'équilibre entre la bioactivation et la détoxification, plusieurs simulations basées sur des perturbations de la concentration du CYP1A2 ont été effectuées. Les résultats suggèrent que la modification de la concentration totale de CYP1A2 dans le modèle entraîne une distorsion complète de l'équilibre entre la bioactivation et la détoxification. Ainsi, une diminution importante de la concentration totale du CYP1A2 entraîne une réduction importante de la bioactivation. Au contraire, une augmentation de la concentration totale du CYP12 affecte l'équilibre en faveur de la bioactivation, bien que cet effet soit modéré à de faibles concentrations de MeIQx. Ce dernier constat peut s'expliquer par l'implémentation des courbes sigmoïdes qui régulent l'équilibre des voies et favorisent la voie de détoxification par C-Hydroxy-MeIQx à de faibles concentrations (environ 1  $\mu$ M).

En conclusion, ces analyses ont démontré que la voie de détoxification prédomine sur une large gamme de concentrations opérant de 0.05 à 100  $\mu$ M renforçant ainsi la preuve de la fonction majeure du foie, qui est la détoxification. En fonction des concentrations de MeIQx, le ratio entre les voies de bioactivation et de détoxification est fortement modulé. Ce ratio représente l'équilibre entre la quantité de MeIQx métabolisé soit en produits de détoxification, soit transformé en Composé-Potentiellement-Génotoxique. De toute évidence, ce modèle fournit une sur-approximation du risque de bioactivation expliqué par le fait qu'il suppose que le système est fermé et que tous les composés non mesurés sont potentiellement des métabolites génotoxiques qui peuvent donc se lier à l'ADN pour former des adduits. Cependant, il suggère que le ratio prédit entre les voies de bioactivation et de détoxification pourrait être considéré comme un indicateur de la génotoxicité du MeIQx à faibles concentrations. Plus important encore, ce modèle illustre clairement l'importance de prendre en considération dans les approches de modélisation *in silico*, les mécanismes de régulation potentiels affectant la régiosélectivité. En effet, comme démontré dans la présente analyse, la régulation d'une seule enzyme (le CYP1A2 dans le cadre du MeIQx) peut fortement impacter sur l'équilibre entre la bioactivation et la détoxification.

Enfin, alors que les approches de plus haut niveau, exposées dans les chapitres 3 et 4 visent à prédire le métabolisme des xénobiotiques et identifier la génotoxicité des métabolites, une méthodologie plus fine permet de démystifier la complexité d'un système biologique. En effet, à la différence du modèle proposé dans cette présente étude, les approches précédemment utilisées sur le MeIQx (chapitres 3 et 4) étaient dans l'incapacité d'évaluer l'impact des sélectivités enzymatiques sur la bioactivation. Ce dernier propos expose parfaitement la complémentarité et l'emploi d'approches différentes pour l'appréhension de l'étude d'un xénobiotique dans toute sa complexité.

Les perspectives de ce travail consisteront principalement à tester la pertinence de cette approche proposée dans le cadre de la modélisation d'une exposition du MeIQx globale. En effet, une limitation principale du modèle est directement liée aux données acquises sur les hépatocytes humains, limitant ainsi l'étude à ce même cadre. Bien que le foie soit connu pour être le principal organe de la détoxification (justifiant

le choix d'avoir développé en premier lieu un modèle spécifique à cet organe) et que les prédictions associées à ce modèle soient conformes aux données issues de l'exposition humaine, il sera très intéressant de comparer la dynamique du MeIQx avec celle dans d'autres organes humains. Une autre limitation évidente du modèle est sa dépendance vis-à-vis du patient. En effet, l'ajustement du modèle était basé sur des jeux de données extraites de patients humains sans information sur les caractéristiques du patient telles que le polymorphisme enzymatique. Dans ce contexte, les données disponibles ne suffisent pas pour l'implémentation d'un modèle à plus large échelle qui soit multi-organes ou à l'échelle d'une population, qui nécessiterait de connaître l'individu moyen. Néanmoins, cette approche constitue une bonne étude préliminaire au développement de modèles complexes [262, 126] et souligne l'importance d'intégrer des mécanismes de régulation enzymatique dans des modèles plus complexes comme les modèles PBPK multi-organes.

## 6.4 Résumé

Les amines aromatiques hétérocycliques (AHA) sont des contaminants environnementaux et alimentaires potentiellement cancérigènes pour l'homme. Le MeIQx, l'une des AHA les plus abondantes dans les viandes grillées, peut être transformée par le CYP1A2 soit en métabolites qui seront détoxifiés, soit en produits bioactivés. Ces derniers sont toxiques et peuvent être à l'origine d'instabilités génomiques après leur fixation à l'ADN induisant la formation d'adduits à l'ADN.

L'enjeu de cette étude est d'étudier l'équilibre entre les voies de bioactivation conduisant à la formation d'adduits à l'ADN à travers la formation de composés potentiellement génotoxiques et celles de détoxification du MeIQx. Il a été montré qu'à faible concentration de MeIQx, l'enzyme CYP1A2 catalyse la transformation du MeIQx en une substance non toxique alors qu'à forte concentration, la même enzyme catalyse le même substrat MeIQx en composés toxiques. Pour pouvoir étudier les mécanismes complexes régissant le métabolisme du MeIQx, la famille de 73 modèles paramétrés précédemment sélectionnés a été simulée afin d'étudier l'importance de mécanismes de régulation complexe du CYP1A2 dans l'équilibre entre les voies de bioactivation et de détoxification.

Cette approche a mis en évidence que la compétition des enzymes au sein des différentes voies métaboliques se traduit par des effets biphasiques, qui fournissent une explication plausible sur la variabilité de la réponse du système observée entre les faibles et fortes concentrations. De plus, cette biphasicité modulant la réponse du CYP1A2 est un paramètre clé du système qui régule directement la bioactivation et donc la formation d'adduits à l'ADN.

# Conclusion et perspectives

## Conclusion

Le travail effectué a été à l'interface de nombreux domaines mêlant des connaissances biologiques, biochimiques, chemo-informatiques et informatiques. L'objectif global de celui-ci était de développer des méthodes informatiques pour étudier l'équilibre entre les voies de bioactivation et celles de détoxification chez les xénobiotiques. Cette problématique a été abordée par le développement d'une méthode intégrative de prédiction de construction *de novo* de graphes métaboliques et par l'implémentation d'un modèle mathématique sous-déterminé modélisant des effets de compétitions enzymatiques.

La construction *de novo* de graphes métaboliques est sujette à de nombreux problèmes. Tout d'abord, une bonne prédiction de graphes métaboliques est caractérisée par des scores de sensibilité et de précision élevés. En d'autres termes, les graphes prédits doivent être le plus en accord possible avec la réalité biologique en prédisant les métabolites expérimentalement validés, tout en évitant les sur-approximations de prédictions. Ainsi, la difficulté majeure n'est pas de prédire correctement les métabolites, mais de limiter la combinatoire inhérente aux méthodes actuelles de construction *de novo* de graphes métaboliques basées sur l'exploitation d'un dictionnaire de transformation ou sur une base de données contenant des réactions biochimiques. De plus, la bonne prédiction d'un réseau métabolique est un bon point, mais son annotation par l'identification des molécules réactives avec l'ADN ou la sélectivité enzymatique par exemple en est un autre important. Ces derniers points sont difficilement traités par les méthodes actuellement développées.

Cette thèse s'est attachée à pallier ces problèmes en proposant une méthode intégrative de construction *de novo* de graphes métaboliques à l'échelle moléculaire (voir 1.3.3). Cette dernière améliore une méthode de prédiction de réactions métaboliques existante tout en intégrant des méthodes de prédiction de sélectivité enzymatique pour les CYP, de sites de métabolisme (SOMs) et de sites de réactivité (SORs). La combinaison de ces méthodes a ainsi permis la construction *de novo* de graphes mé-

taboliques qui intègrent une bonne sensibilité et une bonne précision. De plus, cette dernière a permis l'annotation de ces graphes offrant l'identification des molécules et atomes réactifs avec l'ADN et l'identification de la sélectivité enzymatique vis-à-vis des CYP pour la catalyse de certaines réactions. Cette méthode a été appliquée avec succès pour la prédiction d'une famille complète de xénobiotiques que sont les AHAs et est la première à prédire les cartes métaboliques de ces contaminants alimentaires. Enfin, ce premier travail s'est soldé par l'implémentation d'un logiciel permettant aux biologistes d'effectuer simplement les tâches suivantes : la visualisation, la manipulation, l'exploration et l'analyse des graphes construits *de novo*.

Après avoir appréhendé un système à l'échelle moléculaire, il est nécessaire de comprendre et étudier la dynamique d'un système biologique. Cette étape peut parfois être complexe en raison de la limitation des données biologiques disponibles lorsque l'on souhaite l'étudier en détail (voir 1.3.2). Ce niveau de précision est souvent utile lors de l'étude du métabolisme de substances chimiques ou l'étude de mécanismes fins complexes. La seconde partie de cette thèse consistait à mettre au point une méthode permettant l'exploitation de modèles sous-déterminés afin d'étudier la dynamique du métabolisme de substances chimiques, qui sont régies par des mécanismes de régulation complexes. L'approche développée est basée à la différence des méthodologies classiques sur l'exploitation d'une famille d'ensembles de paramètres estimés décrivant la dynamique du système biologique étudié. Cette méthode appliquée au MeIQx a montré son succès à prédire l'équilibre des voies de bioactivation induisant la toxicité et celles de détoxification ; et à modéliser simplement des phénomènes complexes de compétitions enzymatiques régulés par l'emploi d'une sigmoïde modélisant la distribution des enzymes au sein des voies métaboliques qu'elles catalysent. En conclusion, ce dernier travail propose à la fois un modèle générique simple pour modéliser des compétitions enzymatiques et à la fois une méthodologie pour exploiter des données biologiques limitées lorsqu'une précision plus importante que celle obtenue avec les méthodes par contraintes est nécessaire.

Enfin, l'étude approfondie du MeIQx, qui a servi de fil directeur tout au long de ce document, a montré l'intérêt et la complémentarité des approches développées dans le cadre de cette thèse. Alors que les approches de construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions (chapitres 3 et 4) trouvent tout leur intérêt lors d'études préliminaires du métabolisme et de l'évaluation de la génotoxicité de xénobiotiques peu ou pas documentés, celles quantitatives dynamiques ciblées sur une molécule précise (chapitres 5 et 6) permettent d'en appréhender toute la complexité associée à son métabolisme médiant la toxicité.

Pour conclure, l'enjeu de ce travail était de mêler différentes techniques de modélisation pour prédire et étudier à différentes échelles le métabolisme de xénobiotiques. En effet, afin d'appréhender le métabolisme global d'une substance chimique, il est important de l'étudier à différentes échelles. Alors que les échelles situées au niveau de l'organe ou d'un organisme et les échelles cellulaires peuvent être adaptées pour

élucider la dynamique globale ou les aspects clés du métabolisme, les approches plus focalisées aux échelles cellulaires et moléculaires permettent d'en comprendre son détail et ses subtilités qui sont des points capitaux affectant le métabolisme global et parfois négligés. L'espoir de cette thèse est de voir l'émergence de nouvelles méthodes et frameworks intégratifs pour la prédiction et l'analyse intégrale du métabolisme des xénobiotiques chez l'homme.

## Perspectives

Des perspectives multiples s'ouvrent aux méthodes développées dans cette thèse. Elles touchent de nombreux domaines tels que la gestion des données, les applications et améliorations de modèles, l'intégration de la connaissance générée *in silico*.

### Perspectives de gestion des données

**Enrichissement de la base de données de MetaPrint2D-React** La base de données actuelle de MetaPrint2D-React repose sur l'exploitation de la version 2010.2 de la base de données métabolique Symyx Metabolite Database, maintenant rachetée par Dassault sous la marque Biovia. Un point important serait donc de mettre à jour cette base de données. De plus, un aspect intéressant pourrait être l'enrichissement de cette base de données avec les informations contenues dans d'autres bases de données publiques telles que KEGG [111, 110] et DrugBank [122, 249, 248]. L'avantage de l'utilisation d'autres bases de données est potentiellement d'augmenter la sensibilité des prédictions des métabolites. Néanmoins, plusieurs aspects devront être pris en compte. Premièrement, la redondance inhérente à la combinaison de plusieurs bases de données devra être traitée. Puis, un enrichissement de la base de données de travail pourrait fortement impacter l'explosion de la combinatoire qui devra être traitée par l'implémentation de nouveaux scores statistiques. Enfin, une conséquence directe de cet enrichissement pourrait être potentiellement, la réduction significative de la précision du workflow et donc la nécessité d'intégrer de nouvelles méthodes au sein de celui-ci.

**Automatisation de la construction de la base de données du workflow** Afin d'améliorer la précision, un ensemble de filtres basés sur des logiciels de prédictions de SOMs ont été implémentés. Afin de permettre leur combinaison et l'automatisation de la détection des SOMs, l'exploitation de ces logiciels a été calibrée en utilisant une base de données manuellement curée. De ce fait, en cas d'étude d'une famille de xénobiotiques éloignés de celle des AHAs, une autre base de données devra être créée. De plus, la construction manuelle d'une base de données ne permet pas la considération exhaustive des réactions contenues dans la littérature. Pour remédier à ce défaut et permettre d'améliorer l'utilisation générique du workflow, l'objectif serait de rajouter l'information des scores pour chaque outil dans une base de données contenant un ensemble de réactions métaboliques expérimentalement validées. Puis, lors de l'étude



d'une nouvelle famille de xénobiotiques, une sous-base de données pour l'optimisation de scores basés sur les similarités moléculaires de la famille d'intérêt (groupe chimique, utilisation de descripteurs moléculaires ...) serait automatiquement construite à partir de la base de données d'apprentissage principale.

**Priorisation des voies métaboliques et amélioration de la filtration** A l'heure actuelle, le workflow n'est pas en mesure de différencier les voies métaboliques majeures et mineures. Pour pallier à ce défaut, plusieurs solutions peuvent être implémentées. La première serait de calculer un score commun des résultats retournés par les différents logiciels implémentés dans le workflow afin de classer les voies métaboliques. La seconde solution serait de compiler des descripteurs énergétiques et d'accessibilité à l'image de SMARTCYP [192, 193] afin de classer les métabolites prédits. Une autre solution serait à l'aune de PROXIMAL [257], d'utiliser les activités catalytiques et les abondances enzymatiques de chaque enzyme pour pouvoir distinguer les différentes voies métaboliques. Enfin d'autres méthodes largement implémentées dans les logiciels comme META, Meteor ou encore TIMES peuvent aussi être considérées. Parmi elles, la génération de modèles QSAR pour classer, mais aussi flirter peut être aussi envisagée, ou encore l'utilisation de descripteurs (pouvant être aisément prédits) comme le logP afin d'améliorer l'algorithme actuel basé sur la réactivité permettant d'identifier quand une molécule peut-être considérée comme finale [176].

## Perspectives d'applications et améliorations de modèles

**Application à d'autres xénobiotiques ou médicaments** Le phénomène de catalyse d'un même substrat par une même enzyme produisant deux composés distincts pourrait être un phénomène sous-considéré et plus fréquent que l'on pense. Dans la poursuite du travail effectué sur le MeIQx et pour continuer à éclaircir ce mécanisme particulier, il pourrait être intéressant d'envisager d'appliquer le modèle développé sur d'autres substances toxiques. En effet, des molécules comme le bisphénol A [181] ou le propranolol [149, 150] ont toutes deux une cinétique biphasique particulière. Elles pourraient être potentiellement de bons candidats d'application du modèle développé.

**Application aux modèles multi-organes** Il pourrait être intéressant de comparer l'influence d'un tel mécanisme à l'échelle d'un modèle PBPK multi-organes au sein d'un individu et d'une population. La démarche pourrait être dans un premier temps de comparer un modèle classique avec celui développé dans cette thèse afin d'étudier l'importance de ce mécanisme à une plus large échelle. Puis, il pourrait aussi être intéressant de l'appliquer sur d'autres molécules et de réfléchir à de nouvelles méthodes permettant de prédire les paramètres de régiosélectivité dépendante du substrat. En d'autres termes, la prédiction des paramètres sigmoïdes afin d'établir un modèle générique et faciliter leur utilisation, pourrait être un autre challenge.

### Perspectives d'intégration de la connaissance générée *in silico*

**Identification automatisée des métabolites via spectrométrie de masse** Un avantage majeur de la prédiction des graphes de métabolisme réside dans leurs sensibilités élevées. Cette dernière pourrait être un avantage pour développer de nouvelles techniques d'identification automatique *in silico* de métabolites à l'image de Mass-MetaSite, MetabolitePilot ou encore IMassFrag [177, 261]. Néanmoins, pour se différencier de ces méthodes, une méthode pourrait être l'exploitation de la prédiction du spectre de masse à partir du SMILES en intégrant dans un pipeline des logiciels comme CFM-ID (Competitive Fragmentation Modeling for Metabolite Identification) [9, 7, 8].

**Etude de la dynamique des réseaux métaboliques prédits** Un challenge majeur serait d'apporter un aspect quantitatif à chaque réseau prédit et de les étudier à travers des questions de type : Comment maximise/minimise-t-on les adduits à l'ADN ? ; Dans les conditions X et Y, comment sont impactées les sorties du réseau métaboliques ? Une piste pour aborder ce point serait d'explorer les méthodes basées sur les contraintes et notamment l'utilisation de la dFBA (dynamic FBA) dpFBA (dynamic parsimonious FBA), qui ont été récemment utilisées dans le cadre d'un modèle multi-organes combinant un réseau métabolique humain avec des données métaboliques humaines [124, 224]. De plus, ces réseaux peuvent être aisément contraints par l'exploitation de données de type abondance enzymatique issues de PaxDb [241] ou encore l'utilisation/la prédiction d'intervalles de paramètres cinétiques de type  $V_m$  et  $K_m$ .



# Bibliographie

- [1] What Is Systems Biology (<https://www.systemsbiology.org/about/what-is-systems-biology>).
- [2] IARC Monographs programme on the evaluation of the carcinogenic risk of chemicals to humans. Overall evaluations of carcinogenicity : an updating of IARC Monographs volumes 1 to 42. IARC monographs on the evaluation of carcinogenic risks to humans. Supplement / World Health Organization, International Agency for Research on Cancer, 7 :1-440, 1987.
- [3] 2-nitrofluorene. IARC monographs on the evaluation of carcinogenic risks to humans, 46 :277-289, 1989.
- [4] Accelrys. TOPKAT (TOxicity Prediction by Komputer Assisted Technology). 2015.
- [5] Samuel E. Adams. Molecular Similarity and Xenobiotic Metabolism. PhD thesis, 2010.
- [6] Lovisa Afzelius, Catrin Hasselgren Arnby, Anders Broo, Lars Carlsson, Christine Isaksson, Ulrik Jurva, Britta Kjellander, Karin Kolmodin, Kristina Nilsson, Florian Raubacher, and Lars Weidolf. State-of-the-art tools for computational site of metabolism predictions : comparative analysis, mechanistical insights, and future applications. *Drug Metabolism Reviews*, 39(1) :61-86, 2007.
- [7] Felicity Allen, Russ Greiner, and David Wishart. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics*, 11(1) :98-110, February 2015.
- [8] Felicity Allen, Allison Pon, Russ Greiner, and David Wishart. Computational Prediction of Electron Ionization Mass Spectra to Assist in GC/MS Compound Identification. *Analytical Chemistry*, 88(15) :7689-7697, August 2016.
- [9] Felicity Allen, Allison Pon, Michael Wilson, Russ Greiner, and David Wishart. CFM-ID : a web server for annotation, spectrum prediction and metabolite iden-

- tification from tandem mass spectra. *Nucleic Acids Research*, 42(Web Server issue) :W94–99, July 2014.
- [10] Anna Alonen, Olli Aitio, Kati Hakala, Leena Luukkanen, Moshe Finel, and Risto Kostianen. Biosynthesis of Dobutamine Monoglucuronides and Glucuronidation of Dobutamine by Recombinant Human Udp-Glucuronosyltransferases. *Drug Metabolism and Disposition*, 33(5) :657–663, May 2005.
- [11] Vinicius Alves, Eugene Muratov, Stephen Capuzzi, Regina Politi, Yen Low, Rodolpho Braga, Alexey V. Zakharov, Alexander Sedykh, Elena Mokshyna, Sherif Farag, Carolina Andrade, Victor Kuz'min, Denis Fourches, and Alexander Tropsha. Alarms about structural alerts. *Green chemistry : an international journal and green chemistry resource : GC*, 18(16) :4348–4360, August 2016.
- [12] Assembly of Life Sciences (U S. ) Committee on Amines and David Barringer Clayson. *Aromatic Amines : An Assessment of the Biological and Environmental Effects*. National Academies, January 1981. Google-Books-ID : slArAAAAYAAJ.
- [13] Eric Anderson, Gilman D Veith, David Weininger, and Minn.) Environmental Research Laboratory (Duluth. SMILES, a line notation and computerized interpreter for chemical structures. U.S. Environmental Protection Agency, Environmental Research Laboratory, Duluth, MN, 1987. OCLC : 20862752.
- [14] K. Augustsson, K. Skog, M. Jägerstad, P. W. Dickman, and G. Steineck. Dietary heterocyclic amines and cancer of the colon, rectum, bladder, and kidney : a population-based study. *Lancet (London, England)*, 353(9154) :703–707, February 1999.
- [15] Barbara M. Bakker, Karen van Eunen, Jeroen A. L. Jeneson, Natal A. W. van Riel, Frank J. Bruggeman, and Bas Teusink. Systems biology from micro-organisms to human metabolic diseases : the role of detailed kinetic models. *Biochemical Society Transactions*, 38(5) :1294–1301, October 2010.
- [16] J. H. Barrett, G. Smith, R. Waxman, N. Gooderham, T. Lightfoot, R. C. Garner, K. Augustsson, C. R. Wolf, D. T. Bishop, D. Forman, and Colorectal Cancer Study Group. Investigation of interaction between N-acetyltransferase 2 and heterocyclic amines as potential risk factors for colorectal cancer. *Carcinogenesis*, 24(2) :275–282, February 2003.
- [17] Todd M. Baughman, Christine L. Talarico, and John R. Soglia. Evaluation of the metabolism of propranolol by linear ion trap technology in mouse, rat, dog, monkey, and human cryopreserved hepatocytes. *Rapid Communications in Mass Spectrometry*, 23(14) :2146–2150, July 2009.
- [18] F. A. Beland. *Chemical Carcinogenesis and Mutagenesis I* | Springer. 1990.
- [19] Medjda Bellamri, Ludovic Le Hegarat, Robert J. Turesky, and Sophie Langouët. Metabolism of the Tobacco Carcinogen 2-Amino-9h-pyrido[2,3-b]indole (AαC)

- in Primary Human Hepatocytes. *Chemical Research in Toxicology*, 30(2) :657–668, February 2017.
- [20] C. Belloc, S. Baird, J. Cosme, S. Lecoeur, J. C. Gautier, D. Challine, I. de Waziers, J. P. Flinois, and P. H. Beaune. Human cytochromes P450 expressed in *Escherichia coli* : production of specific antibodies. *Toxicology*, 106(1-3) :207–219, January 1996.
- [21] Pr Dominique Belpomme. Ces maladies créées par l’homme : Comment la dégradation de l’environnement met en péril notre santé. Albin Michel, February 2004.
- [22] Rebecca L. Blanchard, Robert R. Freimuth, Jochen Buck, Richard M. Weinshilboum, and Michael W. H. Coughtrie. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics*, 14(3) :199–211, March 2004.
- [23] M. Blum, A. Demierre, D. M. Grant, M. Heim, and U. A. Meyer. Molecular mechanism of slow acetylation of drugs and carcinogens in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 88(12) :5237–5241, June 1991.
- [24] A. R. Boobis, A. M. Lynch, S. Murray, R. de la Torre, A. Solans, M. Farré, J. Segura, N. J. Gooderham, and D. S. Davies. CYP1a2-catalyzed conversion of dietary heterocyclic amines to their proximate carcinogens is their major route of metabolism in humans. *Cancer Research*, 54(1) :89–94, January 1994.
- [25] A. Bordbar and B. O. Palsson. Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *Journal of Internal Medicine*, 271(2) :131–141, February 2012.
- [26] Aarash Bordbar, Adam M. Feist, Renata Usaitė-Black, Joseph Woodcock, Bernhard O. Palsson, and Iman Famili. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC systems biology*, 5 :180, October 2011.
- [27] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 : Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [28] Kevin S. Brown and James P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review. E, Statistical, Non-linear, and Soft Matter Physics*, 68(2 Pt 1) :021904, August 2003.
- [29] Andrej Bugrim, Tatiana Nikolskaya, and Yuri Nikolsky. Early prediction of drug metabolism and toxicity : systems biology approach and modeling. *Drug Discovery Today*, 9(3) :127–135, February 2004.

- [30] L. M. Butler, R. Sinha, R. C. Millikan, C. F. Martin, B. Newman, M. D. Gammon, A. S. Ammerman, and R. S. Sandler. Heterocyclic amines, meat intake, and association with colon cancer in a population-based study. *American Journal of Epidemiology*, 157(5) :434–445, March 2003.
- [31] M. A. Butler, N. P. Lang, J. F. Young, N. E. Caporaso, P. Vineis, R. B. Hayes, C. H. Teitel, J. P. Massengill, M. F. Lawsen, and F. F. Kadlubar. Determination of CYP1a2 and NAT2 phenotypes in human populations by analysis of caffeine urinary metabolites. *Pharmacogenetics*, 2(3) :116–127, June 1992.
- [32] Mary Ann Butler, F. Peter Guengerich, and Fred F. Kadlubar. Metabolic Oxidation of the Carcinogens 4-Aminobiphenyl and 4,4'-Methylenebis(2-chloroaniline) by Human Hepatic Microsomes and by Purified Rat Hepatic Cytochrome P-450 Monooxygenases. *Cancer Research*, 49(1) :25–31, January 1989.
- [33] William G. Button, Philip N. Judson, Anthony Long, and Jonathan D. Vessey. Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. *Journal of Chemical Information and Computer Sciences*, 43(5) :1371–1377, October 2003.
- [34] Tingting Cai, Lihua Yao, and Robert J. Turesky. Bioactivation of Heterocyclic Aromatic Amines by UDP Glucuronosyltransferases. *Chemical Research in Toxicology*, 29(5) :879–891, 2016.
- [35] Valérie Campagna-Slater, Joshua Pottel, Eric Therrien, Louis-David Cantin, and Nicolas Moitessier. Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by p450s. *Journal of Chemical Information and Modeling*, 52(9) :2471–2483, September 2012.
- [36] Arthur J. Campbell, Michelle L. Lamb, and Diane Joseph-McCarthy. Ensemble-based docking using biased molecular dynamics. *Journal of Chemical Information and Modeling*, 54(7) :2127–2138, July 2014.
- [37] Jerry L. Campbell, Melvin E. Andersen, Paul M. Hinderliter, Kun Don Yi, Timothy P. Pastoor, Charles B. Breckenridge, and Harvey J. Clewell. PBPK Model for Atrazine and Its Chlorotriazine Metabolites in Rat and Human. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 150(2) :441–453, April 2016.
- [38] Jerry L. Campbell, Miyoung Yoon, and Harvey J. Clewell. A case study on quantitative in vitro to in vivo extrapolation for environmental esters : Methyl-, propyl- and butylparaben. *Toxicology*, 332 :67–76, June 2015.
- [39] J. Castañeda-Acosta, P. L. Bounds, and G. W. Winston. Microsomal deacetylation of ring-hydroxylated 2-(acetylamino)fluorene isomers : effect of ring position and molecular mechanics considerations. *Journal of Biochemical and Molecular Toxicology*, 13(5) :279–286, 1999.

- [40] Shu-Ying Chang, Wenyang Li, Sarah C. Traeger, Bei Wang, Donghui Cui, Hongjian Zhang, Bo Wen, and A. David Rodrigues. Confirmation that cytochrome P450 2c8 (CYP2c8) plays a minor role in (S)-(+)- and (R)-(-)-ibuprofen hydroxylation in vitro. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 36(12) :2513–2522, December 2008.
- [41] Ann Chao, Michael J. Thun, Cari J. Connell, Marjorie L. McCullough, Eric J. Jacobs, W. Dana Flanders, Carmen Rodriguez, Rashmi Sinha, and Eugenia E. Calle. Meat consumption and risk of colorectal cancer. *JAMA*, 293(2) :172–182, January 2005.
- [42] William W. Chen, Mario Niepel, and Peter K. Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes & Development*, 24(17) :1861–1875, September 2010.
- [43] T. A. Chiang, W. Pei-Fen, L. S. Ying, L. F. Wang, and Y. C. Ko. Mutagenicity and aromatic amine content of fumes from heated cooking oils produced in Taiwan. *Food and Chemical Toxicology : An International Journal Published for the British Industrial Biological Research Association*, 37(2-3) :125–134, March 1999.
- [44] H. C. Chou, N. P. Lang, and F. F. Kadlubar. Metabolic activation of N-hydroxy arylamines and N-hydroxy heterocyclic amines by human sulfotransferase(s). *Cancer Research*, 55(3) :525–529, February 1995.
- [45] Athel Cornish-Bowden. One hundred years of Michaelis–Menten kinetics. *Perspectives in Science*, 4 :3–9, March 2015.
- [46] Rafael S. Costa, Daniel Machado, Isabel Rocha, and Eugénio C. Ferreira. Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis-Menten and approximate kinetic equations. *Bio Systems*, 100(2) :150–157, May 2010.
- [47] Xavier Coumoul, France Massicot, Jean-Claude Pairon, Pascal Andujar, Armelle Baeza-Squiban, Robert Barouki, Laurent Bodin, Lucie Chevillard, Xavier Declèves, Robert Garnier, and Francelyne Marano. *Toxicologie : Fondamentaux et fiches pratiques*. Dunod, April 2017. Google-Books-ID : QwS4DgAAQBAJ.
- [48] Edward Croom. Metabolism of xenobiotics of human environments. *Progress in Molecular Biology and Translational Science*, 112 :31–88, 2012.
- [49] Gabriele Cruciani, Emanuele Carosati, Benoit De Boeck, Kantharaj Ethirajulu, Claire Mackie, Trevor Howe, and Riccardo Vianello. MetaSite : understanding metabolism in human cytochromes from the perspective of the chemist. *Journal of Medicinal Chemistry*, 48(22) :6970–6979, November 2005.
- [50] Arthur Dalby, James G. Nourse, W. Douglas Hounshell, Ann K. I. Gushurst, David L. Grier, Burton A. Leland, and John Laufer. Description of several che-



- mical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of Chemical Information and Computer Sciences*, 32(3) :244–255, May 1992.
- [51] Na Le Dang, Tyler B. Hughes, Varun Krishnamurthy, and S. Joshua Swamidass. A simple model predicts UGT-mediated metabolism. *Bioinformatics (Oxford, England)*, 32(20) :3183–3189, October 2016.
- [52] F. Darvas. *Metabolexpert : An Expert System for Predicting Metabolism of Substances* In : Kaiser K.L.E. (eds) *QSAR in Environmental Toxicology - II*. Springer, Dordrecht. pages 71–81, 1987.
- [53] Daylight. *Daylight Chemical Information Systems, Inc. of Aliso Viejo, CA ; Daylight Theory Manual*. 2006.
- [54] Victorien Delannée, Sophie Langouët, Nathalie Théret, and Anne Siegel. A modeling approach to evaluate the balance between bioactivation and detoxification of meiqx in human hepatocytes. *PeerJ*, 5 :e3703, September 2017.
- [55] Juan C. Domínguez-Romero, Juan F. García-Reyes, Rubén Martínez-Romero, Paula Berton, Esther Martínez-Lara, María L. Del Moral-Leal, and Antonio Molina-Díaz. Combined data mining strategy for the systematic identification of sport drug metabolites in urine by liquid chromatography time-of-flight mass spectrometry. *Analytica Chimica Acta*, 761 :1–10, January 2013.
- [56] Manfred Drack, Wilfried Apfalter, and David Pouvreau. ON THE MAKING OF A SYSTEM THEORY OF LIFE : PAUL A WEISS AND LUDWIG VON BERTALANFFY'S CONCEPTUAL CONNECTION. *The Quarterly review of biology*, 82(4) :349–373, December 2007.
- [57] D. L. Eaton, E. P. Gallagher, T. K. Bammler, and K. L. Kunze. Role of cytochrome P4501a2 in chemical carcinogenesis : implications for human variability in expression and enzyme activity. *Pharmacogenetics*, 5(5) :259–274, October 1995.
- [58] Birgit Eiermann, Per Olof Edlund, Agneta Tjernberg, Per Dalén, Marja-Liisa Dahl, and Leif Bertilsson. 1- and 3-Hydroxylations, in Addition to 4-Hydroxylation, of Debrisoquine Are Catalyzed by Cytochrome P450 2d6 in Humans. *Drug Metabolism and Disposition*, 26(11) :1096–1101, November 1998.
- [59] Mark J. Embrechts and Sean Ekins. Classification of metabolites with kernel-partial least squares (K-PLS). *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 35(3) :325–327, March 2007.
- [60] Henri A. Favre and Warren H. Powell. *Nomenclature of Organic Chemistry*. December 2013. DOI : 10.1039/9781849733069.
- [61] Amir Feizi and Sergio Bordel. Metabolic and protein interaction sub-networks controlling the proliferation rate of cancer cells and their impact on patient survival. *Scientific Reports*, 3 :3041, October 2013.

- [62] D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *The Biochemical Journal*, 238(3) :781–786, September 1986.
- [63] J S Felton, M Jagerstad, M G Knize, K Skog, and Keiji Wakabayashi. Contents in foods, beverages and tobacco. In : Nagao M, Sugimura T, editors. *Food Borne Carcinogens Heterocyclic Amines*. John Wiley & Sons Ltd.; Chichester, England : 2000. pp. 31–71. 2000.
- [64] Jasmin Fisher and Thomas A. Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11) :1239–1249, November 2007.
- [65] Ori Folger, Livnat Jerby, Christian Frezza, Eyal Gottlieb, Eytan Ruppin, and Tomer Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology*, 7 :501, June 2011.
- [66] National Center for Biotechnology Information. PubChem Compound Database. Cid=62275, 2016.
- [67] Christian Fossen. *Molecular Systems Biology - Department of Biology*.
- [68] null Franklin. The N-glucuronidation of xenobiotics. An aspet-supported symposium held at the 1996 faseb meeting in washington, dc. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 26(9) :829, September 1998.
- [69] C. B. Frederick, C. C. Weis, T. J. Flammang, C. N. Martin, and F. F. Kadlubar. Hepatic N-oxidation, acetyl-transfer and DNA-binding of the acetylated metabolites of the carcinogen, benzidine. *Carcinogenesis*, 6(7) :959–965, July 1985.
- [70] Richard A. Friesner, Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin. Glide : a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7) :1739–1749, March 2004.
- [71] Niranjali Gamage, Amanda Barnett, Nadine Hempel, Ronald G. Duggleby, Kelly F. Windmill, Jennifer L. Martin, and Michael E. McManus. Human sulfotransferases and their role in chemical metabolism. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 90(1) :5–22, March 2006.
- [72] Junfeng Gao, Lynda B. M. Ellis, and Lawrence P. Wackett. The University of Minnesota Biocatalysis/Biodegradation Database : improving public access. *Nucleic Acids Research*, 38(Database issue) :D488–491, January 2010.
- [73] Monika Gibis. *Heterocyclic Aromatic Amines in Cooked Meat Products : Causes, Formation, Occurrence, and Risk Assessment*. *Comprehensive Reviews in Food Science and Food Safety*, 15(2) :269–302, March 2016.

- [74] Hugo Girard, Lesley M. Butler, Lyne Villeneuve, Robert C. Millikan, Rashmi Sinha, Robert S. Sandler, and Chantal Guillemette. UGT1a1 and UGT1a9 functional variants, meat intake, and colon cancer, among Caucasians and African-Americans. *Mutation Research*, 644(1-2) :56–63, September 2008.
- [75] Hugo Girard, Jean Thibaudeau, Michael H. Court, Louis-Charles Fortier, Lyne Villeneuve, Patrick Caron, Qin Hao, Lisa L. von Moltke, David J. Greenblatt, and Chantal Guillemette. UGT1a1 polymorphisms are important determinants of dietary carcinogen detoxification in the liver. *Hepatology (Baltimore, Md.)*, 42(2) :448–457, August 2005.
- [76] Dan Gu, Lynn McNaughton, David Lemaster, Brian G. Lake, Nigel J. Goodeham, Fred F. Kadlubar, and Robert J. Turesky. A comprehensive approach to the profiling of the cooked meat carcinogens 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine, and their metabolites in human urine. *Chemical Research in Toxicology*, 23(4) :788–801, April 2010.
- [77] F. P. Guengerich and C. G. Turvy. Comparison of levels of several human microsomal cytochrome P-450 enzymes and epoxide hydrolase in normal and disease states using immunochemical analysis of surgical liver samples. *The Journal of Pharmacology and Experimental Therapeutics*, 256(3) :1189–1194, March 1991.
- [78] F. Peter Guengerich. Cytochrome p450 and chemical toxicology. *Chemical Research in Toxicology*, 21(1) :70–83, January 2008.
- [79] Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, and James P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology*, 3(10) :1871–1878, October 2007.
- [80] Thomas A. Halgren, Robert B. Murphy, Richard A. Friesner, Hege S. Beard, Leah L. Frye, W. Thomas Pollard, and Jay L. Banks. Glide : a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry*, 47(7) :1750–1759, March 2004.
- [81] George J. Hammons, F. Peter Guengerich, Constance C. Weis, Frederick A. Beland, and Fred F. Kadlubar. Metabolic Oxidation of Carcinogenic Arylamines by Rat, Dog, and Human Hepatic Microsomes and by Purified Flavin-containing and Cytochrome P-450 Monooxygenases. *Cancer Research*, 45(8) :3578–3585, August 1985.
- [82] D. W. Hein. N-Acetyltransferase genetics and their role in predisposition to aromatic and heterocyclic amine-induced carcinogenesis. *Toxicology Letters*, 112-113 :349–356, March 2000.

- [83] D. W. Hein. N-acetyltransferase 2 genetic polymorphism : effects of carcinogen and haplotype on urinary bladder cancer risk. *Oncogene*, 25(11) :1649–1658, March 2006.
- [84] D. W. Hein, M. A. Doll, A. J. Fretland, M. A. Leff, S. J. Webb, G. H. Xiao, U. S. Devanaboyina, N. A. Nangju, and Y. Feng. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 9(1) :29–42, January 2000.
- [85] D W Hein, M A Doll, T D Rustan, K Gray, Y Feng, R J Ferguson, and D M Grant. Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis*, 14(8) :1633–1638, August 1993.
- [86] D. W. Hein, T. D. Rustan, R. J. Ferguson, M. A. Doll, and K. Gray. Metabolic activation of aromatic and heterocyclic N-hydroxyarylamines by wild-type and mutant recombinant human NAT1 and NAT2 acetyltransferases. *Archives of Toxicology*, 68(2) :129–133, 1994.
- [87] David W. Hein. Molecular genetics and function of NAT1 and NAT2 : role in aromatic amine metabolism and carcinogenesis. *Mutation Research*, 506-507 :65–77, September 2002.
- [88] Stephen Heller, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi, and Igor Pletnev. InChI - the worldwide chemical structure identifier standard. *Journal of Cheminformatics*, 5 :7, January 2013.
- [89] Matthias Hennemann, Arno Friedl, Mario Lobell, Jörg Keldenich, Alexander Hillisch, Timothy Clark, and Andreas H. Göller. CypScore : Quantitative prediction of reactivity toward cytochromes P450 based on semiempirical molecular orbital theory. *ChemMedChem*, 4(4) :657–669, April 2009.
- [90] Roberto Heredia Ortiz, Anne Maître, Damien Barbeau, Michel Lafontaine, and Michèle Bouchard. Use of physiologically-based pharmacokinetic modeling to simulate the profiles of 3-hydroxybenzo(a)pyrene in workers exposed to polycyclic aromatic hydrocarbons. *PloS One*, 9(7) :e102570, 2014.
- [91] Toyoko Hiroi, Toshio Chow, Susumu Imaoka, and Yoshihiko Funae. Catalytic Specificity of CYP2d Isoforms in Rat and Human. *Drug Metabolism and Disposition*, 30(9) :970–976, September 2002.
- [92] Kongwang Hu and Feihu Chen. Identification of significant pathways in gastric cancer based on protein-protein interaction networks and cluster analysis. *Genetics and Molecular Biology*, 35(3) :701–708, July 2012.

- [93] Tyler B. Hughes, Na Le Dang, Grover P. Miller, and S. Joshua Swamidass. Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network. *ACS central science*, 2(8) :529–537, August 2016.
- [94] Tyler B. Hughes, Grover P. Miller, and S. Joshua Swamidass. Site of reactivity models predict molecular reactivity of diverse chemicals with glutathione. *Chemical Research in Toxicology*, 28(4) :797–809, April 2015.
- [95] IARC. International Agency for Research on Cancer.. Tobacco smoking. Vol. 38. International Agency for Research on Cancer ; Lyon, France : 1986. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 1986.
- [96] IARC. International Agency for Research on Cancer. Some naturally occurring substances : food items and constituents, heterocyclic aromatic amines and mycotoxins. Vol. 56. International Agency for Research on Cancer ; Lyon, France : 1993. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 1993.
- [97] K. Ikeya, A. K. Jaiswal, R. A. Owens, J. E. Jones, D. W. Nebert, and S. Kimura. Human CYP1a2 : sequence, gene structure, comparison with the mouse and rat orthologous gene, and differences in liver 1a2 mRNA expression. *Molecular Endocrinology (Baltimore, Md.)*, 3(9) :1399–1408, September 1989.
- [98] Masahiko Iwasaki, Yoshinobu Yoshimura, Satoru Asahi, Kimitoshi Saito, Shuichi Sakai, Shigemichi Morita, Osamu Takenaka, Toshio Inoda, Eiji Kashiyama, Akinori Aoyama, Takeshi Nakabayashi, Satoshi Omori, Takashi Kuwabara, Takashi Izumi, Kouichi Nakamura, Kaoru Takanaka, Yukiharu Nakayama, Mitsuaki Takeuchi, Hideki Nakamura, Shunichi Kametani, Yoshiaki Terauchi, Takanori Hashizume, Sekio Nagayama, Toshiyuki Kume, Meguru Achira, Hiroyuki Kawai, Takashi Kawashiro, Akio Nakamura, Yasuhiro Nakai, Akira Kagayama, Toshifumi Shiraga, Takuro Niwa, Takuya Yoshimura, Jun Morita, Fukuichi Oh-sawa, Masato Tani, Nobuo Osawa, Keiichi Ida, and Kiyoshi Noguchi. Functional characterization of single nucleotide polymorphisms with amino acid substitution in CYP1a2, CYP2a6, and CYP2b6 found in the Japanese population. *Drug Metabolism and Pharmacokinetics*, 19(6) :444–452, December 2004.
- [99] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, and Bray F. GLOBOCAN 2012 : Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 v1.0. 2012.
- [100] Nina Jeliaskova and Vedrin Jeliaskov. AMBIT RESTful web services : an implementation of the OpenTox application programming interface. *Journal of Cheminformatics*, 3 :18, May 2011.
- [101] Livnat Jerby and Eytan Ruppin. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clinical Cancer Research : An Official*

- Journal of the American Association for Cancer Research, 18(20) :5572–5584, October 2012.
- [102] Li Jiang, Si-Cheng Liang, Chao Wang, Guang-Bo Ge, Xiao-Kui Huo, Xiao-Yi Qi, Sa Deng, Ke-Xin Liu, and Xiao-Chi Ma. Identifying and applying a highly selective probe to simultaneously determine the O-glucuronidation activity of human UGT1a3 and UGT1a4. *Scientific Reports*, 5, April 2015.
- [103] Zhengwen Jiang, Nadine Dragin, Lucia F. Jorge-Nebert, Martha V. Martin, F. Peter Guengerich, Eleni Aklillu, Magnus Ingelman-Sundberg, George J. Hammons, Beverly D. Lyn-Cook, Fred F. Kadlubar, Shannon N. Saldana, Michael Sorter, Alexander A. Vinks, Nassr Nassr, Oliver von Richter, Li Jin, and Daniel W. Nebert. Search for an association between the human CYP1a2 genotype and CYP1a2 metabolic phenotype. *Pharmacogenetics and Genomics*, 16(5) :359–367, May 2006.
- [104] Kenneth A. Johnson and Roger S. Goody. The Original Michaelis Constant : Translation of the 1913 Michaelis–Menten Paper. *Biochemistry*, 50(39) :8264–8269, October 2011.
- [105] Hm Jones and K. Rowland-Yeo. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT : pharmacometrics & systems pharmacology*, 2 :e63, August 2013.
- [106] F. Kadlubar, R. K. Kaderlik, G. J. Mulder, D. Lin, M. A. Butler, C. H. Teitel, R. F. Minchin, K. F. Ilett, M. D. Friesen, and H. Bartsch. Metabolic activation and DNA adduct detection of PhIP in dogs, rats, and humans in relation to urinary bladder and colon carcinogenesis. *Princess Takamatsu Symposia*, 23 :207–213, 1995.
- [107] F. F. Kadlubar, M. A. Butler, K. R. Kaderlik, H. C. Chou, and N. P. Lang. Polymorphisms for aromatic amine metabolism in humans : relevance for human carcinogenesis. *Environmental Health Perspectives*, 98 :69–74, November 1992.
- [108] Fred F. Kadlubar, James A. Miller, and Elizabeth C. Miller. Hepatic Microsomal N-Glucuronidation and Nucleic Acid Binding of N-Hydroxy Arylamines in Relation to Urinary Bladder Carcinogenesis. *Cancer Research*, 37(3) :805–814, March 1977.
- [109] S. Kaivosaari, M. Finel, and M. Koskinen. N-glucuronidation of drugs and other xenobiotics by human and animal UDP-glucuronosyltransferases. *Xenobiotica*, 41(8) :652–669, Aug 2011.
- [110] M. Kanehisa and S. Goto. KEGG : kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1) :27–30, January 2000.
- [111] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle : back

- to metabolism in KEGG. *Nucleic Acids Research*, 42(Database issue) :D199–205, January 2014.
- [112] R. Kato. Metabolic activation of mutagenic heterocyclic aromatic amines from protein pyrolysates. *Critical Reviews in Toxicology*, 16(4) :307–348, 1986.
- [113] Tony K. L. Kiang, Mary H. H. Ensom, and Thomas K. H. Chang. UDP-glucuronosyltransferases and clinical drug-drug interactions. *Pharmacology & Therapeutics*, 106(1) :97–132, April 2005.
- [114] Hyun Uk Kim, Seung Bum Sohn, and Sang Yup Lee. Metabolic network modeling and simulation for drug targeting and discovery. *Biotechnology Journal*, 7(3) :330–342, March 2012.
- [115] Johannes Kirchmair, Mark J. Williamson, Jonathan D. Tyzack, Lu Tan, Peter J. Bond, Andreas Bender, and Robert C. Glen. Computational prediction of metabolism : sites, products, SAR, P450 enzyme dynamics, and mechanisms. *Journal of Chemical Information and Modeling*, 52(3) :617–648, March 2012.
- [116] Hiroaki Kitano. *Foundations of Systems Biology*. MIT Press, 2001. Google-Books-ID : zfXFQgAACAAJ.
- [117] Hiroaki Kitano. *Systems biology : a brief overview*. *Science (New York, N.Y.)*, 295(5560) :1662–1664, March 2002.
- [118] R. Kiwamoto, A. Spenkelnik, I. M. C. M. Rietjens, and A. Punt. An integrated QSAR-PBK/D modelling approach for predicting detoxification and DNA adduct formation of 18 acyclic food-borne  $\alpha,\beta$ -unsaturated aldehydes. *Toxicology and Applied Pharmacology*, 282(1) :108–117, January 2015.
- [119] G. Klopman, M. Dimayuga, and J. Talafous. META. 1. A program for the evaluation of metabolic transformation of chemicals. *Journal of Chemical Information and Computer Sciences*, 34(6) :1320–1325, December 1994.
- [120] G. Klopman, M. Tu, and J. Talafous. META. 3. A genetic algorithm for metabolic transform priorities optimization. *Journal of Chemical Information and Computer Sciences*, 37(2) :329–334, April 1997.
- [121] S. Knasmüller, M. Murkovic, W. Pfau, and G. Sontag. Heterocyclic aromatic amines—still a challenge for scientists. *Journal of Chromatography B*, 802(1) :1–2, March 2004.
- [122] Craig Knox, Vivian Law, Timothy Jewison, Philip Liu, Son Ly, Alex Frolkis, Allison Pon, Kelly Banco, Christine Mak, Vanessa Neveu, Yannick Djoumbou, Roman Eisner, An Chi Guo, and David S. Wishart. DrugBank 3.0 : a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Research*, 39(Database issue) :D1035–1041, January 2011.

- [123] Jeffrey S. Knutsen, Brent D. Kerger, Brent Finley, and Dennis J. Paustenbach. A calibrated human PBPK model for benzene inhalation with urinary bladder and bone marrow compartments. *Risk Analysis : An Official Publication of the Society for Risk Analysis*, 33(7) :1237–1251, July 2013.
- [124] M. Krauss, S. Schaller, S. Borchers, R. Findeisen, J. Lippert, and L. Kuepfer. Integrating cellular metabolism into a multiscale whole-body model. *PLoS Comput. Biol.*, 8(10) :e1002750, 2012.
- [125] Markus Krauss, Rolf Burghaus, Jörg Lippert, Mikko Niemi, Pertti Neuvonen, Andreas Schuppert, Stefan Willmann, Lars Kuepfer, and Linus Görlitz. Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification. In *Silico Pharmacology*, 1 :6, 2013.
- [126] L. Kuepfer, C. Niederalt, T. Wendl, J.-F. Schlender, S. Willmann, J. Lippert, M. Block, T. Eissing, and D. Teutonico. Applied Concepts in PBPK Modeling : How to Build a PBPK/PD Model. *CPT : pharmacometrics & systems pharmacology*, 5(10) :516–531, October 2016.
- [127] Lars Kuepfer. Towards whole-body systems physiology. *Molecular Systems Biology*, 6 :409, September 2010.
- [128] K. S. Kulp, M. G. Knize, N. D. Fowler, C. P. Salmon, and J. S. Felton. PhIP metabolites in human urine after consumption of well-cooked chicken. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 802(1) :143–153, March 2004.
- [129] K. S. Kulp, M. G. Knize, M. A. Malfatti, C. P. Salmon, and J. S. Felton. Identification of urine metabolites of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine following consumption of a single cooked chicken meal in humans. *Carcinogenesis*, 21(11) :2065–2072, November 2000.
- [130] J. W. Lampe, I. B. King, S. Li, M. T. Grate, K. V. Barale, C. Chen, Z. Feng, and J. D. Potter. Brassica vegetables increase and apiaceous vegetables decrease cytochrome P450 1a2 activity in humans : changes in caffeine metabolite ratios in response to controlled vegetable diets. *Carcinogenesis*, 21(6) :1157–1162, June 2000.
- [131] N. P. Lang, M. A. Butler, J. Massengill, M. Lawson, R. C. Stotts, M. Hauer-Jensen, and F. F. Kadlubar. Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501a2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 3(8) :675–682, December 1994.



- [132] S Langouët, D H Welti, N Kerriguy, L B Fay, T Huynh-Ba, J Markovic, F P Guengerich, A Guillouzo, and R J Turesky. Metabolism of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline in human hepatocytes : 2-amino-3-methylimidazo[4,5-f]quinoxaline-8-carboxylic acid is a major detoxification pathway catalyzed by cytochrome P450 1a2. *Chemical research in toxicology*, 14(2) :211–221, February 2001.
- [133] Sophie Langouët, Axel Paehler, Dieter H. Welti, Nathalie Kerriguy, André Guillouzo, and Robert J. Turesky. Differential metabolism of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine in rat and human hepatocytes. *Carcinogenesis*, 23(1) :115–122, January 2002.
- [134] L. Le Marchand, A. A. Franke, L. Custer, L. R. Wilkens, and R. V. Cooney. Lifestyle and nutritional correlates of cytochrome CYP1a2 activity : inverse associations with plasma lutein and alpha-tocopherol. *Pharmacogenetics*, 7(1) :11–19, February 1997.
- [135] L. Le Marchand, J. H. Hankin, L. R. Wilkens, L. M. Pierce, A. Franke, L. N. Kolonel, A. Seifried, L. J. Custer, W. Chang, A. Lum-Jones, and T. Donlon. Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1a2 phenotypes in increasing colorectal cancer risk. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 10(12) :1259–1266, December 2001.
- [136] Ethan Lee, Adrian Salic, Roland Krüger, Reinhart Heinrich, and Marc W. Kirschner. The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS biology*, 1(1) :E10, October 2003.
- [137] Jeremy A. Leonard, Yu-Mei Tan, Mary Gilbert, Kristin Isaacs, and Hisham El-Masri. Estimating Margin of Exposure to Thyroid Peroxidase Inhibitors Using High-Throughput in vitro Data, High-Throughput Exposure Modeling, and Physiologically Based Pharmacokinetic/Pharmacodynamic Modeling. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 151(1) :57–70, May 2016.
- [138] Jianing Li, Severin T. Schneebeli, Joseph Bylund, Ramy Farid, and Richard A. Friesner. IDSite : An accurate approach to predict P450-mediated drug metabolism. *Journal of Chemical Theory and Computation*, 7(11) :3829–3845, November 2011.
- [139] D. Lin, D. J. Meyer, B. Ketterer, N. P. Lang, and F. F. Kadlubar. Effects of human and rat glutathione S-transferases on the covalent DNA binding of the N-acetoxy derivatives of heterocyclic amine carcinogens in vitro : a possible mechanism of organ specificity in their carcinogenesis. *Cancer Research*, 54(18) :4920–4926, September 1994.

- [140] D J Liska. The detoxification enzyme systems. *Alternative medicine review : a journal of clinical therapeutic*, 3(3) :187–198, June 1998.
- [141] Daniel Machado, Rafael S. Costa, Miguel Rocha, Eugénio C. Ferreira, Bruce Tidor, and Isabel Rocha. Modeling formalisms in Systems Biology. *AMB Express*, 1 :45, December 2011.
- [142] Radhakrishnan Mahadevan, Jeremy S. Edwards, and Francis J. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, 83(3) :1331–1340, September 2002.
- [143] Thomas Maiwald and Jens Timmer. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, 24(18) :2037–2043, September 2008.
- [144] Dipanjali Majumdar, Chirasree Dutta, and Subha Sen. Inhalation exposure or body burden? Better way of estimating risk—An application of PBPK model. *Environmental Toxicology and Pharmacology*, 41 :54–61, January 2016.
- [145] M. A. Malfatti and J. S. Felton. N-glucuronidation of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and N-hydroxy-PhIP by specific human UDP-glucuronosyltransferases. *Carcinogenesis*, 22(7) :1087–1093, July 2001.
- [146] Michael A. Malfatti and James S. Felton. Human UDP-glucuronosyltransferase 1a1 is the primary enzyme responsible for the N-glucuronidation of N-hydroxy-PhIP in vitro. *Chemical Research in Toxicology*, 17(8) :1137–1144, August 2004.
- [147] S. Manabe, K. Tohyama, O. Wada, and T. Aramaki. Detection of a carcinogen, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), in cigarette smoke condensate. *Carcinogenesis*, 12(10) :1945–1947, October 1991.
- [148] Carol A. Marchant, Katharine A. Briggs, and Anthony Long. In silico tools for sharing data and knowledge on toxicity and metabolism : derek for windows, meteor, and vitic. *Toxicology Mechanisms and Methods*, 18(2-3) :177–187, 2008.
- [149] Y. Masubuchi, S. Hosokawa, T. Horie, T. Suzuki, S. Ohmori, M. Kitada, and S. Narimatsu. Cytochrome P450 isozymes involved in propranolol metabolism in human liver microsomes. The role of CYP2d6 as ring-hydroxylase and CYP1a2 as N-desisopropylase. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 22(6) :909–915, December 1994.
- [150] Y. Masubuchi, N. Kagimoto, S. Narimatsu, S. Fujita, and T. Suzuki. Regioselective contribution of the cytochrome P-450 2d subfamily to propranolol metabolism in rat liver microsomes. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 21(6) :1012–1016, December 1993.
- [151] T. Matsumoto, D. Yoshida, and H. Tomita. Determination of mutagens, amino-alpha-carbolines in grilled foods and cigarette smoke condensate. *Cancer Letters*, 12(1-2) :105–110, March 1981.

- [152] Ovanes G. Mekenyan, Sabcho D. Dimitrov, Todor S. Pavlov, and Gilman D. Veith. A systematic approach to simulating metabolism in computational toxicology. I. The TIMES heuristic modelling framework. *Current Pharmaceutical Design*, 10(11) :1273–1293, 2004.
- [153] Michaela Meyer, Sebastian Schneckener, Bernd Ludewig, Lars Kuepfer, and Joerg Lippert. Using expression data for quantification of active processes in physiologically based pharmacokinetic modeling. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 40(5) :892–901, May 2012.
- [154] Victor Moreno, Hansruedi Glatt, Elisabet Guino, Eva Fisher, Walter Meinel, Matilde Navarro, Josep M. Badosa, Heiner Boeing, and Bellvitge Colorectal Cancer Study Group. Polymorphisms in sulfotransferases SULT1a1 and SULT1a2 are not related to colorectal cancer. *International Journal of Cancer*, 113(4) :683–686, February 2005.
- [155] Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. AutoDock4 and AutoDockTools4 : Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16) :2785–2791, December 2009.
- [156] Kenneth C. Morton, Frederick A. Beland, Frederick E. Evans, Nancy F. Fullerton, and Fred F. Kadlubar. Metabolic Activation of N-Hydroxy-N,N'-diacetylbenzidine by Hepatic Sulfotransferase. *Cancer Research*, 40(3) :751–757, March 1980.
- [157] D. R. Mould and R. N. Upton. Basic concepts in population modeling, simulation, and model-based drug development. *CPT : pharmacometrics & systems pharmacology*, 1 :e6, September 2012.
- [158] Smithing M.p and Darvas F. HazardExpert : an expert system for predicting chemical toxicity. ACS Symposium series American Chemical Society, 1992.
- [159] E. Muckel, H. Frandsen, and H. R. Glatt. Heterologous expression of human N-acetyltransferases 1 and 2 and sulfotransferase 1a1 in *Salmonella typhimurium* for mutagenicity testing of heterocyclic amines. *Food and Chemical Toxicology : An International Journal Published for the British Industrial Biological Research Association*, 40(8) :1063–1068, August 2002.
- [160] M. Murkovic. Formation of heterocyclic aromatic amines in model systems. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 802(1) :3–10, March 2004.
- [161] Weijuan Ni, Lynn McNaughton, David M. LeMaster, Rashmi Sinha, and Robert J. Turesky. Quantitation of 13 heterocyclic aromatic amines in cooked beef, pork, and chicken by liquid chromatography-electrospray ionization/tandem mass spectrometry. *Journal of Agricultural and Food Chemistry*, 56(1) :68–78, January 2008.

- [162] S. Nowell and C. N. Falany. Pharmacogenetics of human cytosolic sulfotransferases. *Oncogene*, 25(11) :1673–1678, March 2006.
- [163] Susan Nowell, Brian Coles, Rashmi Sinha, Stewart MacLeod, D. Luke Ratnasinghe, Craig Stotts, Fred F. Kadlubar, Christine B. Ambrosone, and Nicholas P. Lang. Analysis of total meat intake and exposure to individual heterocyclic amines in a case-control study of colorectal cancer : contribution of metabolic variation to risk. *Mutation Research*, 506-507 :175–185, September 2002.
- [164] Won Seok Oh, Doo Nam Kim, Jihoon Jung, Kwang-Hwi Cho, and Kyoung Tai No. New combined model for the prediction of regioselectivity in cytochrome P450/3a4 mediated metabolism. *Journal of Chemical Information and Modeling*, 48(3) :591–601, March 2008.
- [165] Lars Olsen, Chris Oostenbrink, and Flemming Steen Jørgensen. Prediction of cytochrome P450 mediated metabolism. *Advanced Drug Delivery Reviews*, 86 :61–71, June 2015.
- [166] Zoltán N. Oltvai and Albert-László Barabási. Life's complexity pyramid. *Science*, 298(5594) :763–764, 2002.
- [167] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3) :245–248, March 2010.
- [168] A. Orzechowski, D. Schrenk, and K. W. Bock. Metabolism of 1- and 2-naphthylamine in isolated rat hepatocytes. *Carcinogenesis*, 13(12) :2227–2232, December 1992.
- [169] Fatih Oz and Mukerrem Kaya. Heterocyclic Aromatic Amines in Meat. *Journal of Food Processing and Preservation*, 35(6) :739–753, December 2011.
- [170] S. Ozawa, H. C. Chou, F. F. Kadlubar, K. Nagata, Y. Yamazoe, and R. Kato. Activation of 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b] pyridine by cDNA-expressed human and rat arylsulfotransferases. *Japanese Journal of Cancer Research : Gann*, 85(12) :1220–1228, December 1994.
- [171] P. Pais and M. G. Knize. Chromatographic and related techniques for the determination of aromatic heterocyclic amines in foods. *Journal of Chromatography. B, Biomedical Sciences and Applications*, 747(1-2) :139–169, September 2000.
- [172] E. T. Papoutsakis. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering*, 26(2) :174–187, February 1984.
- [173] Andrew Parkinson and Brian W. Ogilvie. Chapter 6. Biotransformation of Xenobiotics | Casarett & Doull's Essentials of Toxicology, 2e | AccessPharmacy | McGraw-Hill Medical. 2010.

- [174] G. Patlewicz, N. Jeliaskova, R. J. Safford, A. P. Worth, and B. Aleksiev. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR and QSAR in environmental research*, 19(5-6) :495–524, 2008.
- [175] Constantin Patrianakos and Dietrich Hoffmann. Chemical Studies on Tobacco Smoke LXIV. On the Analysis of Aromatic Amines in Cigarette Smoke. *Journal of Analytical Toxicology*, 3(4) :150–154, July 1979.
- [176] Megan L Peach, Alexey V Zakharov, Ruifeng Liu, Angelo Pugliese, Gregory Tawa, Anders Wallqvist, and Marc C Nicklaus. Computational tools and resources for metabolism-related property predictions. 1. Overview of publicly available (free and commercial) databases and software. *Future medicinal chemistry*, 4(15) :1907–1932, October 2012.
- [177] Anna Pelander, Elli Tyrkkö, and Ilkka Ojanperä. In silico methods for predicting metabolism and mass fragmentation applied to quetiapine in liquid chromatography/time-of-flight mass spectrometry urine drug screening. *Rapid communications in mass spectrometry : RCM*, 23(4) :506–514, February 2009.
- [178] Przemyslaw Piechota, Mark T. D. Cronin, Mark Hewitt, and Judith C. Madden. Pragmatic approaches to using computational methods to predict xenobiotic metabolism. *Journal of Chemical Information and Modeling*, 53(6) :1282–1293, June 2013.
- [179] Simulation Plus. ADMET Predictor (Simulations Plus Inc). 2014.
- [180] Nathan D. Price, Jason A. Papin, Christophe H. Schilling, and Bernhard O. Palsson. Genome-scale microbial in silico models : the constraints-based approach. *Trends in Biotechnology*, 21(4) :162–169, April 2003.
- [181] J. J. Pritchett, R. K. Kuester, and I. G. Sipes. Metabolism of bisphenol a in primary cultured hepatocytes from mice, rats, and humans. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 30(11) :1180–1185, November 2002.
- [182] Arwa B. Raies and Vladimir B. Bajic. In silico toxicology : computational methods for the prediction of chemical toxicity. *Wiley Interdisciplinary Reviews. Computational Molecular Science*, 6(2) :147–172, March 2016.
- [183] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics (Oxford, England)*, 25(15) :1923–1929, August 2009.
- [184] Jennifer L. Reed and Bernhard Ø Palsson. Thirteen years of building constraint-based in silico models of Escherichia coli. *Journal of Bacteriology*, 185(9) :2692–2699, May 2003.

- [185] Lars Ridder and Markus Wagener. SyGMA : combining expert knowledge and empirical scoring in the prediction of metabolites. *ChemMedChem*, 3(5) :821–832, May 2008.
- [186] Ivonne M. C. M. Rietjens, Jochem Louisse, and Ans Punt. Tutorial on physiologically based kinetic modeling in molecular nutrition and food research. *Molecular Nutrition & Food Research*, 55(6) :941–956, June 2011.
- [187] Michał Rostkowski, Ola Spjuth, and Patrik Rydberg. WhichCyp : prediction of cytochromes P450 inhibition. *Bioinformatics*, 29(16) :2051–2052, August 2013.
- [188] P. Roy, O. Tretyakov, J. Wright, and D. J. Waxman. Stereoselective metabolism of ifosfamide by human P-450s 3a4 and 2b6. Favorable metabolic properties of R-enantiomer. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 27(11) :1309–1318, November 1999.
- [189] Anastasia Rudik, Alexander Dmitriev, Alexey Lagunin, Dmitry Filimonov, and Vladimir Poroikov. SOMP : web server for in silico prediction of sites of metabolism for drug-like compounds. *Bioinformatics (Oxford, England)*, 31(12) :2046–2048, June 2015.
- [190] Anastasia V. Rudik, Vladislav M. Bezhentsev, Alexander V. Dmitriev, Dmitry S. Druzhilovskiy, Alexey A. Lagunin, Dmitry A. Filimonov, and Vladimir V. Poroikov. MetaTox : Web Application for Predicting Structure and Toxicity of Xenobiotics' Metabolites. *Journal of Chemical Information and Modeling*, 57(4) :638–642, April 2017.
- [191] Anastasia V. Rudik, Alexander V. Dmitriev, Alexey A. Lagunin, Dmitry A. Filimonov, and Vladimir V. Poroikov. Metabolism site prediction based on xenobiotic structural formulas and PASS prediction algorithm. *Journal of Chemical Information and Modeling*, 54(2) :498–507, February 2014.
- [192] Patrik Rydberg, David E. Gloriam, and Lars Olsen. The SMARTCyp cytochrome P450 metabolism prediction server. *Bioinformatics (Oxford, England)*, 26(23) :2988–2989, December 2010.
- [193] Patrik Rydberg, David E. Gloriam, Jed Zaretski, Curt Breneman, and Lars Olsen. SMARTCyp : A 2d Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS medicinal chemistry letters*, 1(3) :96–100, June 2010.
- [194] Jennifer E. Sager, Jingjing Yu, Isabelle Ragueneau-Majlessi, and Nina Isoherranen. Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation Approaches : A Systematic Review of Published Models, Applications, and Model Verification. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 43(11) :1823–1837, November 2015.

- [195] K. Saito, Y. Yamazoe, T. Kamataki, and R. Kato. Activation and detoxication of N-hydroxy-Trp-P-2 by glutathione and glutathione transferases. *Carcinogenesis*, 4(12) :1551–1557, December 1983.
- [196] Romeil Sandhu, Tryphon Georgiou, Ed Reznik, Liangjia Zhu, Ivan Kolesov, Yasin Senbabaoglu, and Allen Tannenbaum. Graph Curvature for Differentiating Cancer Networks. *Scientific Reports*, 5 :12323, July 2015.
- [197] Jan-Frederik Schlender, Michaela Meyer, Kirstin Thelen, Markus Krauss, Stefan Willmann, Thomas Eissing, and Ulrich Jaehde. Development of a Whole-Body Physiologically Based Pharmacokinetic Approach to Assess the Pharmacokinetics of Drugs in Elderly Individuals. *Clinical Pharmacokinetics*, 55(12) :1573–1589, December 2016.
- [198] H. A. Schut, F. B. Daniel, K. M. Schenck, T. R. Loeb, and G. D. Stoner. Metabolism and DNA adduct formation of 2-acetylaminofluorene by bladder explants from human, dog, monkey, hamster and rat. *Carcinogenesis*, 5(10) :1287–1292, October 1984.
- [199] H. A. Schut and E. G. Snyderwine. DNA adducts of heterocyclic amine food mutagens : implications for mutagenesis and carcinogenesis. *Carcinogenesis*, 20(3) :353–368, March 1999.
- [200] H. Schweikl, J. A. Taylor, S. Kitareewan, P. Linko, D. Nagorney, and J. A. Goldstein. Expression of CYP1a1 and CYP1a2 genes in human liver. *Pharmacogenetics*, 3(5) :239–249, October 1993.
- [201] Tomer Shlomi, Moran N. Cabili, and Eytan Ruppin. Predicting metabolic biomarkers of human inborn errors of metabolism. *Molecular Systems Biology*, 5 :263, 2009.
- [202] Suresh B. Singh, Lucy Q. Shen, Matthew J. Walker, and Robert P. Sheridan. A model for predicting likely sites of CYP3a4-mediated metabolism on drug-like molecules. *Journal of Medicinal Chemistry*, 46(8) :1330–1336, April 2003.
- [203] R Sinha, N Rothman, S.D. Mark, S Murray, E.D. Brown, O.A. Levander, D S Davies, F F Kadlubar, and R.N. Hoover. Lower levels of urinary 2-amino-3,8-dimethylimidazo[4,5-f]-quinoxaline (MelQx) in humans with higher CYP1a2. *Carcinogenesis*, 16(11) :2859–2861, 1995.
- [204] Keng Cher Soh and Vassily Hatzimanikatis. DREAMS of metabolism. *Trends in Biotechnology*, 28(10) :501–508, October 2010.
- [205] Mariana C. Sousa, Rodolpho C. Braga, Bertilha A.S. Cintra, Valéria de Oliveira, and Carolina H. Andrade. In silico metabolism studies of dietary flavonoids by cyp1a2 and cyp2c9. *Food Research International*, 50(1) :102 – 110, 2013.

- [206] M. Steiner, M. Bastian, W. A. Schulz, T. Pulte, K. H. Franke, A. Röhring, J. M. Wolff, H. Seiter, and P. Schuff-Werner. Phenol sulphotransferase *SULT1a1* polymorphism in prostate cancer : lack of association. *Archives of Toxicology*, 74(4-5) :222–225, July 2000.
- [207] Ralf Steuer. Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry*, 68(16-18) :2139–2151, September 2007.
- [208] W. G. Stillwell, R. J. Turesky, R. Sinha, and S. R. Tannenbaum. N-oxidative metabolism of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline (MeIQx) in humans : excretion of the N2-glucuronide conjugate of 2-hydroxyamino-MeIQx in urine. *Cancer Res.*, 59(20) :5154–5159, Oct 1999.
- [209] Shana J. Sturla and Paul Hollenberg. Systems toxicology : a special issue. *Chemical Research in Toxicology*, 27(3) :311, March 2014.
- [210] T. Sugimura. Mutagens, carcinogens, and tumor promoters in our daily food. *Cancer*, 49(10) :1970–1984, May 1982.
- [211] T. Sugimura. Overview of carcinogenic heterocyclic amines. *Mutation Research*, 376(1-2) :211–219, May 1997.
- [212] T Sugimura, N Nagao, T Kawachi, M Honda, T Yahagi, Y Seino, S Stao, N Matsukura, T Matsushima, A Shirai, M Sawamura, and H Matsumoto. Mutagen-carcinogens in food, with special reference to highly mutagenic pyrolytic products in broiled foods. In : Hiatt HH, Watson JD, Winstein JA, editors. *Origins of Human Cancer, Book C*. Cold Spring Harbor Laboratory ; Cold Spring Harbor, NY : 1977. pp. 1561–1577. 1977.
- [213] Takashi Sugimura, Keiji Wakabayashi, Hitoshi Nakagama, and Minako Nagao. Heterocyclic amines : Mutagens/carcinogens produced during cooking of meat and fish. *Cancer Science*, 95(4) :290–299, April 2004.
- [214] J. Talafous, L. M. Sayre, J. J. Mieyal, and G. Klopman. META. 2. A dictionary model of mammalian xenobiotic metabolism. *Journal of Chemical Information and Computer Sciences*, 34(6) :1326–1333, December 1994.
- [215] Yijin Tang, Fekadu Kassie, Xuemin Qian, Buzayew Ansha, and Robert J. Turesky. DNA adduct formation of 2-amino-9h-pyrido[2,3-b]indole and 2-amino-3,4-dimethylimidazo[4,5-f]quinoline in mouse liver and extrahepatic tissues during a subchronic feeding study. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, 133(2) :248–258, June 2013.
- [216] Ákos Tarcsay and György M. Keseru. In silico site of metabolism prediction of cytochrome P450-mediated biotransformations. *Expert Opinion on Drug Metabolism & Toxicology*, 7(3) :299–312, March 2011.



- [217] Akos Tarcsay, Róbert Kiss, and György M. Keseru. Site of metabolism prediction on cytochrome P450 2c9 : a knowledge-based docking approach. *Journal of Computer-Aided Molecular Design*, 24(5) :399–408, May 2010.
- [218] T. Teorell. STUDIES ON THE DIFFUSION EFFECT UPON IONIC DISTRIBUTION : II. EXPERIMENTS ON IONIC ACCUMULATION. *The Journal of General Physiology*, 21(1) :107–122, September 1937.
- [219] Christoph Thiel, Sebastian Schneckener, Markus Krauss, Ahmed Ghallab, Ute Hofmann, Tobias Kanacher, Sebastian Zellmer, Rolf Gebhardt, Jan G. Hengstler, and Lars Kuepfer. A systematic evaluation of the use of physiologically based pharmacokinetic modeling for cross-species extrapolation. *Journal of Pharmaceutical Sciences*, 104(1) :191–206, January 2015.
- [220] Edine W. Tiemersma, Dorien W. Voskuil, Annelies Bunschoten, Elbert A. Hogenboom, Ben J. M. Witteman, Fokko M. Nagengast, HansRuedi Glatt, Frans J. Kok, and Ellen Kampman. Risk of colorectal adenomas in relation to meat consumption, meat preparation, and genetic susceptibility in a Dutch population. *Cancer causes & control : CCC*, 15(3) :225–236, April 2004.
- [221] John A. Timbrell and Timothy C. Marrs. Biotransformation of Xenobiotics. In *General, Applied and Systems Toxicology*. John Wiley & Sons, Ltd, 2009. DOI : 10.1002/9780470744307.gat004.
- [222] Roberto Todeschini and Viviana Consonni. *Handbook of Molecular Descriptors*. Wiley-VCH Verlag GmbH, 2000. DOI : 10.1002/9783527613106.ch1a.
- [223] F. Toribio, M. T. Galceran, and L. Puignou. Separation of heteroaromatic amines in food products. *Journal of Chromatography. B, Biomedical Sciences and Applications*, 747(1-2) :171–202, September 2000.
- [224] Masood Khaksar Toroghi, William R Cluett, and Radhakrishnan Mahadevan. A multi-scale model of the whole human body based on dynamic parsimonious flux balance analysis. *IFAC-PapersOnLine*, 49(7) :937 – 942, 2016. 11th IFAC Symposium on Dynamics and Control of Process Systems Including Biosystems DYCOPS-CAB 2016.
- [225] T. Totsuka, R. Nishigaki, Sugimura, T., and Wakabayashi K. The possible involvement of mutagenic and carcinogenic heterocyclic amines in human cancer. in : Skog k, alexander j, editors. *acrylamide and other hazardous compounds in heat-treated foods*. woodhead publisher ; boca raton, fl : 2006. pp. 296–515. *Carcinogenesis*, pages 296–515, 2006.
- [226] R. J. Turesky. Metabolism and biodisposition of heterocyclic amines. *Prog. Clin. Biol. Res.*, 347 :39–53, 1990.

- [227] R. J. Turesky. The role of genetic polymorphisms in metabolism of carcinogenic heterocyclic aromatic amines. *Current Drug Metabolism*, 5(2) :169–180, April 2004.
- [228] R. J. Turesky, A. Constable, J. Richoz, N. Varga, J. Markovic, M. V. Martin, and F. P. Guengerich. Activation of heterocyclic aromatic amines by rat and human liver microsomes and by purified rat and human cytochrome P450 1a2. *Chemical Research in Toxicology*, 11(8) :925–936, August 1998.
- [229] R. J. Turesky, R. C. Garner, D. H. Welte, J. Richoz, S. H. Leveson, K. H. Dingley, K. W. Turteltaub, and L. B. Fay. Metabolism of the food-borne mutagen 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline in humans. *Chemical Research in Toxicology*, 11(3) :217–225, March 1998.
- [230] R. J. Turesky, N. P. Lang, M. A. Butler, C. H. Teitel, and F. F. Kadlubar. Metabolic activation of carcinogenic heterocyclic aromatic amines by human liver and colon. *Carcinogenesis*, 12(10) :1839–1845, October 1991.
- [231] R. J. Turesky, V. Parisod, T. Huynh-Ba, S. Langouët, and F. P. Guengerich. Regioselective differences in C(8)- and N-oxidation of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline by human and rat liver microsomes and cytochromes P450 1a2. *Chemical Research in Toxicology*, 14(7) :901–911, July 2001.
- [232] Robert J. Turesky. Formation and biochemistry of carcinogenic heterocyclic aromatic amines in cooked meats. *Toxicology Letters*, 168(3) :219–227, February 2007.
- [233] Robert J. Turesky, F. Peter Guengerich, André Guillouzo, and Sophie Langouët. Metabolism of heterocyclic aromatic amines by human hepatocytes and cytochrome P4501a2. *Mutation Research*, 506-507 :187–195, September 2002.
- [234] Robert J. Turesky and Loic Le Marchand. Metabolism and biomarkers of heterocyclic aromatic amines in molecular epidemiology studies : lessons learned from aromatic amines. *Chemical Research in Toxicology*, 24(8) :1169–1214, August 2011.
- [235] A. L. Upthagrove and W. L. Nelson. Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol analogs : metabolism by CYP1A2. *Drug Metab. Dispos.*, 29(11) :1389–1395, Nov 2001.
- [236] Suzanne J. P. L. van den Berg, Ans Punt, Ans E. M. F. Soffers, Jacques Vervoort, Stephen Ngeleja, Bert Spenklink, and Ivonne M. C. M. Rietjens. Physiologically based kinetic models for the alkenylbenzene elemicin in rat and human and possible implications for risk assessment. *Chemical Research in Toxicology*, 25(11) :2352–2367, November 2012.

- [237] Alejandro F. Villaverde and Julio R. Banga. Reverse engineering and identification in systems biology : strategies, perspectives and challenges. *Journal of the Royal Society, Interface / the Royal Society*, 11(91) :20130505, February 2014.
- [238] K. Vistisen, H. E. Poulsen, and S. Loft. Foreign compound metabolism capacity in man measured from metabolites of dietary caffeine. *Carcinogenesis*, 13(9) :1561–1568, September 1992.
- [239] Ludwig Von Bertalanffy. *General System Theory : Foundations, Development, Applications*. George Braziller, New york, 1968.
- [240] David G. Walters, Philip J. Young, Cynthia Agus, Mark G. Knize, Alan R. Boobis, Nigel J. Gooderham, and Brian G. Lake. Cruciferous vegetable consumption alters the metabolism of the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) in humans. *Carcinogenesis*, 25(9) :1659–1669, September 2004.
- [241] M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner, and C. von Mering. PaxDb, a database of protein abundance averages across all three domains of life. *Molecular & cellular proteomics : MCP*, 11(8) :492–500, August 2012.
- [242] David Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1) :31–36, February 1988.
- [243] E. K. Weisburger and J. H. Weisburger. Chemistry, carcinogenicity, and metabolism of 2-fluorenamine and related compounds. *Advances in Cancer Research*, 5 :331–431, 1958.
- [244] David W. Weisgerber. Chemical Abstracts Service Chemical Registry System : History, scope, and impacts. *Journal of the American Society for Information Science*, 48(4) :349–360, April 1997.
- [245] E. M. P. Widmark. Presence of Cancer-producing Substances in Roasted Food : Abstract : *Nature*. *Nature*, 143 :984–984, June 1939.
- [246] Egon L. Willighagen, John W. Mayfield, Jonathan Alvarsson, Arvid Berg, Lars Carlsson, Nina Jeliakova, Stefan Kuhn, Tomáš Pluskal, Miquel Rojas-Chertó, Ola Spjuth, Gilleain Torrance, Chris T. Evelo, Rajarshi Guha, and Christoph Steinbeck. The Chemistry Development Kit (CDK) v2.0 : atom typing, depiction, molecular formulas, and substructure searching. *Journal of Cheminformatics*, 9 :33, June 2017.
- [247] Stefan Willmann, Karsten Höhn, Andrea Edginton, Michael Sevestre, Juri Solodenko, Wolfgang Weiss, Jörg Lippert, and Walter Schmitt. Development of a physiology-based whole-body population model for assessing the influence of

- individual variability on the pharmacokinetics of drugs. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(3) :401–431, June 2007.
- [248] David S. Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. DrugBank : a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, 36(Database issue) :D901–906, January 2008.
- [249] David S. Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. DrugBank : a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(Database issue) :D668–672, January 2006.
- [250] Olaf Wolkenhauer, Charles Auffray, Olivier Brass, Jean Clairambault, Andreas Deutsch, Dirk Drasdo, Francesco Gervasio, Luigi Preziosi, Philip Maini, Anna Marciniak-Czochra, Christina Kossow, Lars Kuepfer, Katja Rateitschak, Ignacio Ramis-Conde, Benjamin Ribba, Andreas Schuppert, Rod Smallwood, Georgios Stamatakos, Felix Winter, and Helen Byrne. Enabling multiscale modeling in systems medicine. *Genome Medicine*, 6(3) :21, 2014.
- [251] Martin Kin Lok Wong, James Robert Krycer, James Geoffrey Burchfield, David Ernest James, and Zdenka Kuncic. A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems. *FEBS open bio*, 5 :226–239, 2015.
- [252] Gang Xie, Chi C Wong, Ka-Wing Cheng, Liqun Huang, Panayiotis Constantinides, and Basil Rigas. Regioselective oxidation of phospho-nsaids by human cytochrome p450 and flavin monooxygenase isoforms : Implications for their pharmacokinetic properties and safety. 167 :222–32, 04 2012.
- [253] Y Yamazoe and K Nagata. In vitro metabolism. In : Sugimura T, Nagao M, editors. Food Borne Carcinogens Heterocyclic Amines. John Wiley & Sons Ltd. ; Chichester, England : 2000. pp. 74–89. 2000.
- [254] Y. Yamazoe, K. Nagata, S. Ozawa, D. W. Gong, and R. Kato. Activation and detoxication of carcinogenic arylamines by sulfation. *Princess Takamatsu Symposia*, 23 :154–162, 1995.
- [255] C. C. Yang, S. N. Jenq, and H. Lee. Characterization of the carcinogen 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline in cooking aerosols under domestic conditions. *Carcinogenesis*, 19(2) :359–363, February 1998.
- [256] Xiaoxia Yang, Daniel R. Doerge, Justin G. Teeguarden, and Jeffrey W. Fisher. Development of a physiologically based pharmacokinetic model for assessment of human exposure to bisphenol A. *Toxicology and Applied Pharmacology*, 289(3) :442–456, December 2015.

- [257] Mona Yousofshahi, Sara Manteiga, Charmian Wu, Kyongbum Lee, and Soha Hassoun. PROXIMAL : a method for Prediction of Xenobiotic Metabolism. *BMC systems biology*, 9 :94, December 2015.
- [258] Rae Yuan, Soraya Madani, Xiao-Xiong Wei, Kellie Reynolds, and Shiew-Mei Huang. Evaluation of cytochrome P450 probe substrates commonly used by the pharmaceutical industry to study in vitro drug interactions. *Drug Metabolism and Disposition : The Biological Fate of Chemicals*, 30(12) :1311–1319, December 2002.
- [259] Jed Zaretski, Charles Bergeron, Patrik Rydberg, Tao-wei Huang, Kristin P. Bennett, and Curt M. Breneman. RS-predictor : a new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3a4. *Journal of Chemical Information and Modeling*, 51(7) :1667–1689, July 2011.
- [260] Jed Zaretski, Matthew Matlock, and S. Joshua Swamidass. XenoSite : accurately predicting CYP-mediated sites of metabolism with neural networks. *Journal of Chemical Information and Modeling*, 53(12) :3373–3383, December 2013.
- [261] Veronica Zelesky, Richard Schneider, John Janiszewski, Ismael Zamora, James Ferguson, and Matthew Troutman. Software automation tools for increased throughput metabolic soft-spot identification in early drug discovery. *Bioanalysis*, 5(10) :1165–1179, May 2013.
- [262] Xiaomei Zhuang and Chuang Lu. PBPK modeling and simulation in drug research and development. *Acta Pharmaceutica Sinica. B*, 6(5) :430–440, September 2016.

# Table des figures

1.1	Pyramide de la complexité de la vie autour de laquelle les problématiques en biologie des systèmes s'articulent (adaptée de [67]) . . . . .	13
1.2	Les différents niveaux organisationnels d'étude de la biologie des systèmes [1] . . . . .	14
1.3	Construction d'une chaîne SMILES pour décrire une structure 2D d'une molécule chimique . . . . .	17
1.4	Formulation de la molécule MeIQx avec les principaux identifiants moléculaires biologiques et informatiques . . . . .	19
1.5	Equation bilan de l'hydroxylation du MeIQx catalysé par l'enzyme cytochrome P450 (CYP) . . . . .	20
1.6	Evolution énergétique d'une réaction biochimique . . . . .	21
1.7	Application d'une réaction chimique grâce à l'utilisation du langage SMIRKS . . . . .	23
1.8	Les différentes approches de modélisation en fonction de leur échelle d'étude . . . . .	24
1.9	Modélisation PBPK . . . . .	26
1.10	Modélisation à l'échelle cellulaire (adaptée de [207]) . . . . .	28
1.11	Définition d'un problème FBA [167] . . . . .	30
1.12	Prédiction des métabolites . . . . .	34
1.13	Les trois approches majeures pour la prédiction <i>in silico</i> de la toxicité (adaptée de [11, 93]) . . . . .	36
2.1	Métabolisme des carcinogènes chimiques . . . . .	41
2.2	Voies de bioactivation des AHA . . . . .	45
2.3	Métabolisme d'A $\alpha$ C . . . . .	47
2.4	Métabolisme du MeIQx . . . . .	48
2.5	Métabolisme du PhIP . . . . .	49
2.6	Synthèse des approches de modélisation aux différentes échelles d'étude	55

3.1	Pipeline pour la prédiction des voies métaboliques des xénobiotiques et l'identification des métabolites réactifs à l'ADN . . . . .	61
3.2	Evaluation des ratios décrivant le nombre de métabolites restants après filtrations et le nombre de métabolites expérimentalement identifiés présents dans la première base de données en fonction des différents scores . . . . .	65
3.3	Prédiction du métabolisme d'A $\alpha$ C par le pipeline . . . . .	69
3.4	Prédiction du métabolisme du MeIQx par le pipeline . . . . .	70
3.5	Prédiction du métabolisme du PhIP par le pipeline . . . . .	71
3.6	Exemple de la prédiction individuelle des SOM par Way2Drug SOMP et Xenosite Metabolism 1.0 . . . . .	75
4.1	Réactivité avec l'ADN des métabolites résultant de la combinaison des réactions d'Hydroxylation (Hy), de Glucuronidation (Glc), d'Acétylation (Ac), de Sulfatation (Su) et d'Oxydation (Ox) appliquée au métabolisme des 30 AHA . . . . .	83
4.2	Interface de contrôle pour la génération de graphe pour une AHA sélectionnée par l'utilisateur . . . . .	86
4.3	Interface d'exploration du graphe généré . . . . .	87
4.4	Interface d'exploration en détail d'un nœud . . . . .	88
4.5	Exemple de génération d'un graphe pour A $\alpha$ C avec une profondeur de 1 et les information de réactivité et des types de réactions . . . . .	89
4.6	Graphe métabolique du MeIQx prédit avec deux niveaux de profondeur (extrait de l'interface web) . . . . .	90
5.1	Métabolisme du MeIQx (adaptée de [132]) . . . . .	96
5.2	Distribution des métabolites du MeIQx en fonction de sa concentration initiale dans les hépatocytes humains (adaptée de [132]) . . . . .	97
5.3	Modèle réduit du métabolisme du MeIQx . . . . .	100
5.4	Un modèle générique des compétitions enzymatiques au sein de réactions catalysées par la même enzyme . . . . .	101
5.5	Variation des paramètres estimés . . . . .	108
5.6	Docking du MeIQx avec le CYP1A2 par l'utilisation du logiciel Swiss-Dock . . . . .	109
5.7	Métabolisme du C-Hydroxy-MeIQx pour des concentrations initiales de MeIQx de 0.05, 0.1, 0.5 $\mu$ M après 5, 6, 7, 8, 9, 10h d'exposition . . . . .	110
5.8	Métabolisme du HON-MeIQx-N <sup>2</sup> -Gl pour des concentrations initiales de 0.5 (A), 1 (B), 1.5 (C) $\mu$ M après 12h d'exposition . . . . .	112
6.1	Biotransformation du MeIQx en produits de bioactivation (rouge) et en produits de détoxification (vert) pour des concentrations initiales différentes de MeIQx . . . . .	118

6.2	<b>Variation du coefficient de distribution <math>\alpha</math> du CYP1A2 (A) et du coefficient de distribution <math>\beta</math> d'UGT (B) modélisés par une fonction sigmoïde dépendante de la concentration du MeIQx . . . . .</b>	119
6.3	<b>Evolution du ratio Bioactivation/Detoxification en fonction de la concentration de MeIQx . . . . .</b>	121
6.4	<b>Analyse sensitive 2D réalisée à l'aide de PottersWheel . . . . .</b>	123
6.5	<b>Evaluation de la formation des produits de bioactivation et de détoxification en fonction des changements de concentrations totales de CYP1A2 . . . . .</b>	124





## Publications

**Article de journal** Victorien Delannée, Sophie Langouët, Nathalie Théret, and Anne Siegel. A modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes. *PeerJ*, 5 :e3703, September 2017.

**Poster** Victorien Delannée, Sophie Langouët, Nathalie Théret, and Anne Siegel. In silico prediction of Heterocyclic Aromatic Amines metabolism susceptible to form new DNA adducts in humans. *14th European ISSX Meeting*, June 2017.

**Article de journal** *En soumission imminente à Cheminformatics*. Victorien Delannée, Sophie Langouët, Nathalie Théret, and Anne Siegel. In silico prediction of Heterocyclic Aromatic Amines metabolism susceptible to form new DNA adducts in humans.



# Annexes

Le chapitre 3 présente une méthode pour la construction automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions. Cette méthode est basée sur l'implémentation de différentes approches informatiques pour la prédiction des réactions métaboliques, les sites de métabolisme (SOMs) et les sites de réactivité (SORs). Un des enjeux de l'approche développée est la construction des graphes métaboliques avec une forte précision, tout en ne modifiant pas la sensibilité. Ainsi, une optimisation des paramètres des différentes méthodes a été effectuée, basée sur l'apprentissage de paramètres à partir de molécules expérimentalement validées, contenues dans des bases de données.

La première base de données d'apprentissage nommée "base de données d'optimisation des paramètres pour la prédiction des SOMs" consiste en une liste curée manuellement de 15 molécules contenant des amines aromatiques extraites de la littérature et subissant des réactions d'hydroxylation et/ou de glucuronidation (voir Annexe Table 1). Ces 15 molécules sont impliquées dans 49 réactions de 5 types différents (C-Hydroxylation, N-Hydroxylation, C-Glucuronidation, N-Glucuronidation et O-Glucuronidation). Ces types de réactions représentent celles qui sont majoritairement subies par les différentes amines aromatiques. Cette base de données est utilisée pour calibrer les paramètres des méthodes de prédiction des SOMs pour CYP et UGT.

Name	IUPAC Name	Hydrox- num- ber	Glucur- num- ber	References
2-Acetylaminofluorene (2-AAF)	N-(9H-fluoren-2-yl)acetamide	8	1	[18, 39, 198, 243, 3]
2-Acetylaminophenanthrene	N-phenanthren-2-ylacetamide	1	0	[18]
2-Aminofluorene (2-AF)	9H-fluoren-2-amine	8	1	[18, 12, 39, 81, 243]

3,2'-Dimethyl-4-aminobiphenyl	2-methyl-4-(2-methylphenyl)aniline	0	1	[18, 173]
2-Naphthylamine (2-NA)	naphthalen-2-amine	2	2	[18, 108, 81, 168]
4-Acetylamino-4-aminobiphenyl (4-ABP)	N-[4-[(E)-2-phenylethenyl]phenyl]acetamide	1	0	[18]
17-beta-Estradiol	4-phenylaniline	1	1	[18, 108, 234, 32]
Benzidine (Bz)	(8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decahydrocyclopenta[a]phenanthrene-3,17-diol	0	2	
Debrisoquine	4-(4-aminophenyl)aniline	3	3	[18, 68, 69, 156]
Dobutamine	3,4-dihydro-1H-isoquinoline-2-carboximidamide	7	0	[58, 91, 221]
Desacetylcinobufagin	4-[2-[4-(4-hydroxyphenyl)butan-2-ylamino]ethyl]benzene-1,2-diol	0	3	[10]
Morphine	(3 $\beta$ ,9 $\xi$ ,15 $\beta$ ,16 $\beta$ )-3,16-Dihydroxy-14,15-epoxybupafenone	0	2	[102]
N-Methyl-4-aminoazobenzene	(4R,4aR,7S,7aR,12bS)-3-methyl-2,4,4a,7,7a,13-hexahydro-1H-4,12-methanobenzofuro[3,2-e]isoquinoline-7,9-diol	0	2	[173]
Propofol	N-methyl-4-[(3-methylphenyl)diazanyl]aniline	2	0	[18]
Propranolol	2,6-di(propan-2-yl)phenol	0	1	[173]
S-Oxazepam	1-naphthalen-1-yloxy-3-(propan-2-ylamino)propan-2-ol	3	4	[17, 55]
Tripelennamine	(3S)-7-chloro-3-hydroxy-5-phenyl-1,3-dihydro-1,4-benzodiazepin-2-one	0	1	[173]
	N'-benzyl-N,N-dimethyl-N'-pyridin-2-ylethane-1,2-diamine	0	1	[173]
<b>Total</b>		<b>36</b>	<b>23</b>	

**TABLE 1 – Base de données 1 : Liste des molécules utilisées pour optimiser les scores des outils associés à la prédiction des SOMs pour CYP et UGT**

Le métabolisme intégral de ces 15 molécules contenant 59 réactions est utilisé pour l'optimisation des scores

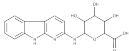
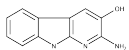
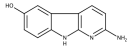
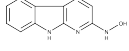
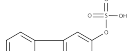

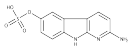
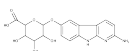
La deuxième base de données, nommée "base de données d'optimisation des paramètres pour l'identification des adduits à l'ADN" est une liste de 30 amines aromatiques curées manuellement et expérimentalement validées comme inductrices d'adduits à l'ADN (voir Annexe Table 2). Cette base de données est utilisée pour définir les différents seuils associés aux prédictions réalisées par Xenosite Reactivity 2.0 afin de classer les métabolites en fonction de leurs réactivités vis-à-vis de l'ADN.

Name	IUPAC Name	DNA Reactivity probability	References
N-Hydroxyl-IQ	N-{3-methyl-3H-imidazo[4,5-f]quinolin-2-yl}hydroxylamine	0.96	[232]
N-Acetoxy-IQ	{3-methyl-3H-imidazo[4,5-f]quinolin-2-yl}amino acetate	0.91	[232]
N-Sulfonyloxy-IQ	[{(3-methyl-3H-imidazo[4,5-f]quinolin-2-yl)amino}oxy]sulfonic acid	0.91	[232]
N-Hydroxyl-2-NA	N-(naphthalen-2-yl)hydroxylamine	0.91	[18]
N-Acetoxy-2-NA	(naphthalen-2-yl)amino acetate	0.91	[18]
N-Sulfonloxy-2-NA	{[(naphthalen-2-yl)amino]oxy} sulfonic acid	0.92	[18]
N-Glucuronosyloxy-2-NA	3,4,5-trihydroxy-6-[[{(naphthalen-2-yl)amino]oxy]oxane-2-carboxylic acid	0.80	[18, 108]
N'-Hydroxy-N-Acetyl-Bz	N-[4'-(hydroxyamino)-[1,1'-biphenyl]-4-yl]acetamide	0.87	[18]
N'-Acetoxy-N-Acetyl-Bz	4'-acetamido-[1,1'-biphenyl]-4-ylamino acetate	0.72	[18]
N'-Sulfonyloxy-N-Acetyl-Bz	[(4'-acetamido-[1,1'-biphenyl]-4-ylamino)oxy]sulfonic acid	0.92	[18]
N-Hydroxyl-4-ABP	N-[[1,1'-biphenyl]-4-yl]hydroxylamine	0.89	[18, 234]

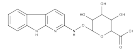
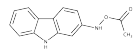
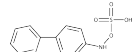
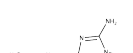
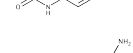
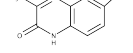
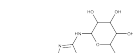

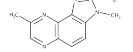
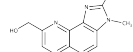
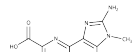
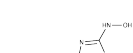
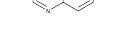
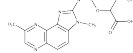
N-Acetoxy-4-ABP	N-acetyl-N-[1,1'-biphenyl]-4-ylacetamide	0.78	[18, 234]
N-Acetoxy-N-Acetyl-4-ABP	[1,1'-biphenyl]-4-ylamino acetate	0.22	[18, 234]
N-Sulfonyloxy-N-Acetyl-4-ABP	1-([1,1'-biphenyl]-4-yl)(sulfo)amino)ethan-1-one	0.36	[18, 234]
N-Hydroxyl-AF	N-(9H-fluoren-2-yl)hydroxylamine	0.90	[18, 198, 243, 3]
N-Sulfonyloxy-AF	{[(9H-fluoren-2-yl)amino]oxy)sulfonic acid	0.92	[18, 3]
N-Acetoxy-AF	(9H-fluoren-2-yl)amino acetate	0.80	[18, 3]
N-Glucuronyloxy-AF	6-[(9H-fluoren-2-yl)amino]oxy-3,4,5-trihydroxyoxane-2-carboxylic acid	0.78	[18, 3]
N-Sulfonyloxy-AAF	acetylaminofluorene-N-sulfate	0.40	[18, 3]
N-Glucuronyloxy-AAF	6-[[N-(9H-fluoren-2-yl)acetamido]oxy]-3,4,5-trihydroxyoxane-2-carboxylic acid	0.09	[18]
N-Hydroxyl-1-aminopyrene	N-(pyren-1-yl)hydroxylamine	0.93	[18]
N-Hydroxyl-6-aminochrysene	N-(chrysen-6-yl)hydroxylamine	0.93	
N-Acetoxy-4-Acetylaminostilbene	N-{4-[(E)-2-phenylethenyl]phenyl}acetamido acetate	0.41	[18]
N-Hydroxyl-2-Acetylaminophenanthrene	N-(phenanthren-2-yl)hydroxylamine	0.91	[18]
N-Acetoxy-2-Acetylaminophenanthrene	N-(phenanthren-2-yl)acetamido acetate	0.23	[18]
N-Sulfonyloxy-2-Acetylaminophanthrene	[2-oxo-1-(phenanthren-2-yl)propoxy]sulfonic acid	0.43	[18]
N-Sulofnyloxy-N-Methyl-4-aminoazobenzene	{[methyl(4-[(E)-2-phenyldiazen-1-yl]phenyl)amino]oxysulfonic acid	0.12	[18]

O-Sulofnyloxy-N-4-aminoazobenzene	[{(4-[(E)-2-phenyldiazen-1-yl]phenyl)amino)oxy]sulfonic acid	0.12	[18]
N-Hydroxyl-3,2'-Dimethyl-4-aminobiphenyl	N-{2',3-dimethyl-[1,1'-biphenyl]-4-yl}hydroxylamine	0.90	[18]
N-Acetoxy-3,2'-Dimethyl-4-aminobiphenyl	{2',3-dimethyl-[1,1'-biphenyl]-4-yl}amino acetate	0.77	[18]

TABLE 2 – Base de données 2 : Liste des métabolites utilisés pour déterminer les seuils de réactivités avec l'ADN

Image	Name	IUPAC Name	Experimentally reactive with DNA (predicted score)
	AαC-N <sup>2</sup> -Gluc	3,4,5-trihydroxy-6-({9H-pyrido[2,3-b]indol-2-yl}amino)oxane-2-carboxylic acid	(0.04)
	AαC-3-OH	2-amino-9H-pyrido[2,3-b]indol-3-ol	(0.87)
	AαC-6-OH	2-amino-9H-pyrido[2,3-b]indol-6-ol	(0.88)
	AαC-HN <sup>2</sup> -OH	N-{9H-pyrido[2,3-b]indol-2-yl}hydroxylamine	yes (0.94)
	AαC-3-O-SO <sub>3</sub> H	{2-amino-9H-pyrido[2,3-b]indol-3-yl}oxidanesulfonic acid	(0.86)
	AαC-3-O-Gluc	6-({2-amino-9H-pyrido[2,3-b]indol-3-yl}oxy)-3,4,5-trihydroxyoxane-2-carboxylic acid	(0.49)
	AαC-6-O-SO <sub>3</sub> H	{2-amino-9H-pyrido[2,3-b]indol-6-yl}oxidanesulfonic acid	(0.81)
	AαC-6-O-Gluc	6-({2-amino-9H-pyrido[2,3-b]indol-6-yl}oxy)-3,4,5-trihydroxyoxane-2-carboxylic acid	(0.53)



	AαC-HN <sup>2</sup> -O-Gluc	3,4,5-trihydroxy-6-[(9H-pyrido[2,3-b]indol-2-yl)amino]oxy]oxane-2-carboxylic acid	yes (0.83)
	N-Acetoxy-AαC	(9H-carbazol-2-yl)amino acetate	yes (0.80)
	N-Sulfonyloxy-AαC	{[(9H-carbazol-2-yl)amino]oxy}sulfonic acid	yes (0.93)
	7-oxo-MeIQx	2-amino-3,8-dimethyl-3H,6H,7H-imidazo[4,5-f]quinoxalin-7-one	(0.92)
	N-desmethyl-7-oxo-MeIQx	2-amino-8-methyl-3H,6H,7H-imidazo[4,5-f]quinoxalin-7-one	(0.93)
	MeIQx-N <sup>2</sup> -Gluc	6-({3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl)amino)-3,4,5-trihydroxyoxane-2-carboxylic acid	(0.07)
	MeIQx-N <sup>2</sup> -SO <sub>3</sub> H	N-{3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl}sulfamic acid	(0.13)
	8-CH <sub>2</sub> OH-IQx	{2-amino-3-methyl-3H-imidazo[4,5-f]quinoxalin-8-yl}methanol	(0.93)
	IQx-8-COOH	2-amino-3-methyl-3H-imidazo[4,5-f]quinoxaline-8-carboxylic acid hydrate	(0.92)
	HON-MeIQx	N-{3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl}hydroxylamine	(0.95)
	HON-MeIQx-N <sup>2</sup> -Gluc	6-(N-{3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl}hydroxyamino)-3,4,5-trihydroxyoxane-2-carboxylic acid	(0.31)
	N-Acetoxy-MeIQx	{3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl}amino acetate	yes (0.90)
	N-Sulfonyloxy-MeIQx	[{[(3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-yl)amino]oxy}sulfonic acid	yes (0.95)
	HON-PhIP	N-{1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-yl}hydroxylamine	yes (0.96)

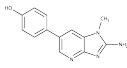
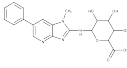
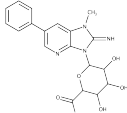
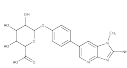
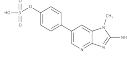
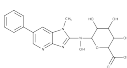
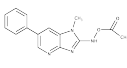
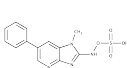
	4'-HO-PhIP	4-{2-amino-1-methyl-1H-imidazo[4,5-b]pyridin-6-yl}phenol	(0.93)
	PhIP-N2-Gluc	3,4,5-trihydroxy-6-((1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-yl)amino)oxane-2-carboxylic acid	0.06
	PhIP-N3-Gluc	3,4,5-trihydroxy-6-(2-imino-1-methyl-6-phenyl-1H,2H,3H-imidazo[4,5-b]pyridin-3-yl)oxane-2-carboxylic acid	(0.78)
	4'-OGluc-PhIP	6-(4-{2-amino-1-methyl-1H-imidazo[4,5-b]pyridin-6-yl}phenoxy)-3,4,5-trihydroxyoxane-2-carboxylic acid	(0.73)
	4'-OSO3H-PhIP	(4-{2-amino-1-methyl-1H-imidazo[4,5-b]pyridin-6-yl}phenyl)oxidanesulfonic acid	(0.89)
	HON-PhIP-N <sup>2</sup> -Gluc	3,4,5-trihydroxy-6-(N-{1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-yl}hydroxyamino)oxane-2-carboxylic acid	(0.26)
	N-Acetoxy-PhIP	{1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-yl}amino acetate	yes (0.90)
	N-Sulfonyloxy-PhIP	[{1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-yl}amino]oxy]sulfonic acid	yes (0.96)

TABLE 3 – Base de données 3 : Liste des métabolites d'AαC, PhIP et MeIQx utilisés pour la validation du workflow





## Résumé

Prédire, modéliser et analyser le métabolisme de xénobiotiques, substances étrangères à un organisme, à l'aide de méthodes informatiques est un challenge majeur mobilisant la communauté scientifique depuis de nombreuses années. Cette thèse vise à implémenter des méthodes informatiques multi-échelles pour prédire et analyser le métabolisme des xénobiotiques. Un premier axe de cette étude portait sur la construction et l'annotation automatique *de novo* de graphes métaboliques combinant fortes sensibilités et précisions. Ces graphes fournissent ainsi la prédiction du métabolisme de xénobiotiques chez l'homme, ainsi que la génotoxicité des molécules et atomes qui le composent. Puis, le travail s'est orienté sur l'implémentation d'un modèle mathématique dynamique modélisant des effets de compétition enzymatique à travers le développement d'une méthodologie permettant l'exploitation de données biologiques restreintes tout en limitant les biais inhérents.

## Mots clés

Bioinformatique ; Chemoinformatique ; Modélisation multi-échelle ; Prédiction ; Métabolisme ; Toxicité ; Pipeline ; Dynamique ; Quantitatif ; Xénobiotiques.

## Abstract

Predicting, modelling and analysing the metabolism of xenobiotics, substances foreign to an organism, using computer methods, has been a major challenge for the scientific community for many years. This thesis aims to implement multiscale computing methods for predicting and analyzing the metabolism of xenobiotics. A first focus of this study was on the construction and automatic *de novo* annotation of metabolic graphs combining high sensitivity and precision. These graphs thus provide the prediction of the metabolism of xenobiotics in humans, as well as the genotoxicity of the molecules and atoms that make up xenobiotics. Then, the work focused on the implementation of a dynamic mathematical model modelling enzymatic competition effects through the development of a methodology allowing the exploitation of limited biological data while limiting inherent biases.

## Keywords

Computational Biology ; Chemoinformatics ; Multi-scale modelling ; Predicting ; Metabolism ; Toxicity ; Pipeline ; Dynamic ; Quantitative ; Xenobiotics.