



**HAL**  
open science

# Étude et applications de nouveaux modèles géométriques des canaux d'accès au site actif de certains cytochromes P450 humains par des ligands volumineux

Lydia Benkaidali

► **To cite this version:**

Lydia Benkaidali. Étude et applications de nouveaux modèles géométriques des canaux d'accès au site actif de certains cytochromes P450 humains par des ligands volumineux. Chimie théorique et/ou physique. Université Pierre et Marie Curie - Paris VI, 2016. Français. NNT : 2016PA066377 . tel-01388012v2

**HAL Id: tel-01388012**

**<https://theses.hal.science/tel-01388012v2>**

Submitted on 6 Mar 2017

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

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



χ Physique  
χ Analytique  
Paris Centre  
ED 388



# Université Pierre et Marie Curie

Ecole Doctorale :

Chimie physique et chimie analytique de Paris Centre

*Laboratoire* : ITODYS, CNRS UMR 7086, Paris

**Etude et applications de nouveaux modèles géométriques des canaux d'accès au site actif de certains cytochromes P450 humains par des ligands volumineux.**

Par : **Lydia Benkaidali**

Thèse de doctorat en chimie théorique

Dirigée par **François Maurel**  
Co-dirigée par **Michel Petitjean**

**Présentée et soutenue publiquement le 15 septembre 2016**

Devant un jury composé de :

ANDRÉ	François	Chercheur	Examineur
CHOMILIER	Jacques	Directeur de recherches	Président
DANSETTE	Patrick	Directeur de recherches	Rapporteur
MAUREL	François	Professeur	Directeur de thèse
PETITJEAN	Michel	Chargé de recherches	Co-directeur de thèse
TANGOUR	Bahoueddine	Professeur	Examineur
URBAN	Philippe	Chargé de recherches	Rapporteur



Except where otherwise noted, this work is licensed under  
<http://creativecommons.org/licenses/by-nc-nd/3.0/>



*A mes Parents...*



## REMERCIEMENTS

Albert Einstein disait : *Ma vie extérieure et intérieure dépend du travail de mes contemporains et de celui de mes ancêtres et je dois m'efforcer de leur fournir la même proportion de ce que j'ai reçu et de ce que je reçois encore.*

J'espère, par ce travail, apporter à mes encadrants, à mes professeurs, à ma famille, à mon université, à mes collègues, à mes amis, une infime proportion de ce que j'ai reçu et de ce que je continue à recevoir de leur part...

Ce travail de thèse a été réalisé au laboratoire ITODYS, de l'Université Paris Diderot, sous la direction du professeur François Maurel, avec le co-encadrement du Docteur Michel Petitjean du laboratoire MTi.

Je remercie le Programme Doctoral International "Modélisation des Systèmes Complexes" de l'IRD et l'UPMC pour le financement, ainsi que l'entreprise IBC pour le co-financement, qui m'ont permis de réaliser ce doctorat.

J'adresse mes remerciements aux membres du jury qui ont accepté d'évaluer mes travaux. Je voudrais ainsi exprimer ma reconnaissance au **Dr. Patrick Dansette** et au **Dr. Philippe Urban** d'avoir accepté de juger ce travail, ainsi qu'au **Dr. Jacques Chomilier** d'avoir accepté de présider cette soutenance.

Par ce travail je clos plusieurs années d'efforts, de volonté à réaliser un vœu, celui de faire de la recherche. Je quitte un confort matériel et affectif, pour une aventure incertaine et difficile. L'expatriation studieuse n'a pas été un long fleuve tranquille, mais elle a surtout été animée par une détermination de faire "le travail que je veux".

*...Je ne crois pas qu'il y ait de bonne ou de mauvaise situation. Moi, si je devais résumer ma vie aujourd'hui avec vous, je dirais que c'est d'abord des rencontres, des gens qui m'ont tendu la main, peut-être à un moment où je ne pouvais pas, où j'étais seul chez moi. Et c'est assez curieux de se dire que les hasards, les rencontres forgent une destinée... Parce que quand on a le goût de la chose, quand on a le goût de la chose bien faite, le beau geste, parfois on ne trouve pas l'interlocuteur en face, je dirais, le miroir qui vous aide à avancer. Alors ce n'est pas mon cas, comme je le disais là, puisque moi au contraire, j'ai pu; et je dis merci... Mais demain, qui sait, peut-être seulement à me mettre au service de la communauté... (Édouard Baer). <https://www.youtube.com/watch?v=O2S-0twv64M>*

### Remerciements particuliers :

J'ai une reconnaissance sans pareil envers le **Pr. François Maurel** qui m'a accordé cette chance, pourtant sans me connaître, de m'avoir accueillie au sein de son laboratoire et d'avoir encadré cette thèse. Une personne qui ne s'arrête pas à un "standard administratif" de recrutement, mais qui donne la chance à ceux qui la demandent. Je note son humilité et sa disponibilité (la porte de François est toujours ouverte ...quand il est là!). Je note également le dépassement du préjugé dont j'ai bénéficié, sans lequel ce travail n'aurait jamais eu lieu. Si d'autres après moi pourraient en bénéficier, mon expérience aurait servi au moins à ça. Ce que j'ai reçu je saurai le mettre à disposition à mon tour.

Si un mot plus puissant que merci existe, je le dirai au **Dr. Michel Petitjean**. J'aurai eu la chance de rencontrer cette personnalité scientifique dans ma carrière, et

j'ai déjà tout gagné rien qu'à l'avoir rencontré et travaillé avec lui. Sa compétence est indiscutable, son dévouement et l'impeccabilité de sa mission sont exemplaires. Je garde en mémoire la fois où il demanda un autographe à une personne qui utilisait le descripteur qui porte son nom ! Je raconte cette anecdote qui représente ce qu'il est et sa richesse intérieure. Michel a été un père scientifique. Merci à Michel pour son coaching incessant "d'aller à l'essentiel", qui a eu raison de mon extrémisme à ce sujet. Merci pour l'aide à la rédaction de cette thèse au mot près. Je le remercie pour m'avoir poussé, avec bonté et fermeté, à m'améliorer durant ces années de thèse et pour m'avoir accordé sa confiance. Il m'a accompagné avec patience et pédagogie et pour ses conseils et ses orientations qui m'ont été précieux.

Je remercie très spécialement le **Dr. François André** de m'avoir fait confiance pour porter le projet dont il a été un des pionniers, pour sa gentillesse et son aide, pour ses encouragements et de m'avoir donné de l'assurance dans la qualité de mes résultats et mon travail. Je le remercie pour la lecture critique et corrections apportées au manuscrit. J'ai une reconnaissance pour François et Michel de m'avoir fait confiance, de m'avoir soutenue et aider à me relever après le mauvais épisode que j'ai vécu avant ma thèse. Je ne l'oublie pas, j'espère avoir été à la hauteur.

J'ai pour le **Dr. Pridi Siregar** une immense gratitude d'avoir accepté de cofinancer ma thèse, d'avoir cru au projet et de m'avoir fait confiance. Pridi porte en tout le sens du *buisness angel*. Merci d'avoir apporté ton œil externe à cette problématique des P450s et de tes propositions. Merci pour les discussions fructueuses sur l'ADME/Tox, le business scientifique ou autres vastes discussions qui vont de l'ontologie aux religions. J'ai eu du plaisir à discuter avec Pridi, un mélange de pragmatisme matériel et de penseur philosophe.

Je suis extrêmement reconnaissante envers le **Pr. Bahoueddine Tangour** qui a suivi cette thèse, pour son étonnante effervescence d'idées pour le sujet, et pour son accueil chaleureux à Tunis. B. Tangour prouve une fois de plus que la passion dépasse le matériel. Je le remercie infiniment de s'être investi pour régler ma situation administrative.

J'ai adoré mon sujet, l'attachement à ce sujet m'a fait dépasser toutes les barrières. Après avoir donné et tapé à toutes les portes pour pouvoir travailler sur la problématique des canaux P450s, je n'oublie pas le sentiment où je me lance dans le sujet et que je me retrouve devant "*un paquet de nouilles*" (droit d'auteur à Florent) en parlant du réseau de canaux adapté à de petites molécules, le "*Waouh, qu'est-ce que je fais avec ça!*" a été dur à encaisser. Mais avec de la ténacité, et une bonne philosophie de vie et de la chance (humaine et scientifique), j'étais loin de penser arriver à ces résultats, jusqu'à être éprise de la volonté d'aller toujours plus loin.

Plus qu'une thèse, j'ai l'impression d'avoir vécu une aventure, scientifique et humaine. J'ai rencontré des gens ouverts, sympathiques, compétents et humains. Comme je l'ai dit souvent "l'équilibre de la nature s'est révélé dans la compensation de mes difficultés par mon entourage". J'ai adoré mon environnement, je dis merci à toutes ces personnes. La thèse m'a fait connaître des gens exceptionnels, par leur savoir, leur personnalités diverses, mais l'humilité commune que tout le monde partage. J'ai eu de la chance, au sein de mon laboratoire ITODYS, puis l'équipe du MTi, au CEA, à Alger, à Oran et à Tunis.

Je voudrais remercier ici les personnes grâce à qui ces années (dans et hors du laboratoire) ont été des moments agréables de vie et d'échanges.

Un remerciement particulier pour le **Dr. Gautier Moroy** pour la lecture critique du manuscrit, pour ses explications et sa disponibilité. Je le remercie pour ses cours si bien préparés et présentés.

Je remercie du fond du cœur les **Professeurs Meziane Brahim, Safia Kellou, Boubekour Maouche, Mourad Lounis** pour leur contribution incontestée au dénouement administratif de cette thèse. Je grave dans ma mémoire les moult appels de Mourad à se soucier de mon avancement et des divers contacts et s'est démené pour les solutions proposées, son dévouement sincère ressenti, celui d'un ami, d'un frère.

Je remercie toute la bande de **collègues du PDI MSC**, des 4 coins du monde. Vous êtes tous sans exception sympathiques, merci pour l'ambiance festive épicée par la particularité régionale de chacun, du plus folklorique au plus réservé. Je n'ai pas toujours compris vos modélisations environnementales ou de réseaux multi-agents (moi qui fait de la chimie-biologie :-/), mais j'ai surtout compris que la qualité scientifique est répartie à travers le globe, et des personnes animées par la science et la volonté d'apporter chacun son petit grain de sable à son domaine et à l'avancée des problématiques de son pays. Merci au **comité du PDI** d'avoir permis cette cohésion et d'avoir réussi le pari. Merci aussi pour les journées péniches organisées chaque année, qui nous permettent en toute convivialité de partager des moments d'amitié et de science.

Je remercie l'ED 387 et son directeur le **Pr. Germain Trugnan** qui a accepté mon inscription, pour son accueil et son "*good will*" pour cette thèse. Je remercie l'ED 388 et sa directrice, **Pr. Hélène Gérard** d'avoir accepté le transfert de mon inscription dans son ED jusqu'à ma soutenance. Merci au **Dr. Jean-Michel El Hage Chahine**, pour avoir accepté de siéger dans mon comité de thèse.

Les personnes qui ont participé à ce travail : **Dr. Christophe Cambier**, le facilitateur à la *positive-attitude* et à l'esprit *tant qu'il y a du travail allons-y*; **Ikram Allam** pour son travail ardu sur la deuxième version de CCCPP et sa pédagogie qui a profité à mon mémoire; **Jean-Paul Lemaire** pour toutes les fois où il s'est démené pour me rendre service et dépassant le cadre de ma thèse; **Aymeric Noël** pour son aide informatique dévouée; et le **Dr. Claire Minoletti** pour son intérêt pour mon sujet qui s'est manifesté par sa collaboration industrielle avec Sanofi, et le recrutement de **Sonia Ziada**, qui ensemble ont beaucoup apporté au projet. L'apport de Sonia à la thèse est indéniable. C'est elle qui m'a éclairé sur les canaux, et sur les cross-validations. Elle a été aussi de très bons conseils, qui dépassent le cadre scientifique, marquée par beaucoup de maturité et de générosité. Plus qu'une collègue ça a été aussi une amie. Je pense que la thèse aurait été autre si Sonia n'étais pas là; et ceci est grâce à Claire et à François André.

Je veux remercier toutes ces personnes que j'ai côtoyées dans mon environnement scientifique, et qui ont participé à égayer ma thèse, que ce soit sur le plan scientifique que sur le plan humain. Je vous remercie de votre sympathie et gentillesse, et investissement pour certains. Pas dans l'ordre : au CEA Annie Barrand-Frelet et Marcel Delaforge, à l'ITODYS : Mahamadou Seydou pour sa gentillesse mémorable et ses qualités authentiques puisées de l'esprit africain de collectivité; mes collègues de bureau adorables Laurence Ourry (l'animatrice, ça ne s'arrête pas et c'est normal vu son ouverture d'esprit),

Arnaud Fihey (le studieux et chouette), Karima Lassoued et Ester Papa (dynamique et vivace), leur souplesse de caractère a été d'un grand apport à la convivialité, Florent Barbault, notamment pour ses éclairages en modélisation biologique, toujours dévoué à apporter ses connaissances concernant la "communauté drug-bio-momo", Aurélie Perrier, Jean-Claude Pertays, Philippe Decosse, Philippe Lang, Mohamed Jouini, Pierre-François Quenin, Faïna Mamerri et Souad Ammar pour leurs tempéraments. Au MTi : Anne-Claude Camproux, Olivier Taboureau, Delphine Flatters, Leslie Regad, Anne Badel, Colette Geneix, et les enseignants du Master ISDD qui m'ont permis de me mettre à niveau en biologie structurale. A Oran : Abdelkader Chebbah, Aïcha Derdour, Mohamed Benyettou, Noureddine Tchouar et Rachid Kessas pour leurs disponibilités et leur bonne volonté, puis Yasmine Derdour pour son implication sincère; puis les cools et atypiques Lillois : Stéphane Aloïse et Nacer Idrissi, sans oublier les agents de sécurité du bâtiment Lavoisier qui me permettaient de rester travailler et qui m'ouvraient le grillage jusqu'à des heures très tardives (minuit j'avoue).

Je ne veux pas terminer ces remerciements scientifiques sans souligner le mérite de toutes les personnes qui font de la recherche dans les pays du Sud, et restent animées malgré le manque de confort qu'on connaît au Nord. La recherche est un métier qui nécessite une dynamique constante pour "rester branché", et j'ai mon total respect aux chercheurs du Sud qui maintiennent le cap malgré une dynamique à cinétique plus faible et des états de transition beaucoup plus hauts :-). Je reste certaine que le secret est qu'ils portent le soleil dans leurs coeurs et leurs esprits.

Je souhaite exprimer ma gratitude à ceux qui m'ont soutenue et qui ont été indulgents avec moi durant la rédaction de cette thèse (clin d'œil...).

Je remercie mes **amis**, ou ma famille de substitution en France. Vous avez été là pour les bons moments mais aussi j'ai su compter sur vous lorsque j'en ai eu besoin. Merci pour ces dîners, soirées, sorties, etc. Merci aussi pour la présence et la sécurité que vous m'avez apporté. Je suis désolée de n'avoir pas été très disponible cette dernière année, entre la thèse et "mes trucs" d'à côté j'ai eu du mal à répondre sous une semaine à vos appels/messages :-p Merci à celles et à ceux qui m'ont compris.

J'ai une tendre pensée pour mes **ami(e)s "lointain(e)s"**. A ceux qui me posent toujours la question "tu viens quand ?" et qui ont compris que mon "bientôt" ne présageait rien d'imminent. 1,2,3 viva l'Algérie dans mon cœur toujours :-)

Je finis par remercier avec affection, ma **Famille** qui, d'Alger, m'a soutenue durant mes études. Je remercie mes amours de parents, qui de tout leur cœur ont fait le maximum pour m'assister, me soulager, m'encourager, et surtout me comprendre pour mes longs moments d'absence ou d'indisponibilité. Ma soeur Yasmin et mon frère Anis, puis l'adorable tante Aziza qui suscite tous ses contacts à mon moindre coup de grippe.

Me voilà arrivée!

# SOMMAIRE

<b>LISTE DES FIGURES</b>	<b>12</b>
<b>LISTE DES TABLES</b>	<b>15</b>
<b>1 INTRODUCTION</b>	<b>17</b>
1.1 Que savons-nous des cytochromes P450 ?	17
1.2 Découverte du P450 : historique, avancées et problématiques	18
1.3 Objectif de la thèse, originalité	19
1.4 Etat de l'art. Contexte et enjeux du projet	21
1.4.1 Pertinence et caractère stratégique du projet	21
1.4.2 Cytochromes P450 : le paradigme de la multi-spécificité dans le paysage conformationnel	22
1.4.3 Hypothèses sur le comportement polyvalent des CYPs	23
1.5 Topologie structurale des cytochromes P450 et des canaux	24
1.5.1 Repliement structural global et des cavités des P450s	24
1.5.2 Régions de reconnaissance et de liaison de substrat et du partenaire redox	28
1.5.3 Différences structurales et fonctionnelles entre les P450	29
1.6 L'hypothèse de canalisation	30
1.7 Simulations de dynamique moléculaire	34
1.8 Importance de la lipophilie des substrats des P450s	38
1.8.1 Rôle de la lipophilie dans l'affinité de liaison	38
1.8.2 Les facteurs contributifs dans les relations de lipophilie	40
1.8.3 Lipophilie et sélectivité du P450	41
1.9 Ingénierie sur les substrats des isoenzymes pour le contrôle du métabolisme	42
1.9.1 Une conception rationnelle pour une approche <i>Early-ADME</i>	42
1.9.2 Stratégies de contrôle du métabolisme par les CYPs	44
1.10 Isoformes P450 humains : interactions spécifiques avec les substrats. Sélectivité	46
1.10.1 Diversité de formes et de fonctions. Modes d'interactions isoformes-substrats	47
1.10.2 Structure de l'isoenzyme CYP3A4 et de ses ligands	49
Topologie du CYP3A4	51
Substrats et inhibiteurs du CYP3A4 : corrélations et mécanisme de liaison au site actif	55
Liaison au substrat/inhibiteur et coopérativité dans le CYP3A4	56
1.10.3 Résumé des problèmes posés par l'isoforme 3A4	60
1.10.4 Structure de l'isoenzyme CYP1A2 et de ses ligands	61
Substrats du CYP1A2 : corrélations et mécanisme de liaison au site actif	62
<b>2 METHODES THEORIQUES D'EXPLORATION DES CAVITES ET CANAUX DANS LES PROTEINES</b>	<b>63</b>
2.1 Définitions, terminologie	63

2.2	Etat de l'art : approches existantes, méthodologie et algorithmes. . . . .	67
2.2.1	Méthodes géométriques à ligand explicite . . . . .	69
2.2.2	Méthodes géométriques à flux minimal . . . . .	69
2.3	Caractéristiques originales du calcul des canaux dans CCCPP . . . . .	71
2.3.1	Modélisation géométrique des protéines avec CCCPP . . . . .	72
	Première version de CCCPP : définition du canal . . . . .	72
	Seconde version de CCCPP : chemins de coûts minimaux . . . . .	76
2.3.2	Les deux modes de visualisations des cavités . . . . .	78
2.4	Echantillonnage conformationnel des ligands . . . . .	78
2.5	Préparation des données . . . . .	79
<b>3</b>	<b>RESULTATS ET DISCUSSION . . . . .</b>	<b>81</b>
3.1	Génération des canaux et des chemins de coûts minimaux du CYP3A4 . . . . .	81
3.1.1	Préparation des données . . . . .	82
	Reconstruction des acides aminés manquants . . . . .	82
	Préparation du fichier PDB . . . . .	82
	Alignement dans un référentiel commun . . . . .	82
	Arrivée dans le site actif . . . . .	82
3.1.2	Exécution de CCCPP . . . . .	83
3.1.3	Génération du graphe de l'enveloppe convexe . . . . .	85
3.1.4	Génération du réseau complet des canaux . . . . .	85
3.1.5	Extraction et marquage des composantes connexes . . . . .	88
3.1.6	Etablissement des chemins d'accès au site actif . . . . .	93
3.2	Description des canaux du CYP3A4 . . . . .	94
3.2.1	Canal 2a : le canal principal . . . . .	98
3.2.2	Canal 2f . . . . .	101
3.2.3	Canal S . . . . .	102
3.2.4	Canal 2e . . . . .	104
3.2.5	Topologie des canaux : résumé . . . . .	105
3.3	Position des canaux du CYP3A4 par rapport à la membrane . . . . .	105
3.4	Conformations ouvertes ou fermées : les trois états du CYP3A4 associés à l'ouverture des canaux . . . . .	107
3.4.1	Diagramme de prédiction des ouvertures du CYP3A4 . . . . .	109
3.4.2	Forme <i>Closed</i> : C . . . . .	114
3.4.3	Forme <i>Open 1</i> : O1 . . . . .	115
3.4.4	Forme <i>Open 2</i> : O2 . . . . .	116
3.5	Conformation de la forme apo du CYP3A4 . . . . .	117
3.6	Inhibition du CYP3A4 par les analogues du ritonavir . . . . .	123
3.6.1	Le ritonavir et ses analogues . . . . .	123
3.6.2	Relations structure-activité des analogues du ritonavir . . . . .	125
3.6.3	Influence des groupes latéraux et terminaux du ritonavir sur l'ouverture du CYP3A4 . . . . .	127
3.6.4	Influence des groupes de tête sur la stabilité du complexe . . . . .	131
3.6.5	Rôle du thiazole de tête et de l'isopropylthiazole terminal . . . . .	132
	Effet des interactions non covalentes de la tête thiazolique . . . . .	134
	Rôle du groupement isopropylthiazole dans la stabilisation de la boucle FF' . . . . .	136

3.6.6	Rôle des groupes hydrophobes latéraux . . . . .	137
	Importance du premier phényle dans l'affinité et l'inhibition . . . . .	138
	L'élargissement du ligand provoque l'ouverture en forme O1 . . . . .	139
	Ouverture en forme O2 pour l'accueil de deux molécules sur deux sites de liaison . . . . .	141
	La flexibilité du désoxyritonavir augmente la stabilité de la forme O2	145
3.6.7	Conclusion sur les analogues du ritonavir . . . . .	147
3.7	Interactions du CYP3A4 avec différents substrats et inhibiteurs . . . . .	147
3.7.1	En forme C . . . . .	147
	Liaison de la progestérone sur un site périphérique . . . . .	148
	Interaction avec une petite molécule : la métyrapone . . . . .	150
	Interaction avec une molécule allongée et flexible : la bromoergocryptine . . . . .	151
3.7.2	En forme O1 . . . . .	153
3.7.3	En forme O2 . . . . .	154
	Interaction du CYP3A4 avec deux molécules de kétoconazole . . . . .	154
	Liaison à une molécule volumineuse : l'érythromycine . . . . .	160
3.8	L'hypothèse multi-canaux dans l'accueil multi-substrats du 3A4 . . . . .	163
3.8.1	Mouvement concerté des canaux du CYP3A4 et canalisation du substrat . . . . .	163
3.8.2	Flexibilité du 3A4 dans l'accueil multi-substrats . . . . .	164
<b>4</b>	<b>PRESENTATION DE L'ARTICLE [Benkaidali et al. 2014]</b> . . . . .	<b>167</b>
<b>5</b>	<b>CONCLUSION ET PERSPECTIVES</b> . . . . .	<b>177</b>
5.1	Conclusion . . . . .	177
5.2	Perspectives . . . . .	178
	<b>ANNEXE : PRODUCTION SCIENTIFIQUE</b> . . . . .	<b>181</b>
	<b>BIBLIOGRAPHIE</b> . . . . .	<b>183</b>

## LISTE DES FIGURES

1.1	Carte topographique des éléments de structures secondaires du P450BMP : figure reprise de [Graham & Peterson 1999]. . . . .	26
1.2	Eléments de structures secondaires du cytochrome P450 BM-3. . . . .	28
1.3	Canaux définis par CAVER pour le CYP1A1 (modèle par homologie basé sur la structure cristalline du CYP1A2). . . . .	32
1.4	Canaux définis par CAVER pour le CYP3A4. . . . .	32
1.5	Exemple d'un docking de dynamique moléculaire dirigée (SMD) de la testostérone dans le canal du CYP3A4 transitoirement ouvert, défini par les hélices F'-G'. . . . .	35
1.6	Pourcentages de composés et de médicaments métabolisés [Rendic 2002]. . . . .	47
1.7	Visualisations des structures secondaires et tertiaires du CYP3A4. . . . .	51
1.8	Evolution des vitesses de réaction $V_i$ en fonction de la concentration en substrat. . . . .	57
1.9	Evolution des vitesses de réaction $V_i$ en fonction de la concentration en substrat pour une enzyme allostérique de type K à effet coopératif du substrat. . . . .	59
2.1	Exemple de graphe non orienté comportant une seule composante connexe. . . . .	64
2.2	Cas le plus simple d'une triangulation de Delaunay dans le plan (avec les cercles circonscrits). . . . .	65
2.3	Exemple de triangulation de Delaunay et de diagramme de Voronoï dans le plan. . . . .	67
2.4	Grille CAVER, selon [Petřek <i>et al.</i> 2006]. . . . .	70
2.5	Faces externes et internes de l'enveloppe convexe du CYP3A4 triangulé. . . . .	73
2.6	Recherche d'un chemin dans le graphe facial. . . . .	74
2.7	Modélisation géométrique par CCCPP du CYP3A4 (code PDB : 1TQN). . . . .	75
2.8	Représentation du trajet 2a à l'aide de trois MCPs. . . . .	77
3.1	Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4 : 1 <sup>ere</sup> partie. . . . .	84
3.2	Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4 : 2 <sup>eme</sup> partie. . . . .	85
3.3	Répartition des tétraèdres internes et externes dans les canaux pour diverses structures du CYP3A4 : 1 <sup>ere</sup> partie. . . . .	87
3.4	Répartition des tétraèdres internes et externes dans les canaux pour diverses structures du CYP3A4 : 2 <sup>eme</sup> partie. . . . .	88
3.5	Répartition des tétraèdres dans le canal principal du CYP3A4 à CV=6Å. . . . .	91
3.6	Importance du canal principal du CYP3A4 au sein de l'enveloppe convexe : 1 <sup>ere</sup> partie. . . . .	92
3.7	Importance du canal principal du CYP3A4 au sein de l'enveloppe convexe : 2 <sup>eme</sup> partie. . . . .	92
3.8	Réseau de canaux pour le CYP3A4 (code PDB 2V0M). . . . .	96
3.9	Cycle catalytique du cytochrome P450. . . . .	97

3.10	Graphe facial du réseau de canaux (à $CV=5.75\text{\AA}$ ) du CYP3A4 code PDB 2V0M, comportant deux molécules de kétoconazole co-cristallisées. . . . .	99
3.11	Chemins de coûts minimaux 2a et 2f, calculés pour $CV=5.75\text{\AA}$ à l'intérieur des canaux du CYP3A4, code PDB 2V0M. . . . .	100
3.12	Superposition des réseaux de canaux de deux structures du CYP3A4, codes PDB 1TQN et 2V0M. . . . .	100
3.13	Trajets 2a et 2e calculés par CCCPP respectivement à $CV=6\text{\AA}$ et à $CV=5.75\text{\AA}$ sur le CYP3A4 code PDB 1TQN. . . . .	104
3.14	CYP (ancré dans la membrane plasmique), avec les canaux 2a, 2e, 2f et S calculés par CCCPP. . . . .	108
3.15	Diagramme des formes du CYP3A4 en fonction du paramètre structural et physicochimique du ligand. . . . .	110
3.16	Graphe facial du CYP3A4 (code 1TQN) à $CV_{lim}=6\text{\AA}$ . . . . .	118
3.17	Trajets 2a, 2e et S du CYP3A4 (code 1TQN) calculés à $CV=5\text{\AA}$ . . . . .	119
3.18	Graphe facial à $CV_{lim}$ du 1TQN (en vert) et du 1W0E (en bleu). . . . .	120
3.19	Graphe facial du CYP3A4 (code 1TQN) à $CV_{lim}=6\text{\AA}$ : superposition des trajets 2a et S. . . . .	120
3.20	Trajets S pour le CYP3A4 en forme apo (code 1TQN) et holo (code 4K9U). . . . .	121
3.21	Structure du ritonavir et de ses analogues. . . . .	124
3.22	Structure du ritonavir en interaction avec les acides aminés voisins du CYP3A4 (code PDB 3NXU). . . . .	126
3.23	Graphe facial du canal 2a à $CV_{lim}=6.25\text{\AA}$ . . . . .	128
3.24	Maillage de la surface de van der Waals du ritonavir ( $777\text{\AA}^2$ ). . . . .	129
3.25	Trajet 2a à $CV_{lim}=6.25\text{\AA}$ . . . . .	130
3.26	Complexe du CYP3A4 avec GS2 (code PDB 4I4G). . . . .	131
3.27	Interactions non covalentes du composé GS3 avec le CYP3A4 (complexe 4I4H). . . . .	133
3.28	Trajet 2a dans le complexe du CYP3A4 avec GS3 (code PDB 4I4H). . . . .	133
3.29	Complexe du CYP3A4 avec le DTMCR (code PDB 3TJS). . . . .	134
3.30	Interactions non covalentes du DTMCR avec le CYP3A4 (complexe 3TJS). . . . .	135
3.31	Réseau des canaux du CYP3A4 lié au DTMCR à $CV_{lim}=6.25\text{\AA}$ . . . . .	136
3.32	Positionnement de l'analogue GS8 du ritonavir dans le complexe du CYP3A4 avec GS8 (code PDB 4K9X). . . . .	137
3.33	Superposition du complexe CYP3A4-ritonavir avec le complexe CYP3A4-GS8. . . . .	138
3.34	Structure des désoxyritonavirs GS4 à G8 et orientation par rapport à l'hème. . . . .	139
3.35	Interactions du composé GS6 avec les acides aminés voisins du CYP3A4 (code PDB 4K9V). . . . .	140
3.36	Interactions du composé GS6 avec son environnement dans le complexe 4K9V formé avec le CYP3A4. . . . .	141
3.37	Visualisation du canal 2a à $CV_{lim}=6.5\text{\AA}$ dans le complexe du composé GS6 avec le CYP3A4 (code PDB 4K9V). . . . .	141
3.38	Graphe facial à $CV_{lim}=6.75\text{\AA}$ du complexe entre deux molécules du composé GS5 et le CYP3A4 (code PDB 4K9U). . . . .	142
3.39	Visualisation du canal 2f à $CV=6.25\text{\AA}$ dans le complexe GS5-CYP3A4 (code PDB 4K9U). . . . .	143
3.40	Détails du complexe 4K9U formé entre le CYP3A4 et le composé GS5. . . . .	144

3.41	Les 3 trajets S, 2f et 2a du complexe 4K9U, calculés à $CV=6.75\text{\AA}$ , $6.25\text{\AA}$ , et $5.75\text{\AA}$ , respectivement. . . . .	144
3.42	Réseau de canaux du complexe 4K9T formé entre le CYP3A4 et le composé GS4; graphe facial à $CV_{lim}=6.75\text{\AA}$ . . . . .	145
3.43	Complexe du CYP3A4 avec la progestérone (code PDB 1W0F). . . . .	149
3.44	Trajets 2a et 2e calculés pour le complexe du CYP3A4 avec la métyrapone (code PDB 1W0G). . . . .	151
3.45	Structure chimique de la bromoergocryptine (BEC) et trajet 2a calculé dans le complexe 3A4-BEC (code PDB 3UA1). . . . .	152
3.46	Formule structurale du ketoconazole. . . . .	154
3.47	Graphe facial du canal du 3A4 holo (code PDB 2V0M). . . . .	156
3.48	Trajets 2a et 2f bordés par les structures secondaires du complexe 2V0M. . . . .	157
3.49	Comparaisons des formes apo et holo. . . . .	158
3.50	Trajets 2a, 2f, et 2e calculés avec 2V0M pour les valeurs de CV respectivement égales à $7\text{\AA}$ , $5.75\text{\AA}$ , et $5.5\text{\AA}$ . . . . .	159
3.51	Formule structurale de l'érythromycine. . . . .	160
3.52	Trajets 2f et 2e avec leurs structures secondaires en bordure. . . . .	161
3.53	Graphe facial des canaux calculés à $CV=6.5\text{\AA}$ pour le complexe CYP3A4-érythromycine (code PDB 2J0D). . . . .	162
3.54	Trajet 2f calculé pour le complexe CYP3A4-érythromycine (code PDB 2J0D). . . . .	162

# LISTE DES TABLES

1.1	Intervalles et moyennes des valeurs de $\log(P)$ pour des substrats P450 humains des familles CYP1, CYP2 et CYP3, d'après [Lewis <i>et al.</i> 2004]. . .	40
1.2	Séquence d'acides aminés des structures secondaires, des SRS, et des acides aminés du site actif et cluster Phe dans le CYP3A4. . . . .	53
2.1	Comparaison entre les triangulations de Delaunay 2D et 3D. . . . .	66
2.2	Acides aminés manquants dans les structures secondaires (rayons X) du 3A. . . . .	80
3.1	Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4. . . . .	84
3.2	Réseau complet des canaux du CYP3A4 à $CV=6\text{\AA}$ . . . . .	86
3.3	Caractéristiques de la principale composante connexe, définissant le principal canal du CYP3A4, à $CV=6\text{\AA}$ . . . . .	89
3.4	Caractéristiques de la composante connexe définissant le canal principal du CYP3A4 à $CV_{lim}$ . . . . .	90
3.5	Description du canal 2a du CYP3A4 en forme fermée : apparition du premier canal. . . . .	93
3.6	Description du canal 2e du CYP3A4 en forme fermée : apparition du nouveau canal. . . . .	94
3.7	Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du premier canal. . . . .	94
3.8	Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du deuxième canal (2f). . . . .	95
3.9	Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du troisième canal (2e). . . . .	95
3.10	Localisation des canaux significatifs du CYP3A4 calculés par CCCPP, avec leurs caractéristiques. . . . .	106
3.11	Les 16 structures holo étudiées du CYP3A4. . . . .	111
3.12	Paramètres de 3 substrats prototypes du CYP3A4. . . . .	113
3.13	Mode d'ouverture des canaux dans les complexes du CYP3A4 avec le ritonavir et ses analogues. . . . .	148



# 1 INTRODUCTION

Tout comme dans la littérature [[Cojocaru et al. 2007](#)], dans ce mémoire le terme substrat sera abusivement utilisé pour désigner soit un ligand (substrat ou inhibiteur) soit un métabolite, c'est à dire tout composé interagissant avec une isoforme.

## 1.1 Que savons-nous des cytochromes P450 ?

Le mouvement des xénobiotiques à l'intérieur et à l'extérieur du corps humain est régi par leur physicochimie. Alors que les portes d'entrée incluent les muqueuses nasales et les poumons, beaucoup de xénobiotiques accèdent au corps par la voie gastro-intestinale. Quelques pénétrations sont possibles pour les molécules hydrosolubles par cette voie à travers les pores aqueux ou les jonctions serrées (absorption paracellulaire). Cependant, le transfert rapide du xénobiotique est uniquement réalisé par la traversée du cœur lipidique de la membrane plasmique de la voie gastro-intestinale. Cette absorption transcellulaire requiert que les molécules soient lipophiles. Une fois les molécules présentes dans le système, ces molécules sont libres de diffuser à travers les organes et les tissus pour rejoindre les récepteurs et enzymes, etc., et influencer ces cibles de manière bénéfique ou nocive.

L'élimination des xénobiotiques se fait par le rein. Les composés circulant dans le sang sont filtrés par le glomérule. Les composés ioniques peuvent être activement transportés dans le tubule rénal par des systèmes de transporteurs. La membrane du rein, comme celles décrites plus haut, permet le passage des composés lipophiles, et les composés concentrés dans le tubule peuvent faire une rétrodiffusion (réabsorption passive) pour équilibrer le sang. Ainsi, seules les molécules hydrosolubles démontrent des taux d'élimination rénale substantiels [[Smith et al. 1996](#)].

Les composés trop lipophiles sont directement métabolisés par le corps en molécules plus hydrophiles, pour être filtrés par le rein puis être éliminés. Le métabolisme des xénobiotiques est généralement divisé en réactions de phase I et de phase II. La phase I est une phase d'attaque oxydative et la phase II celle de conjugaison de la molécule avec l'acide glucuronique, l'acide sulfurique, etc. Bien qu'il y ait de nombreuses enzymes impliquées dans la phase du métabolisme de phase I (ex : alcool déshydrogénation, xanthine oxydase et les amines oxydases), l'enzyme centrale dans le métabolisme des xénobiotique est le cytochrome P450. Cette enzyme a été démontrée comme métabolisant une large gamme de substrats, de la cyclosporine à l'acétone [[Smith et al. 1996](#)]. L'intérêt des CYPs est leur implication dans le métabolisme de phase I des composés exogènes auxquels l'Homme est exposé, incluant les composés pharmaceutiques, agrochimiques, et les polluants environnementaux.

Le processus de métabolisme dans les P450s a été identifié en 3 étapes [[Waller et al. 1996](#)] :

1. Phase de reconnaissance : elle s'effectue à longue distance (jusqu'à 8Å), au niveau de l'entrée des canaux entre les substrats et l'enzyme avec des interactions électrostatiques principalement [[Sevrioukova & Poulos 2013a](#)].
2. Phase d'accessibilité de l'extérieur au site actif enfoui par les canaux : importance de la dimension des canaux et de leurs hydrophaties relatives (via les acides aminés bordant les canaux), dans la discrimination des substrats.

3. Phase de réactivité par oxydation une fois que le substrat a atteint le site actif : importance de la structure et des propriétés physicochimiques du couple principaux acteurs (enzyme et substrat) dans l'interaction à chacune des 3 étapes.

## 1.2 Découverte du P450 : historique, avancées et problématiques

Les événements importants concernant la découverte des P450s remontent aux années 50, mais il semble difficile de s'accorder sur une date précise [Estabrook 2003]. En 1943, une nouvelle fraction cellulaire appelée "microsomes" a été isolée par centrifugation fractionnée de tissus animaux homogénéisés [Claude 1943]. La caractérisation des microsomes a conduit à découvrir la présence de plusieurs enzymes de transport d'électrons dans cette nouvelle fraction cellulaire, qui a été identifiée en 1956 comme étant les vésicules membranaires fragmentées du reticulum endoplasmique [Palade & Siekevitz]. Selon [Estabrook 2005], la protéine P450 a été découverte en 1955 à Philadelphie, indépendamment par deux chercheurs du même Institut dirigé par Britton Chance : Martin Klingenberg, qui travaillait sur les microsomes de rat [Klingenberg 1958], et David Garfinkel, qui travaillait sur les microsomes de porc [Garfinkel 1958].

La réaction d'oxygénation a été prouvée pour la phénolase chez un champignon en 1955 [Mason *et al.* 1955], et la même année pour la pyrocatechase chez la bactérie *Pseudomonas* [Hayaishi *et al.* 1955], ainsi que l'incorporation d'oxygène dans la molécule du substrat à l'occasion de l'hydroxylation des stéroïdes [Hayano *et al.* 1955]. L'introduction d'oxygène atmosphérique est impliquée dans diverses activités métaboliques chez les animaux : métabolisme du tryptophane [Tanaka & Knox 1959], biosynthèse des hormones stéroïdiennes à partir d'halostérol [Hayano *et al.* 1955, Ryan & Engel 1956], métabolisme oxydatif des médicaments dans le foie [Axelrod 1955], etc. Les oxygénases impliquées dans le métabolisme du tryptophane sont solubles. Elles ont été purifiées et caractérisées. Par contre, les oxygénases impliquées dans la biosynthèse des hormones stéroïdiques et dans le métabolisme oxydatif des médicaments sont associées à des fractions membranaires, et résistent à la solubilisation et à la purification. Lorsque les fractions membranaires sont traitées avec des détergents pour solubiliser les oxygénases, les activités enzymatiques sont perdues. La nature de ces oxygénases est restée inconnue jusqu'à la découverte des P450s.

Un article pionnier sur le P450 a été publié en 1962 [Omura & Sato 1962], et la recherche biochimique sur cette nouvelle hémoprotéine s'est étendue rapidement dans les années 60 et 70. La première résolution de structure date de 1985 et concerne le P450cam (CYP101) [Poulos *et al.* 1985]. Il faudra attendre 1993 pour disposer d'une seconde structure bactérienne, celle du P450BM3 (CYP102) [Ravichandran *et al.* 1993]. D'autres ont été élucidées: P450terp (CYP108) [Hasemann *et al.* 1994], P450eryF (CYP107A1) [Cupp-Vickery & Poulos 1995], P450nor (CYP55A1) [Park *et al.* 1997], P450sca (CYP105A3) [Palma *et al.* 1994], qui catalyse l'étape finale de la biosynthèse de la pravastatine [Serizawa & Matsuoka 1991], ainsi que d'autres structures de P450cam et de P450BMP dans différents états, avec et sans substrat [Poulos *et al.* 1986, Poulos *et al.* 1987, Li & Poulos 1997], et des formes mutées [Raag *et al.* 1991, Vidakovic *et al.* 1998], etc.

Environ 2000 articles sur les P450s sont publiés annuellement [Omura 2011]. Les CYPs mammaliens sont surtout des enzymes liées à la membrane [Black 1992, Williams *et al.* 2000, Johnson *et al.* 2002, Headlam *et al.* 2003], ce qui les rend difficiles à caractériser structuralement et dynamiquement. Quelques modèles de CYPs liés à la membrane existent, dont deux pour le 2C9 [Berka *et al.* 2011, Cojocaru *et al.* 2011] et un pour le 3A4 [Denisov *et al.* 2012]. Ils sont cohérents avec la plupart des données expérimentales. Beaucoup de travaux ont été effectués pour comprendre le mécanisme au niveau du site actif, mais les études sur la traversée d'une molécule de l'extérieur vers le site actif à travers les canaux n'ont été abordées que tardivement. De ce fait, on a peu de connaissances sur les chemins de passage des molécules et le rôle de ces chemins dans la réactivité. En effet, à notre connaissance, aucune méthode expérimentale ne permet de suivre le trajet d'une molécule au sein des CYPs, nous avons donc eu recours à la modélisation moléculaire pour tenter d'élucider ce processus peu connu du mécanisme catalytique.

### 1.3 Objectif de la thèse, originalité

L'un des principaux intérêts de l'étude des cytochromes P450 (CYPs) est la prédiction des substrats et inhibiteurs, réactifs avec chacune des isoformes, ou autrement dit, comprendre la spécificité. La compréhension de la spécificité des substrats chez les P450s impliqués dans le processus de détoxification de phase I, est une question importante avec des applications immédiates en toxicologie et dans la conception de médicaments. Notre idée première lors du développement de ce projet est que la variété de ligands de chaque isoforme naturelle dans la superfamille P450 et les interactions médicamenteuses potentielles, dépendent fortement du nombre, des caractéristiques structurales et de la composition chimique de ses canaux. Cependant, il y a un manque évident dans la littérature d'outils capables d'apporter une vue précise et exhaustive de ces canaux d'accès. De tels outils auraient été, bien sûr, d'utilisation générale pour comprendre la large spécificité des substrats dans des enzymes multispécifiques, ayant un site actif enfoui, dont l'accès est totalement contrôlé par un ou plusieurs canaux.

Les canaux agissent comme des filtres et impactent à la fois la spécificité au substrat et le mécanisme catalytique.

De nombreuses approches ont été décrites pour aborder la grande spécificité des substrats et les effets de coopérativité multi-substrats chez certains CYPs humains clés. Des études ultérieures en docking de substrats sur des isoformes humains [Lafite *et al.* 2007, Nguyen *et al.* 2008, Goel *et al.* 2008], basées sur des trajectoires décrites au moyen de simulations de dynamique moléculaire sous contraintes, montrent que (i) la spécificité ne réside pas uniquement dans l'interaction au niveau du site actif, mais une partie clé du processus de reconnaissance de ligand par les CYPs réside dans le déchiffrement du réseau de canaux d'accès potentiels pour chaque substrat, et (ii) qu'aucune stratégie *in silico* prédictive réussie, ne peut être prévue sans résoudre à la question de la canalisation de substrat dans la structure du CYP. Les sites actifs des CYPs sont en effet profondément enfouis, bien que les isoformes considérées (microsomales ou bactériennes) et leur accès au substrat soient en grande partie conduits par le réseau déformable des canaux présents dans la structure.

Le but du projet dans lequel s'insère ma thèse est de faire un outil qui prédit l'accessibilité (ou non) d'un composé dans une isoforme donnée du P450. Ce filtre virtuel

prédirait les substrats improbables des isoformes d'un côté, et les canaux empruntés par les substrats probables de l'autre. Autrement dit, pour une isoforme déterminée, l'outil développé, serait capable de nous dire si un composé pénètre ou non dans le CYP jusqu'au site actif, et si oui, de nous dire par quel chemin. Le but ultime aurait été de répondre à cette question avec le programme CCCPP (*Computing Cavities, Channels, Pores and Pockets*), développé au laboratoire MTi [Benkaidali *et al.* 2014], c'est à dire attribuer à chaque ligand un chemin spécifique en se basant sur des critères géométriques (taille et forme) et chimiques (composition en acides aminés).

L'idée initiale lors du développement du programme CCCPP, était d'attribuer à chaque ligand selon sa dimension, un réseau de canaux accessibles calculés à partir de la dimension du ligand. Le critère physico-chimique viendrait ensuite affiner cette sélection de canaux plausibles par le biais d'un docking itératif. Ainsi on imaginait que seul le critère géométrique permettrait la sélectivité, et qu'une molécule soit accessible à un canal du moment où sa dimension le lui permettent. Mais l'application de notre méthode à plusieurs isoformes et plusieurs ligands nous a révélé un autre résultat, qui s'est avéré pertinent pour la prédiction des routes empruntés par les ligands dans le mécanisme d'entrée ou de sortie du site actif. Ce résultat sera discuté dans de ce manuscrit dans le cadre de l'étude du plus important isoforme, le CYP3A4, puis du CYP1A2. Les perspectives de développement du programme seront aussi proposées pour aboutir à un outil de prédiction complet.

Le projet dans lequel s'insère ma thèse propose de fournir un outil de calcul, CCCPP, pour la prédiction des substrats des P450s. C'est un projet interdisciplinaire novateur et le fruit d'une collaboration de plusieurs équipes aux compétences complémentaires, allant du développement mathématique des concepts à la validation par modélisation moléculaire et expérimentale, en passant par la chimie quantique. Mon travail de thèse au sein de ce projet global est la validation de notre programme CCCPP, en l'appliquant à plusieurs isoformes (CYP1A, 2A et 3A) et à plusieurs substrats. Dans cette thèse nous avons étudié la possibilité de passage d'un ligand à travers les canaux et les règles déterminant sa reconnaissance et son accessibilité. Il s'agit également d'utiliser en complément divers outils de modélisation moléculaire pour définir plus précisément le paysage conformationnel des isoformes des CYPs étudiés, de reconstruire les structures CYPs manquantes dans la base PDB (<http://www.rcsb.org>), et de paramétrer par des descripteurs simples les géométries des petites molécules ou ligands.

Ma thèse comprend plusieurs aspects originaux, présentés comme suit :

- Calculer avec CCCPP les canaux d'accès au site actif de certains CYPs humains les plus impliqués dans le métabolisme de xénobiotiques, en y associant des caractéristiques jamais publiées telles que la paroi (d'où surface et volume), à partir des structures enregistrées dans la base PDB. Cela exclut les méthodes de type robotique avec parcours au sein d'obstacles, ainsi que les méthodes de type CAVER [Petřek *et al.* 2006, Chovancova *et al.* 2012], MOLE [Petřek *et al.* 2007, Berka *et al.* 2012], qui avaient été utilisées dès leur apparition pour produire une nomenclature des voies usuelles d'accès des substrats dans les CYP [Cojocaru *et al.* 2007]. Malheureusement, cette nomenclature abusivement utilisée dans la littérature récente est basée sur une description très vague et non rigoureuse des canaux.
- Prise en compte des formes non sphériques des ligands, par l'introduction du modèle cylindrique que nous venons de démontrer comme étant physiquement plus réaliste [Benkaidali *et al.* 2014] : les ligands globulaires sont rares, voire inexistant

parmi les médicaments aujourd’hui sur le marché. Dans cette optique nouvelle, un canal ou un réseau de canaux est décrit pour une famille donnée de substrats.

- Prise en compte de la flexibilité des ligands sans utiliser un nombre fixe de conformères de chaque ligand (tel que 50 ou 100), mais un nombre calculé de conformères représentatifs couvrant au mieux l’espace conformationnel. Dans ce domaine, nous avons développé une méthodologie originale à l’occasion d’une étude préliminaire sur des substrats du CYP3A4 [Meslamani *et al.* 2009, Benkaidali *et al.* 2012].
- Les structures disponibles dans la base PDB peuvent comporter des parties manquantes, qui peuvent être reconstruites par des approches de modélisation par homologie.
- La flexibilité de la protéine peut être prise en compte par un échantillonnage des conformations disponibles dans la PDB pour une même isoforme, et celle du substrat par l’échantillonnage conformationnel.

## 1.4 Etat de l’art. Contexte et enjeux du projet

Nous décrivons ici le contexte biochimique dans lequel se déroule le projet. Les objectifs algorithmiques et les verrous technologiques sont décrits dans le chapitre 2 de ce manuscrit.

### 1.4.1 Pertinence et caractère stratégique du projet

Les CYPs ont un rôle central dans la fonction de détoxification cellulaire, vitale pour l’organisme, en métabolisant de nombreux composés endo et exogènes lipophiles, à bas poids moléculaire. La toxicité d’une molécule exogène est étroitement reliée à l’efficacité de ce processus, car si la molécule n’est pas métabolisée ou inhibe les cytochromes P450, elle devient potentiellement toxique pour son organisme hôte. Ce mécanisme joue donc un rôle important dans le phénomène de biodisponibilité d’un composé organique au sein d’un organisme. L’analyse et la compréhension de l’activité des CYPs est une question récurrente lors du développement d’un médicament ou d’un cosmétique. Le foie humain est particulièrement riche en P450, dont une des fonctions majeures est l’hydroxylation de divers substrats exogènes. En particulier, l’isoforme CYP3A4 (multi-spécifique, pouvant présenter des phénomènes d’allostérie ou d’inhibition (17 structures dans la base PDB dont 13 cocristallisées avec médicament), joue un rôle majeur dans la biotransformation des médicaments. Les chimiothèques représentent des centaines de milliers de substrats potentiels. En raison du coût considérable des tests préalables à la commercialisation de médicaments, il est crucial d’éliminer le maximum de candidats inutiles, d’où le recours au criblage virtuel à haut débit.

Un problème majeur est lié aux tests biologiques à mettre en oeuvre préalablement à la commercialisation de composés chimiques au contact avec l’homme, médicaments, cosmétiques, produits d’usage domestique, pesticides, additifs alimentaires : ces tests sont particulièrement coûteux pour l’industrie, il est par conséquent crucial de ne proposer aux pharmacologues que des substrats potentiellement actifs, et donc d’éliminer le maximum de candidats inutiles. L’ensemble représente des centaines de milliers de molécules synthétisées !

Dans certains cas, l'industrie pharmaceutique privilégie la synthèse de molécules au contraire peu métabolisables, donc interagissant peu avec les CYPs du foie humain. A l'inverse, dans le cas de l'activation de pro-médicaments (exemple : l'activation du cyclophosphamide par le CYP2B6 dans le traitement local des tumeurs), notre outil CCCPP peut s'avérer utile pour prédire la capacité de reconnaissance et le canal d'accès le plus plausible, ce qui permet d'anticiper des problèmes de polymorphisme génétique si une ou plusieurs mutations perturbent le canal considéré. Le profil génétique de métabolisation du patient peut ainsi être pris en compte pour les CYPs dont la structure existe dans la PDB, ce qui permet la modélisation fiable de mutants. D'autres applications prédictives intéressent le domaine de la cancérogénèse chimique : certains P450s humains sont connus pour modifier des molécules en un ou plusieurs passages de métabolisme hépatique et génèrent un métabolite final qui s'avère être un puissant cancérigène. Or les classes structurales des molécules potentiellement médicamenteuses et celles des molécules potentiellement pro-carcinogéniques diffèrent. Cette discrimination est potentiellement détectable par CCCPP s'il y a, pour un CYP donné, une affinité classe chimique/canal caractérisée. CCCPP a donc des applications plus larges que le métabolisme, puisqu'il permettrait de classer les agents chimiques pour chaque type de CYP, et parmi eux les agents de la cancérogenèse chimique.

Une caractéristique majeure des CYPs est que le site actif (hème ferrique) n'est pas en surface, mais au contraire est enfoui à l'intérieur de l'enzyme. Dès lors, l'un des filtres possibles lors du criblage réside dans l'incapacité du substrat potentiel à pénétrer dans les canaux menant au site actif pour des raisons purement stériques. Dans ce contexte, l'utilisation de l'ordinateur et des approches numériques prédictives est indispensable : le criblage virtuel à haut débit de chimiothèques sur divers CYPs présente un intérêt économique considérable, et l'impact sociétal de la recherche assistée par ordinateur de composés médicamenteux est indéniable en termes de santé publique et environnemental (extension aux CYPs des plantes et à l'étude des biomasses).

### **1.4.2 Cytochromes P450 : le paradigme de la multi-spécificité dans le paysage conformationnel**

Les cytochrome P450s définissent une superfamille d'enzymes impliquées dans des événements biologiques divers, surtout en catalysant la mono-oxygénation de molécules organiques endogènes et exogènes. Ils appartiennent à l'une des familles enzymatiques les plus riches, la superfamille de gènes codant P450 comprend plus de 18000 membres connus aujourd'hui, ce nombre étant en croissance rapide. Ils sont aussi célèbres pour être des enzymes "polyvalentes" qui agissent sur un tableau énorme de composés chimiques [Coon 2005]. De cette façon, ils détiennent une place particulière dans l'univers enzymatique. Les CYPs représentent un groupe divers de protéines hème-thiolate trouvées dans presque tous les organismes. Nous ne comprenons qu'une fraction de ces CYPs. L'oxydoréductase flavoprotéine se situant du côté proximal du CYP est un partenaire redox. Le CYP eucaryote est une enzyme ancrée dans la membrane plasmique, alors que le CYP procaryote est soluble dans la cellule. Le métabolisme des xénobiotiques par l'Homme est un processus qui change la structure moléculaire d'un xénobiotique d'une structure absorbable par la membrane plasmique (lipophile, capable de rester au centre de cette membrane plasmique et de la traverser) pour devenir une structure qui peut être éliminée à travers le sang (hydrophile, incapable de pénétrer la membrane plasmique),

particulièrement par le rein. Cette enzyme peut attaquer des composés sur des carbones ou sur des hétéroatomes (N, S, etc.). Les produits d'oxydation sont des carbinols, phénols, époxydes, oxydes d'azote, sulfoxydes, sulfones, et des produits de désalkylation (via une  $\alpha$ -oxydation d'un hétéroatome). Cette production diversifiée de composés, associée à la grande diversité de structures de ligands, dénote de l'importance de cette enzyme [Smith & Jones 1992, Smith *et al.* 1996].

En effet, la plupart des enzymes (au sens général) catalysent des réactions avec une faible diversité de substrats, donnant lieu à un seul type de produit. De telles enzymes exposent fréquemment une plutôt bonne affinité de substrat et des grands nombres de cycles, avec un complexe d'état de transition bien défini, contrôlé principalement par des paramètres thermodynamiques et agissant comme un facteur moteur majeur pour l'activité et la sélectivité du substrat. Leur cinétique est généralement bien décrite par le modèle de Michaelis-Menten. A l'inverse, quelques enzymes peuvent reconnaître une beaucoup plus grande gamme de substrats. Elles sont fréquemment moins efficaces en termes de taux de catalyse, mais contribuent grandement à la diversité de réactions chimiques promues par des enzymes naturelles, une propriété particulièrement pertinente pour des processus de désintoxication dans des cellules spécialisées. D'habitude ces enzymes ont une cinétique déviant significativement du comportement Michaelien standard, et peuvent être le siège de réactions catalysées distinctes, bien qu'ayant lieu dans le même site actif. Leur large spécificité est fréquemment accompagnée par la possibilité que plusieurs substrats rivalisent pour le même site actif, et dans quelques cas la réaction elle-même est accélérée quand des liaisons se produisent avec de multiples substrats. Les CYPs tombent dans cette catégorie complexe d'enzymes, que les visions mécanistiques standards de liaison au ligand échouent à décrire parfaitement.

Cependant, il est apparent que des variations inter-individuelles existent, en plus des facteurs pharmacogénétiques établis qui sont rencontrés dans des populations ethnogéographiques, qui ont des ramifications importantes pour le développement d'un médicament [Rodrigues & Rushmore 2004]. Puisque ces enzymes métabolisant les médicaments sont fréquemment impliquées dans les interactions médicamenteuses (un facteur important à prendre en compte dans le développement pharmaceutique de médicament), il est crucial de comprendre et de décrire les règles qui gouvernent les liaisons multiples au substrat des CYPs.

### 1.4.3 Hypothèses sur le comportement polyvalent des CYPs

Pour l'enzymologiste, le point fascinant est que cette grande diversité de spécificités de substrats soit basée sur une échelle structurale fortement conservée parmi les membres différents de la superfamille. Ceci est presque unique dans le domaine structure protéique. Comment des structures tridimensionnelles semblables peuvent-elles reconnaître et métaboliser une telle diversité de substrats ? Le problème est obscurci par le fait que, pour les P450s, le site actif n'est pas accessible de l'extérieur de la protéine. Le substrat doit trouver sa voie à travers des canaux non facilement visibles pour atteindre le fer du groupe prosthétique héminique.

Tous les P450s ont un repliement structural conservé, bien que les identités de séquence parmi ces protéines puissent être inférieures à 20%, et avec seulement trois acides aminés absolument conservés à travers cette superfamille de gènes [Hasemann *et al.* 1995]. Dans le repliement structural il semble y avoir un cœur fortement conservé, déterminé dès

la comparaison des six premières structures de P450s cristallisées, solubles [Graham & Peterson 1999]. Cependant, le mode d'interaction avec des substrats, et les donneurs d'électrons et des membranes varient. Tandis que les repliements structuraux des P450s sont identiques pour toutes les structures, ils offrent une diversité notable dans les séquences primaires, secondaires et tertiaires. Parmi ces régions variables, certaines sont associées à la reconnaissance des substrats, à la liaison avec des substrats spécifiques, à la liaison avec un partenaire redox, et peuvent cibler l'emplacement cellulaire de la protéine (soit soluble, soit associée à la membrane, mitochondriale ou microsomale).

Plusieurs acides aminés plus ou moins loin du site actif peuvent être potentiellement impliqués dans la modulation structurale de l'enzyme pour le transport moléculaire de l'extérieur de la protéine au centre réactionnel, via les canaux. La spécificité du substrat ne dépend pas uniquement de l'environnement chimique du site actif, mais plus probablement de la somme de ces mécanismes, avec les acides aminés bordant les canaux et qui participent à l'échange cinétique entre substrats et produits dans et hors du site catalytique, ce qui confère à ces enzymes un haut degré de plasticité fonctionnelle.

Il semble évident que, si les substrats ont de nombreux trajets possibles pour accéder au site actif, ils peuvent rivaliser librement pour l'accès dans le cas de substrats différents (un ligand peut emprunter plusieurs trajets, et un même trajet peut servir à plusieurs ligands). Nous pensons que l'existence de canaux multiples et leurs variations structurales (par des mutations), peuvent rendre compte en même temps de la multi-spécificité (par l'affinité discriminée par chaque canal) et de l'aspect multi-substrat (par l'accès concomitant de plusieurs molécules dans le site actif). Ceci est l'hypothèse de départ du projet.

## 1.5 Topologie structurale des cytochromes P450 et des canaux

La structure secondaire des CYPs est décrite par une nomenclature définie par [Poulos *et al.* 1987] (voir aussi figure 1, [Graham & Peterson 1999]), qui est largement utilisé dans la communauté scientifique P450 et que nous utilisons dans ce manuscrit.

### 1.5.1 Repliement structural global et des cavités des P450s

Dès la disponibilité des premières structures, la conservation du repliement structural au sein des P450s entre procaryotes et des eucaryotes, a été une évidence, bien que l'identité de séquence puisse être inférieure à 20% et avec seulement trois acides aminés absolument conservés [Hasemann *et al.* 1995]. Le repliement structural des P450s a été décrit comme la partie du squelette (carbones  $\alpha$ ) structuralement déterminée, qui, après superposition optimale des  $C\alpha$ , présente une déviation d'au plus 2Å [Hasemann *et al.* 1995, Peterson & Graham 1998]. Mais on parle aussi de similarité de séquence; ainsi, il y a environ 20% identité de séquence et 47% de similarité spatiale entre P450cam et P450BMP [Graham & Peterson 1999]. Finalement, tandis que le repliement structural du P450 est le même pour toutes les structures déterminées, elles offrent une certaine diversité de séquences primaires, secondaires et tertiaires, afin d'avoir des substrats spécifiques et des partenaires redox, et pour cibler l'emplacement cellulaire de la protéine (soluble ou

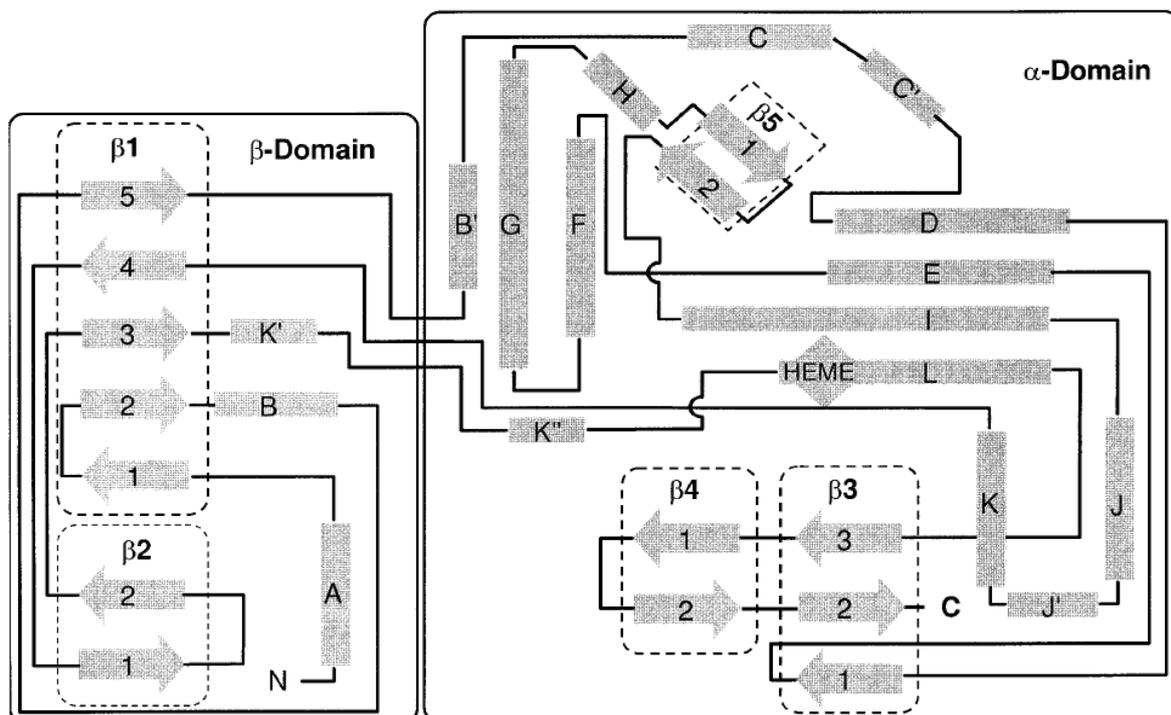
associé à la membrane, mitochondriale ou microsomale). Les P450s diffèrent donc par la composition de leurs sites actifs et de leurs canaux, ou plus généralement de leurs cavités.

La carte topographique du P450BMP [Graham & Peterson 1999], figure 1.1, est utilisée à ce jour comme référence pour tous les P450s. Nous utiliserons cette carte topographique et un de ruban annoté représentant la face distale pour illustrer les ressemblances et les différences entre les structures des P450s. Il y a généralement quatre feuillets  $\beta$  et environ 13 hélices  $\alpha$ , avec insertion d'une hélice notée J' dans P450BM-P, qui, selon les alignements de séquences, semble aussi exister dans P450s des eucaryotes. La présence du feuillet  $\beta_5$  est variable. Le cœur structural conservé du P450 consiste en un ensemble composé des hélices D, E, I, et L et des hélices J et K. L'hélice I contient le Thr fortement conservé, avec un acide aminé dont l'azote terminal est positionné sur le cycle  $\beta$ -pyrrole dans le site actif. L'hélice K contient l'enchaînement E-x-x-R absolument conservé, qui peut être impliqué dans la stabilisation de la structure du cœur, et se trouve du côté proximal de l'hème (qui est le côté où se lie le partenaire redox supposé). L'hélice L fait partie de la région de liaison à l'hème. Il y a deux ensembles de feuillets  $\beta$  structuralement conservés : le feuillet- $\beta_1$ , contenant 5 brins, et le feuillet  $\beta_2$ , contenant deux brins. Ces feuillets aident à former le canal d'accès du substrat hydrophobe.

Comme c'est le cas pour tous les P450s, il y a une séquence consensus structuralement conservée sur la face proximale de l'hème contenant la cystéine. Cette séquence absolument conservée, qui est le cinquième ligand du fer héminique, explique le pic d'absorption Soret caractéristique à 450nm rencontré pour les protéines liées au monoxyde de carbone CO. Une différence marquée entre P450s microsomaux et mitochondriaux survient au début de l'hélice L, pour laquelle tous les P450s mitochondriaux identifiés jusqu'à présent contiennent deux acides aminés chargés positivement, habituellement de l'arginine [Ozols *et al.* 1985]. Finalement, il y a une bobine nommée le *méandre*, fortement conservée dans la structure et localisée aussi sur la face proximale de la protéine. Avant les déterminations structurales en cristallographie X, cette région avait été décrite comme étant la région aromatique conservée dans les deux P450s qui avaient été séquencés [Ozols *et al.* 1985]. Donc, les éléments structuraux les plus variables sont les hélices A, B, B', F, G, H, K', les feuillets  $\beta_3$  et  $\beta_4$ , et les boucles. Une bonne partie des régions impliquées dans la spécificité au substrat est contenue dans les régions variables.

La région de l'hème et la partie de l'hélice I avoisinant la poche de liaison sont généralement bien conservées parmi les isoformes des P450s, facilitant l'activation du dioxygène sur le fer héminique. Les régions importantes sont l'hélice I, les hélices F et G, et la région  $\beta$ , qui comprend une courte hélice B', diverses boucles longues entre les hélices B et C, et 5 brins  $\beta_1$  (figure 1.1). Les hélices F et G et la région  $\beta$  variable forment le dôme de la poche du site actif (du côté distal), et modulent la liaison et la sélectivité des différents ligands grâce à leur composition chimique en acides aminés et à leur déplacement. Le déplacement de ces régions variables permet l'ouverture et la fermeture des canaux d'accès à la poche héminique. En raison des différences de repliement des parties de la protéine formant les cavités, la forme des canaux et celle de la cavité du site actif diffèrent considérablement chez les P450s. Les différences de formes des cavités affecteraient les voies de sélectivité du substrat et la catalyse enzymatique, et celle-ci, à son tour, reflèterait les différences de structure secondaire et tertiaire des protéines.

La structure principale du P450 consiste en hélices  $\alpha$  (désignées de A à L), et en feuillets  $\beta_1$  à  $\beta_4$ , avec le site catalytique autour de l'hème contenant des régions



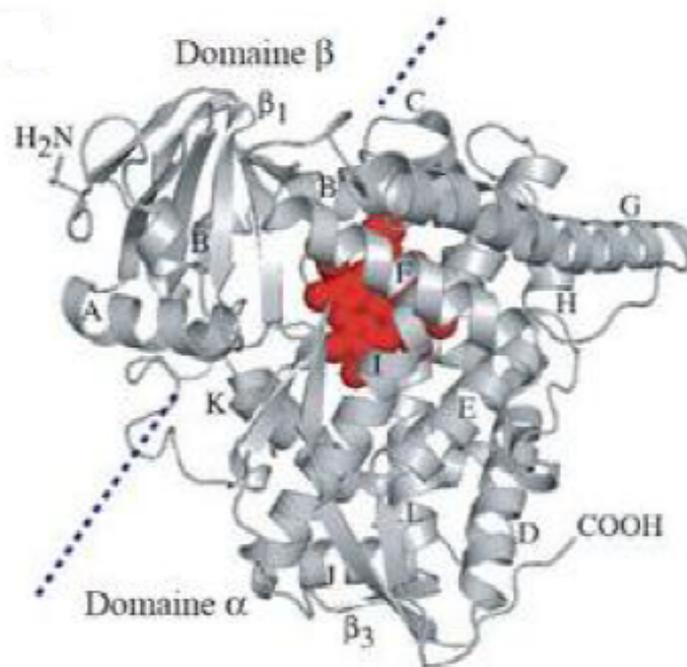
**Figure 1.1** : Carte topographique des éléments de structures secondaires du P450BMP : figure reprise de [Graham & Peterson 1999]. Les dimensions des structures secondaires ne sont pas en proportion de leur longueur de séquence primaire. Les hélices  $\alpha$  sont représentées par les rectangles. Les brins individuels des feuillets  $\beta$  sont indiqués par des flèches. Les boucles sont représentées par des lignes entre les éléments structuraux mentionnés précédemment. La région  $\alpha$ , à droite de la figure, a une prédominance d'éléments en hélices  $\alpha$ , alors que la région  $\beta$  riche en feuillets  $\beta$  est du côté gauche de la figure.

relativement petites nommées SRS (Sites de Reconnaissance du Substrat), impliqués dans la liaison du substrat et dans la catalyse (liaison au partenaire redox). Ces SRS montrent une grande variation de séquence en acides aminés, sous-jacente à la capacité des différents P450s à se lier à divers types de substrats. Les alignements structuraux de multiples P450s révèlent une flexibilité conformationnelle et indiquent que trois régions sont à l'origine de la plupart des variations de structure du squelette : la région FG (C-terminale pour l'hélice F et N-terminale pour l'hélice G), la région  $\beta$  (boucles entre les hélices B/B', boucle BC et brin 5 du feuillet  $\beta_1$ ), et la région  $\beta_4$  (feuillet  $\beta_4$  ou boucle C-terminale pour le CYP3A4) [Schuler & Sligar 2007]. La région FG est généralement plus longue dans les P450 eucaryotes, car elle interagirait avec la membrane. Les hélices FG contribuent au canal d'accès du substrat pour la plupart des P450s et, avec l'hélice B', elles peuvent présenter des mouvements d'ouverture/fermeture au niveau de la liaison au substrat. En plus de ce canal d'accès conventionnel, la boucle BC peut former un canal différent en s'éloignant de l'hélice G pour former une fissure le long de l'hélice I, ainsi que cela a été observé pour le CYP154C1 de *Streptomyces coelicolor* [Poulos & Johnson 2005]. Le CYP51B1 du *Mycobacterium tuberculosis* a un pliage unique dans l'hélice I, qui crée une fissure perpendiculaire à l'hème le long de la boucle BC et de l'hélice C, et cela augmente le nombre de chemins possibles d'entrée et de sortie [Poulos & Johnson 2005].

La région du canal 2 est groupée dans des sous-classes nommées 2a, 2b, etc., selon les éléments de structures secondaires bordant la route du ligand, telle qu'elle apparaît à la surface de protéine. Le bloc BC est défini comme s'étendant du début de l'hélice B à la fin de l'hélice C, et la région FG comme s'étendant du début de l'hélice F à la fin de l'hélice G. Bien que le repliement global des CYPs soit bien conservé, la longueur et la structure secondaire des blocs BC et FG varient considérablement à l'intérieur de la famille des P450s. La flexibilité de ces deux blocs est la source principale du changement conformationnel pour une isoforme donnée [Bridges *et al.* 1998].

Des simulations de dynamique moléculaire sur divers P450s fournissent des résultats concordant avec l'importance des régions identifiées par des études structurales des routes d'entrée et de sortie du substrat. Les données indiquent aussi de possibles différences entre les canaux des P450s bactériens et ceux des mammifères [Cojocaru *et al.* 2007]. Les simulations de dynamique moléculaire du P450cam et du P450BM-3 indiquent que la liaison au substrat et la sortie du produit exigent un déplacement substantiel autour du canal d'accès du substrat.

Les comparaisons entre les structures cristallines des P450s bactériens et celles des mammifères ont révélé une différence structurale commune : le domaine  $\beta$  est déplacé vers le côté proximal du plan de l'hème dans toutes les enzymes des mammifères autres que la famille du CYP51 (stérol déméthylase). Ceci provient de petites insertions de peptides dans les hélices F' et G', qui interagissent avec la membrane des P450s des mammifères [Denisov *et al.* 2012]. Ainsi, étant donné la diversité de substrats et les partenaires redox dans la superfamille, les changements structuraux accompagnant la catalyse sont probablement plus complexes que de simples ouvertures et fermetures des P450s, ces interactions avec la membrane elle-même permettant la reconfiguration structurale du P450 pour l'accès du substrat hydrophobe au site actif. Le domaine  $\beta$  est un domaine ancré dans la membrane plasmique, très impliqué dans le recrutement des ligands car il contient les bouches des canaux. Les domaines  $\alpha$  et  $\beta$  sont visualisés figure 1.2.



**Figure 1.2 :** Eléments de structures secondaires du cytochrome P450 BM-3.

La première topologie d'un P450 a été établie sur CYPBMP [Graham & Peterson 1999], puis a été étendue aux autres P450s en adaptant cette topologie générique aux différences entre P450s au niveau de certaines structures secondaires, présentes ou non. Cette topologie devient la référence dans l'étude structurale de tous les P450s. Nous nous baserons dessus pour notre étude sur le 3A4 et le 1A2.

### 1.5.2 Régions de reconnaissance et de liaison de substrat et du partenaire redox

Les régions non conservées ou variables mentionnées précédemment sont d'habitude associées à la liaison au substrat et au partenaire redox. Des régions impliquées dans la reconnaissance du substrat ont été identifiées [Gotoh 1992], et sont désignées par des numéros SRS (Sites de Reconnaissance du Substrat). Les SRS ont été déterminés sur la base des comparaisons de séquences entre la famille CYP2 et le P450cam, sur lesquels on se réfère dans la littérature des P450.

Les régions variables associées à la liaison du substrat sont les hélices A, B, B', F et G et leurs boucles adjacentes. Les boucles B-B' et B'-C bordent le site actif (SRS-1), et les hélices F et G et leur boucle font partie du canal d'accès et du plafond du site actif (SRS-2 et SRS-3). Le coude  $\beta$  à l'extrémité du feuillet  $\beta_4$  pointe dans le site actif (SRS-6), tout comme la région N-terminale du brin  $\beta_{1-4}$  (SRS-5). L'autre région associée à la liaison du substrat et d'importance dans le mécanisme réactionnel (mais conservée au sein de la famille des P450s), est la partie centrale de l'hélice I (SRS-4).

Tandis que la plupart des régions ci-dessus sont dans la poche héminique des P450s, et impliquées dans l'orientation du substrat dans la poche; l'hélice A, la boucle FG et les brins  $\beta_1-1$  et  $\beta_1-2$  semblent être impliquées dans la reconnaissance du substrat (voir section 1.10.2). Des mutations dans ces zones altèrent la liaison aux substrats et produisent des mutants [Graham-Lorence *et al.* 1997, Oliver *et al.* 1997]. L'hypothèse émise sur les P450s eucaryotes microsomaux associés à la membrane, est que la boucle FG et des parties des feuillettes  $\beta_1$  et  $\beta_2$  peuvent être associées à la membrane au raison de leur hydrophobie accrue, augmentant probablement ainsi l'accès au substrat via la membrane [Graham-Lorence *et al.* 1995].

Chez les P450s eucaryotes, le paradigme des interactions avec le substrat propose que les parties du site de liaison soient flexibles et ajustent leur forme pour accueillir le substrat. De plus, il semble y avoir un potentiel de déplacement significatif autour de la bouche du canal d'accès.

On sait depuis longtemps que le nombre de molécules de P450 sur la membrane du réticulum endoplasmique est nettement supérieur au nombre de molécules de la NADPH-P450 réductase [Peterson *et al.* 1976]. Ceci, ajouté au fait que la population de P450s est hétérogène, a permis de savoir comment s'effectue l'interaction entre les P450s et l'unique NADPH-P450 réductase [Backes *et al.* 1998]. Comprendre cette interaction est important pour comprendre la production potentielle de radicaux anions superoxydes et/ou du peroxyde d'hydrogène par les P450s, qui sont réduits en l'absence de substrat lié. Ainsi, l'interaction substrat-P450 peut provoquer un changement conformationnel qui altère la liaison du P450 à la réductase [Li & Poulos 1997]. Cette réaction de découplage pourrait mener à des dommages tissulaires via des processus nécrotiques ou apoptotiques.

Le classement des acides aminés en fonction de leur importance évolutive relative, effectué à la surface des P450s, indique que l'interaction du partenaire redox s'effectue par la face proximale du P450 et que la liaison entre le P450 et la réductase diffère selon les familles des P450 [Lichtarge *et al.* 1996]. La comparaison des distributions de charges entre les P450s mitochondriaux et microsomaux, révèle des plaques hydrophobes et cationiques uniques sur la face proximale du P450 [Peterson & Graham 1998]. Ce sont bien la face et la région proximales qui sont impliquées dans l'interaction avec le partenaire redox, selon des études sur le P450cam et la putidaredoxine [Stayton & Sligar 1990, Unno *et al.* 1996], et par mutagenèse dirigée des P450s microsomaux et mitochondriaux [Shen & Strobel 1993, Bridges *et al.* 1998, Wada & Waterman 1992, Pikuleva *et al.* 1999].

### 1.5.3 Différences structurales et fonctionnelles entre les P450

Il y a une séparation évolutive entre les P450s procaryotes et eucaryotes : les premiers sont des enzymes cytoplasmiques solubles et les seconds sont des protéines intégrées à la membrane plasmique [Munro *et al.* 2007]. Une extension peptidique transmembranaire de 30 à 50 acides aminés environ, au niveau de la région N-terminale des P450s eucaryotes, les lie du côté de la matrice mitochondriale (pour certaines formes de métaboliseurs de stéroïdes) ou du réticulum endoplasmique [Poulos & Johnson 2005]. La première structure d'un P450 eucaryote résolue aux rayons X est celle de la progésterone 21-hydroxylase CYP2C5. Pour cristalliser la structure, la région de l'ancre N-terminale transmembranaire a été enlevée ( $\Delta$  3-21), des mutations ont été introduites dans la région N-terminale pour augmenter l'expression, et 5 acides aminés ont été modifiés en ceux trouvés dans le CYP2C3 apparenté (la progésterone 16 $\alpha$ -hydroxylase) afin d'améliorer la

solubilité [Williams *et al.* 2000]. Ainsi, le noyau de la structure du P450 a été conservé, mais des changements clés ont été apportés pour améliorer la cristallisation du 2C5. Actuellement, il y a 83 structures uniques de P450 déposées dans la Protein Data Bank. Le repliement a une forme de prisme triangulaire, avec l'hème enfoui au centre de la molécule et à peu près parallèle à l'une des faces triangulaires (figure 2, [Munro *et al.* 2013]). Des variations subtiles dans la position et la flexibilité d'éléments structuraux secondaires et dans la longueur des régions d'interconnexion de boucle, contribuent à la diversité des sites catalytiques et à la réactivité des P450s (figures 2 et 4, [Munro *et al.* 2013]), mais pas seulement, car ces variations contribueraient à des différences dans les cavités et notamment dans les canaux, et ce qui différencierait l'accessibilité des ligands aux P450s.

Les hélices I et L et la région de liaison à l'hème sont fortement conservées. Le méandre et les feuillets  $\beta_1$  et  $\beta_2$  sont généralement conservés parmi les P450s solubles. Les régions FG et BC ont montré une variabilité parmi les P450s solubles.

Pour conclure, la structure générale des P450s et leur cœur structural semble être conservés. Cependant, le mode d'interaction avec les substrats, les donneurs d'électrons et les membranes varient. Ainsi, il reste de nombreux champs de recherche.

## 1.6 L'hypothèse de canalisation

Les P450s contribuent au stress oxydant par un mécanisme complexe impliquant la formation de  $\text{H}_2\text{O}_2$  et de  $\text{O}_2^-$  [Guengerich 2006]. Pour catalyser une oxydation par un CYP, le substrat doit se lier dans le site actif du côté distal de l'hème. Bien que les étapes d'oxydation soient examinées depuis longtemps [White & Coon 1980, Sligar *et al.* 1984, Guengerich 2003, Isin & Guengerich 2008], l'entrée et la sortie des composés vers et depuis le site actif restent peu clairs. Des points ci-dessus nous suggérons un rôle décisif des canaux dans la fonction des CYPs. Ils peuvent expliquer la grande diversité de spécificité du substrat même pour deux isoformes proches en structure primaire. Les canaux d'accès contrôlent la spécificité du substrat pour les CYP1A [Urban *et al.* 2015]. La flexibilité structurale est essentielle pour permettre aux composés chimiques d'aller vers et depuis le site actif, et il a été montré qu'elle se corrèle avec des préférences de substrat pour plusieurs CYPs, y compris pour le 3A4 [Skopalik *et al.* 2008]. Il a été montré que la flexibilité du CYP est proportionnelle à sa spécificité et à la diversité de ses substrats ("*promiscuity*") [Skopalik *et al.* 2008]. Cette étude a évalué la flexibilité de trois isoformes : le 2A6, qui accepte une faible diversité des substrats, le 2C9, connu pour être modérément spécifique, et le 3A4, qui accepte la plus grande diversité de substrats. Sur les dix régions flexibles trouvées, qui varient significativement suivant ces trois isoformes, les régions les plus rigides étaient dans le 2A6, et les plus flexibles dans le 3A4.

Dans cette thèse, on se propose de tester l'hypothèse de la canalisation pour certaines isoformes, afin d'arriver à une meilleure compréhension de la reconnaissance du substrat. L'idée est d'explorer la variabilité de canaux au sein d'une même isoforme, et entre isoformes proches, ainsi que leur localisation et leur morphologie respective. Cette exploration se fera à la lumière de la palette de substrats que chacune des isoformes peuvent reconnaître et métaboliser. Mais avec quels outils ?

Il existe des dizaines de logiciels de calculs de cavités et de canaux, qui donnent une variété de résultats en raison de la diversité de structures de données de sortie [Benkaidali

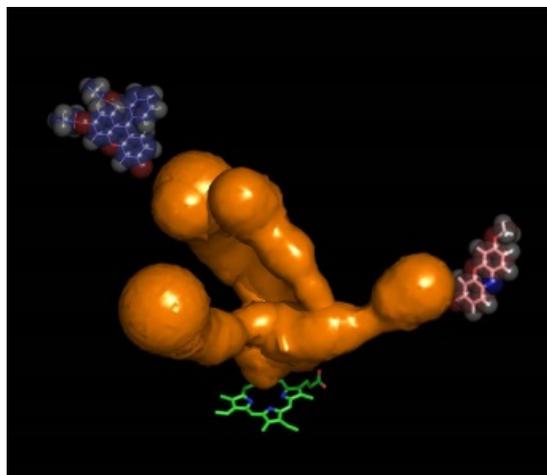
*et al.* 2014]. Aussi, il est plutôt difficile de comparer ces résultats. Par exemple, la nomenclature basée sur celle établie par [Cojocaru *et al.* 2007] avec CAVER sur toutes les structures cristallisées des CYPs disponibles dans la PDB en mars 2006, soient au total 143 fichiers PDB, 192 chaînes et 26 CYPs, ont indiqué 14 canaux, tandis que le CYP3A4 seul (code PDB 1TQN, 1 chaîne, incluse dans l'étude de 2006), a donné lieu en 2012 à lui tout seul à 21 canaux [Berka *et al.* 2012] avec MOLEonline 2.0, un des successeurs de CAVER. En fait seulement 3 canaux ont été attribués au 3A4 dans l'étude de 2006. De telles contradictions ne sont pas inhabituelles. Elles sont apparues aussi entre les canaux du P450cam (CYP101) calculés avec CAVER et MOLEonline et ceux calculés par [Jankowsky *et al.* 2011], et la situation reste peu claire malgré l'aide de plusieurs simulations de dynamique moléculaire [Lüdemann *et al.* 1997, Lüdemann *et al.* 2000a, Lüdemann *et al.* 2000b]. Les développeurs de MOLE 2.0. ont récemment établi une anatomie des canaux de 4306 enzymes, par identification des caractéristiques géométriques et physico-chimiques, telles que le nombre, la longueur, et l'hydrophobicité des canaux. Cette étude a défini une composition générique en acides aminés des canaux, commune à tous les enzymes, caractérisée par les différences des compositions entre le site actif, la surface et l'intérieur de la protéine. La composition des parois du canal est signalée comme étant différente (acides aminés aromatiques, chargés et polaires), de la composition moyenne du cœur des enzymes (acides aminés aliphatiques hydrophobes), aussi bien que de la celle de la surface de la protéine [Pravda *et al.* 2014].

Les expérimentateurs sont laissés face à des dizaines de logiciels publiés, avec lesquels ils obtiennent de nombreux résultats, pouvant être contradictoires, sans pouvoir toujours conclure quels sont ceux devant être retenus. Des comparaisons simples sont donc nécessaires. Une revue que nous venons de publier en 2014 sur les algorithmes géométriques de calcul de canaux a permis d'en recenser pas moins de 35 [Benkaidali *et al.* 2014], alors qu'une revue de 2010 en mentionnait déjà une vingtaine [Pérot *et al.* 2010]. Ces algorithmes peuvent aider dans la définition du nombre de canaux accessibles, par exemple en utilisant une sonde sphérique d'un rayon compatible avec la taille d'une molécule d'eau. Ces méthodes butent sur deux écueils, parfois simultanés :

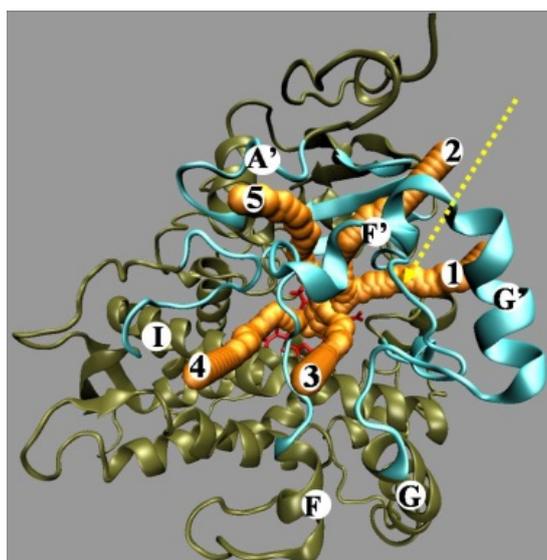
1. Soit les canaux sont "universels" (ex. CAVER, MOLE), c'est à dire qu'ils ne tiennent pas compte de la taille et de la forme du ligand : c'est comme si l'on considérait que les cavités accessibles à une molécule d'eau sont les mêmes que celles accessibles à un aromatique polycondensé ou à un macrolide.
2. On utilise comme modèle de ligand une sphère de rayon arbitraire à fixer par l'utilisateur (ex., les méthodes de type alpha-shape).

Les canaux définis par CAVER pour le CYP1A1 sont représentés figure 1.3 pour le CYP1A1 et figure 1.4 pour le CYP3A4. Quatre et cinq canaux sont détectés respectivement, mais il n'est pas possible de certifier s'ils sont pertinents. Il est néanmoins intéressant de noter que le canal labellisé 5 dans le CYP3A4 correspond à un canal dans lequel la deuxième molécule de kétoconazole est partiellement engagée dans la structure cristalline [Ekroos & Sjögren 2006].

Or, dans notre étude [Benkaidali *et al.* 2014], nous avons démontré que les formes des ligands étaient plutôt plates ou allongées que globulaires, et donc que le modèle sphérique était inadapté. Ce dernier n'offre comme avantage que sa simplicité de calcul et de traitement, mais il est physiquement irréaliste. Il faut remarquer l'extrême difficulté du problème : en premier lieu, une protéine apparaît géométriquement sous forme d'une



**Figure 1.3** : Canaux définis par CAVER pour le CYP1A1 (modèle par homologie basé sur la structure cristalline du CYP1A2).



**Figure 1.4** : Canaux définis par CAVER pour le CYP3A4.

liste de points dans l'espace 3D, et définir des poches et des canaux dans ces conditions est compliqué, surtout lorsqu'il faut identifier les parois pour des tests de gêne stérique (reconnaissance des formes), ainsi que les surfaces et volumes des poches et des canaux. Les méthodes de la littérature sont inadéquates de ce point de vue.

Pour conclure, à l'appui de l'hypothèse des canaux, il est intéressant de citer deux autres faits expérimentaux. D'abord, les cycles particulièrement lents des CYPs humains que nous avons déjà mentionnés. Dans la première phase du cycle catalytique, schématiquement, le CYP produit une seule espèce réactive en attente de l'approche de n'importe quelle molécule acceptrice d'électrons pouvant faire avancer la réaction chimique. La compétition entre substrats accepteurs face à un seul donneur fait de la discrimination du substrat un événement contrôlé cinétiquement plutôt que thermodynamiquement. Dans la littérature, de nombreuses données indiquent que les constantes cinétiques ne sont pas limitantes en comparaison à l'acte catalytique, dont les fréquences s'étendent de 1 à 20 min<sup>-1</sup> [Isin & Guengerich 2006]. D'autres études [Kim *et al.* 2006], basées sur des effets isotopiques, montrent que, chez les mammifères, l'association de substrats et la dissociation des CYPs sont des étapes rapides et que le l'étape limitante est l'acte catalytique lui-même (CYP1A2 et CYP2E1). Chez le lapin, pour le CYP1A2 la liaison au substrat est très rapide ( $k > 10^6 \text{M}^{-1} \cdot \text{s}^{-1}$ ), et l'étape de sortie du produit n'est pas limitante [Guengerich *et al.* 2004]. Ces résultats suggèrent que le parcours des canaux peut être considéré comme libre pour l'accès de substrat au site actif, à condition que leur forme géométrique soit compatible.

D'autre part, une série d'études sur le CYP3A4 [Kenworthy *et al.* 1999, Kenworthy *et al.* 2001], basées sur les mesures de compétition de 34 composés (surtout des médicaments) et sur 10 réactions métaboliques tests, a mené à l'identification de trois groupes principaux de comportement d'inhibition, identifiés par des techniques de clustering. Cette classification, qui définit les familles non compétitives de substrats (types: testostérone, midazolam, et nifédipine), est utilisée par certaines industries pharmaceutiques pour des autorisations de mises sur le marché. Nous reviendrons sur cette analyse dans le chapitre 3. Si cette classification biochimique est pertinente et confirmée par d'autres laboratoires, il pourrait être intéressant de la corrélérer à la sélectivité par canalisation puis d'utiliser ces données pour validation par une approche informatique.

Finalement, pour traiter la plasticité des P450s, l'une des équipes impliquées dans notre projet [Bouvier *et al.* 2008], a appliqué l'analyse de modes normaux à différents isoformes de CYPs humains, pour examiner la tendance de leurs structures à de grandes transitions conformationnelles impliquant un canal alternativement ouvert et fermé. De cette analyse, on peut conclure qu'il n'y a aucune tendance marquée du CYP3A4 pour de grandes transitions conformationnelles qui pourraient expliquer l'adaptation aux substrats volumineux. Ceci suggère que les mouvements des molécules dans ou hors du site actif soient essentiellement conduits par des processus de diffusion à travers les chaînes latérales bordant les canaux.

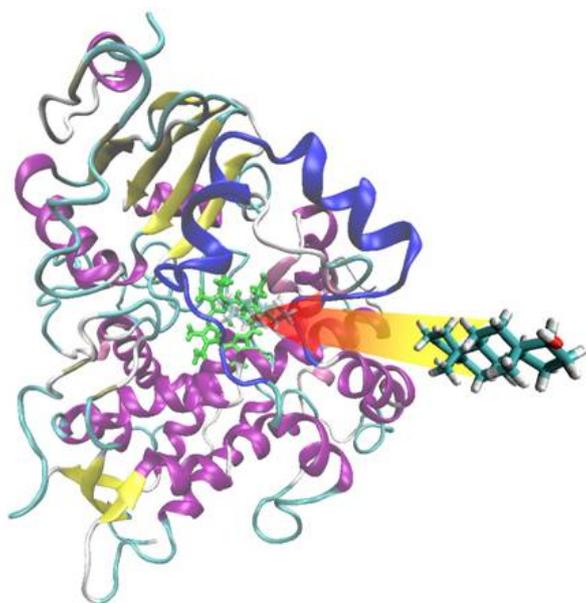
Ainsi, plusieurs canaux d'accès peuvent intervenir simultanément dans l'activité multi-substrat du CYP3A4, permettant ainsi la liaison du substrat et le métabolisme non compétitif. Il y a un intérêt majeur à développer des outils de calcul rapides pouvant discriminer des substrats dans une base de données chimique, qui peuvent avoir accès au site actif d'un CYP donné. Les règles d'interaction entre les substrats et le site actif de l'enzyme (cavité qui contient l'hème) restent à établir, notamment sur la base de critères énergétiques. La flexibilité des CYPs est établie, mais la prédiction des activités

catalytiques de chaque CYP reste difficile [Guengerich 2013]. Les calculs de déformabilité et de flexibilité des ligands et surtout des cytochromes P450s, habituellement effectués par des simulations en dynamique moléculaire (ou plus rarement par des calculs de modes normaux), ont une durée rédhibitoire dans un contexte de criblage virtuel de chimiothèques de centaines de milliers de substrats potentiels. Ces calculs doivent être réservés à l'étude spécifique de certains complexes CYP-ligand, et ne peuvent donner lieu à l'établissement d'un protocole simple de simulation de dynamique moléculaire qu'avec une connaissance préalable suffisante des canaux menant au site actif.

## 1.7 Simulations de dynamique moléculaire

Les études par simulation de dynamique moléculaire des canaux de certains CYPs ont été effectuées [Wade *et al.* 2004, Schleinkofer *et al.* 2005, Skopalik *et al.* 2008, Fishelovitch *et al.* 2009, Cojocar *et al.* 2012]. Plusieurs études de modélisation moléculaire et de simulation de dynamique moléculaire ont été effectuées pour étudier les interactions substrats-canaux dans le CYP3A4 [Lafite *et al.* 2007, Nguyen *et al.* 2008, Goel *et al.* 2008]. Dans chaque simulation, le substrat est guidé de l'extérieur vers le site actif comportant le groupement hème par des contraintes énergétiques douces (figure 1.5). Cette méthode a l'avantage d'être basée sur des calculs d'énergie et aide à quantifier l'interaction (et donc l'affinité) entre le ligand et chaque canal [André *et al.* 2004]. Mais l'examen de la spécificité possible du ligand à chaque canal qui peut guider le substrat au site actif hémique, implique que la simulation en dynamique moléculaire puisse être répétée sur un temps suffisant pour donner des résultats statistiquement interprétables. Pourtant chaque simulation nécessite un long temps de calcul dans le cas d'un solvant explicite (la trajectoire du ligand peut aller jusqu'à 40Å). Les simulations de dynamique moléculaire dirigées (SMD : *Steered Molecular Dynamics*) ont d'autres inconvénients : la contrainte appliquée sur le ligand pour simuler le parcours vers l'hème est un biais énergétique qui affecte l'évaluation de l'affinité ligand/canal, tandis que la complexation du ligand par l'enzyme est plus probablement un processus de diffusion libre à travers les divers canaux de la protéine. Les simulations de dynamique moléculaire à accélération aléatoire (RAMD : *Random Acceleration Molecular Dynamics*) ont été appliquées aux cytochromes bactériens P450cam et P450-BM3 [Lüdemann *et al.* 2000a, Lüdemann *et al.* 2000b], et permettent de voir la sortie du ligand sur un échelle de quelques nanosecondes à quelques picosecondes, mais cela ne correspond pas aux échelles de temps de sortie du ligand. Un protocole combinant simulation de dynamique moléculaire conventionnelle et SMD a été utilisé [Lüdemann *et al.* 2000a, Lüdemann *et al.* 2000b, Li *et al.* 2011b], et appliqué sur divers CYPs mammaliens, dont les isoformes humaines 2A6 [Li *et al.* 2011b] et 2E1 [Shen *et al.* 2012].

Dans les simulations classiques de dynamique moléculaire, on applique aux atomes un champ de force empirique qui prend en compte les interactions électrostatiques et Coulombiennes. Ces simulations considèrent la migration du ligand et la flexibilité de la protéine. Il existe une revue générale récente sur les simulations de dynamique moléculaire de sortie des ligands dans le cas des CYPs mammaliens [Yu *et al.* 2013]. Ces simulations sont parfois associées à des calculs géométriques des canaux, comme pour le 3A4 avec MolAxis [Yaffe *et al.* 2008a, Yaffe *et al.* 2008b]. Le principal problème que l'échelle de



**Figure 1.5 :** Exemple d'un docking de dynamique moléculaire dirigée (SMD) de la testostérone dans le canal du CYP3A4 transitoirement ouvert, défini par les hélices F'-G'.

temps nécessaire pour lier (ou délier) le ligand est très supérieure à ce qui est faisable avec des simulations classiques, de l'ordre de 10 à 100 ns.

Parfois le ligand est suffisamment petit pour que des modifications mineures, telles que changer la température du ligand lié à la protéine, peut suffire à favoriser la sortie du ligand puis à identifier le canal [Arroya-Mañez *et al.* 2011]. Les méthodes de Monte Carlo grand canonique, associée à la dynamique moléculaire ou à la dynamique Brownienne, ont aussi été utilisées pour suivre le mouvement de petits ions dans les canaux [Im *et al.* 2000, Woo *et al.* 2004, Boda *et al.* 2014]. Ces méthodes ont été utilisées pour simuler de petits ligands (ions et eau) passant dans les canaux, mais n'ont pas encore été appliquées à de plus gros ligands.

A ce jour, la plupart des simulations de migration de ligands dans les canaux utilisent des méthodes soumettant le ligand à une force externe depuis le site actif. L'utilisation de contraintes dépendant du temps dans la SMD ou dans la TMD (*Targeted Molecular Dynamics*) pour forcer un déplacement dans une direction donnée est utile dans les problèmes où les simulations sans contraintes prennent trop de temps. La SMD ou la TMD ont été appliquées à l'extraction de ligands de divers récepteurs, fournissant ainsi des informations sur les voies possibles d'échappement des ligands (voir [Wolf 2015] et références citées). Par contre, lorsque le ligand est totalement enfermé dans la protéine, ce qui est le cas pour les CYPs, la direction à choisir pour l'échappement est ambiguë. Une alternative est alors la REMD (*Random Expulsion Molecular Dynamics*) : une force de direction aléatoire est appliquée au ligand jusqu'à ce qu'il sorte de la protéine [Lüdemann *et al.* 2000a]. Si le ligand se heurte à une barrière, la direction de la force est modifiée, et ainsi le ligand peut "sonder" la protéine pour identifier les canaux d'évacuation. Chaque essai réussi permet d'identifier un trajet possible pour l'évacuation du ligand. De multiples essais sont nécessaires pour explorer les protéines ayant plusieurs canaux, et on comptabilise le nombre et la fréquence des passages du ligand dans chaque canal

[Lüdemann *et al.* 2000a]. Les simulations par REMD ont été appliquées aux isoformes CYP101, CYP102A1, et CYP107A1 [Winn *et al.* 2002]. Les structures des canaux a été explorée par REMD [Wade *et al.* 2004] : elles sont classées comme, soit étant fermées, soit ayant plusieurs ouvertures, soit étant largement ouvertes, suivant que la surface accessible au solvant (calculée avec un sonde sphérique de 1.4Å représentant une molécule d'eau), présente le site actif isolé (fermé) ou connecté (plus ou moins ouvert) sur l'espace contenant le solvant.

Dans la méthode IterTunnel [Kingsley & Lill 2014a], qui s'appuie sur l'approche géométrique de prédiction de canaux MolAxis [Yaffe *et al.* 2008a, Yaffe *et al.* 2008b], on utilise la SMD (*Steered Molecular Dynamics*), pour attirer le ligand le long de canaux pré-calculés, qui sont périodiquement mis à jour au fur et à mesure que les mouvements du ligand dans la protéine ouvrent ou ferment de nouveaux canaux [Kingsley & Lill 2014b]. Le potentiel dont dérive la force moyenne appliquée au ligand passant dans un canal, est calculé par une méthode d'échantillonnage de type *umbrella sampling*, combinée à une analyse WHAM *Weighted Histogram Analysis Method* capable de prendre en compte les états intermédiaires, et ainsi on évalue énergétiquement chaque canal [Kingsley & Lill 2014b]. Le type de résultat diffère de celui obtenu par la méthode REMD. Pour cette dernière, les multiples simulations pour un canal donné fournissent une image plus complète des événements durant le transit du ligand dans le canal, mais on n'accède pas explicitement à l'information énergétique. A l'inverse, les simulations répétées sont effectuées par IterTunnel en différents points du canal (*umbrella sampling*), à l'opposé d'un échantillonnage répété du transit pour le canal entier, cet échantillonnage n'étant fait qu'une fois par IterTunnel. Le résultat fourni par IterTunnel est un profil détaillé d'énergie libre de cet événement spécifique du transit, mais ne porte pas sur d'autres événements possibles le long du même trajet de migration du ligand.

IterTunnel a été appliqué au CYP2B6 comme système modèle de la migration du ligand chlorophenylimidazole, puis les canaux ont été comparés à ceux prédits à l'aide un ensemble de 100 structures générées par une simulation de dynamique moléculaire utilisant CAVER 3.0 [Chovancova *et al.* 2012]. Des différences significatives ont été trouvées [Kingsley & Lill 2014b] : 11 voies d'échappement potentiel ont été trouvées par IterTunnel, contre 5 avec CAVER 3.0. En incluant la migration du ligand dans le procédé d'identification du canal, l'ouverture du canal a été observée comme étant un résultat des changements conformationnels induits par le ligand dans la protéine. Pour l'ensemble de 100 structures, de tels canaux n'ont pas été vus car ces changements conformationnels ne peuvent survenir sans inclure le ligand à différents points du chemin de sortie. IterTunnel a donc permis de prouver, d'une part l'influence de la malléabilité protéique sur la forme et l'accessibilité des canaux, et d'autre part l'influence du ligand sur le reformattage des canaux en raison de son interaction avec la protéine. Il résulte de ce procédé l'exposition de nouveaux canaux et la fermeture de canaux préexistants au fur et à mesure de la migration du ligand depuis le site actif.

Des simulations ont été faites sur six isoformes majeures métabolisant des médicaments (2A6, 2B4, 2C9, 2D6, 2E1, et 3A4), sur 100ns et davantage en solution aqueuse [Hendrychova *et al.* 2012], et une conversion de la forme ouverte à la forme fermée a été observée. Durant les simulations, les canaux des ligands restent ouverts dans le 2C9 et le 2B4, fermés temporairement dans le 3A4 et le 2D6, et fermés dans le 2A6 et le 2E1, et il est indiqué que la flexibilité du site actif et de la boucle FG contribuent aux mouvements les plus importants dans les 6 CYPs.

Des simulations par RAMD et SMD ont montré que la rotation des acides aminés aromatiques en entrée des canaux des ligands, ainsi que la formation et la dissociation des liaisons hydrogène accompagnées de mouvements du squelette, contribuent significativement à la sortie du ligand des canaux (voir [Yu *et al.* 2013] et références citées).

Les simulations vues jusqu'ici pour chaque CYP ont été faites en milieu aqueux. La modélisation des CYPs liés à leur membrane peut renseigner sur la dynamique protéique, la cinétique de liaison au ligand, et la sélectivité de liaison des CYPs. Il n'y a pas de structure de CYP déterminée expérimentalement dans une bicouche lipidique, et la simulation d'un tel système reste difficile. Un modèle de CYP mammalien lié à la membrane existe, combinant une approche à gros grains et une simulation atomique détaillée [Cojocaru *et al.* 2011]. Dans le modèle du 2C9, la boucle FG est insérée dans la membrane de telle manière que les canaux des ligands 2a et 2f mènent dans la membrane. Les autres canaux, dont 2ac, 2b, 2c, 2d, 2e, et S, mènent à la solution aqueuse. Cette orientation du 2C9 au-dessus de la membrane, telle que certains des canaux potentiels du ligand pointent vers la membrane alors que les autres pointent vers le solvant aqueux, suggèrent que le canal préféré par un ligand donné dépend de son hydrophobie. La membrane peut affecter les boucles flexibles des blocs BC et FG (le bloc BC est en interaction partielle avec la bicouche et la région FG est inséré dans la bicouche). Dans ces situations, la mobilité de ces régions se réduit à la forme liée à la membrane, par comparaison à la forme soluble [Berka *et al.* 2011, Cojocaru *et al.* 2011, Denisov *et al.* 2012]. Cela suggère que la membrane pourrait contraindre les changements conformationnels à ces régions protéiques et masquer la conversion entre les états ouverts et fermés. Il existe d'autres études permettant de voir l'influence de la membrane sur la dynamique des CYPs mammaliens et leurs canaux. Des simulations de la forme soluble du 3A4 et de sa forme liée à la membrane montrent que le canal 2b reste ouvert dans les deux formes, alors que le canal 2e reste ouvert dans la forme soluble, et fermé dans la forme liée à la membrane [Denisov *et al.* 2012]. Le mouvement des hélices F et G provient de la fermeture du canal 2e, et n'a été observé que dans la simulation de la forme liée à la membrane [Denisov *et al.* 2012]. La simulation du 2C9 dans une bicouche phospholipidique indique que le canal du solvant reste le plus ouvert et le plus adéquat pour la sortie du produit [Berka *et al.* 2011].

Le principal inconvénient des simulations intensives de dynamique moléculaire est qu'elles demandent des ressources de calcul importantes. Dans toutes les variantes de ces méthodes, un seul canal est identifié à la fois, et donc plusieurs simulations sont nécessaires lorsque les substrats ou produits peuvent entrer ou sortir par plusieurs canaux. De plus, ces simulations ne marchent pas toujours et peuvent indiquer des canaux irréalistes lorsqu'elles sont utilisées de manière hasardeuse. Bien que ces méthodes soient capables d'apporter des informations détaillées sur le comportement des ligands dans les canaux, elles peuvent décourager des utilisateurs occasionnels.

Finalement, l'échelle de temps accessible en simulation de dynamique moléculaire est de l'ordre de la dizaine voire de la centaine de nanosecondes, ce qui est très disproportionné en comparaison avec la lenteur caractéristique des temps de réactions enzymatiques des CYPs microsomaux. Nous avons ainsi besoin de nouveaux outils informatiques, plus rapides et plus orientés "haut débit", pour pouvoir aborder le problème avec des centaines ou des milliers de substrats différents. Ce nouvel outil, basé sur des approches géométriques, est détaillé dans le chapitre 2. De tels outils de modélisation de canaux dans

les structures protéiques sont inhabituels, car les problèmes de mécanismes enzymatiques sont rarement élaborés en termes de canalisation de ligand.

## 1.8 Importance de la lipophilie des substrats des P450s

Le rôle de la lipophilie d'un composé dans l'activité biologique et les interactions est bien établi [Kennedy 1997, van de Waterbeemd & Testa 1987, Kubinyi 1997, Livingstone 2000, Clark 2001, Cronin *et al.* 2001, Kerns & Di 2003].

L'importance du caractère lipophile pour une découverte de médicament a été soulignée dans le célèbre concept de la "règle des cinq" [Lipinski *et al.* 1997]. Pour les P450, l'importance de la lipophilie d'un substrat pour sa traversée dans un canal a été citée très tôt dans la littérature [Smith *et al.* 1997a]. La lipophilie d'un composé a une importance clé pour l'affinité de liaison au P450 et à la sélectivité de l'enzyme ou à la clairance [Smith 1994, Smith *et al.* 1997a, Smith *et al.* 1997b, Smith & van de Waterbeemd, 1999], avec une certaine influence sur le taux catalytique relatif [Lewis & Dickins 2002, Lewis & Dickins 2003]. La lipophilie a été discutée concernant les enzymes métabolisant les médicaments pour les P450s humains des familles CYP1, CYP2 et CYP3 [Lewis *et al.* 2004]. A partir d'une compilation étendue de valeurs du  $\log(P)$  pour des substrats du P450, la pertinence de la lipophilie et d'autres facteurs se rapportant à l'affinité de liaison au P450 sont expliquées, menant à la formulation de relations de lipophilie dans les substrats de chaque enzyme P450 humaine impliquée dans le métabolisme du médicament. En outre, des valeurs du  $\log(P)$  pour des substrats du P450 semblent représenter des marqueurs pour la sélectivité enzymatique. En plus du rôle important des liaisons hydrogènes et des énergies d'interaction  $\pi-\pi$ , la désolvatation du site actif P450 apporte une contribution majeure à l'énergie de liaison de substrat.

Chaque famille et sous-famille de P450s présente une sélectivité pour des substrats d'un type particulier, bien qu'il puisse y avoir un certain degré de chevauchement selon la classe chimique des composés en question [Lewis 2003a]. Des détails sont connus grâce à plusieurs études QSAR (corrélations quantitatives entre structures et activités) sur des substrats des P450s, où le  $\log(P)$  est un descripteur principal ou unique [Hansch & Zhang 1993, Gao & Hansch 1996]. Le  $\log(P)$  est aussi supposé être important dans l'évaluation des demi-vies métaboliques chez l'homme, selon une étude chez le rat [Sarver *et al.* 1997]. La lipophilie serait donc le composant principal de la liaison du substrat aux P450s, probablement en raison de la désolvatation, où le  $\log(P)$  du composé (ou le  $\log(D_{7.4})$ ) représente un marqueur utile du caractère lipophile global [Lewis 2003b]. La lipophilie est un des paramètres clés pour évaluer les propriétés ADME/Tox (Absorption, Distribution, Métabolisme, Elimination/Toxicité) de nouvelles entités chimiques [Hansch 1969, Leo *et al.* 1971, van de Waterbeemd *et al.* 2001, Modi 2003], et donc cette quantité est aussi très pertinente pour déterminer parmi des composés les substrats plausibles du P450.

### 1.8.1 Rôle de la lipophilie dans l'affinité de liaison

En plus du rôle de la lipophilie discuté auparavant, il y a plusieurs facteurs contribuant à l'énergie totale de liaison enzyme-substrat. L'affinité de liaison substrat-

P450 peut être formulée dans l'équation (1), qui estime l'énergie de liaison  $\Delta G_{bind}$  des P450 à partir des énergies de solvation, de translation et de rotation:

$$\Delta G_{bind} = \Delta G_{part} + \Delta G_{hb} + \Delta G_{\pi-\pi} + \Delta G_{rot} + \text{constante} \quad (1)$$

$\Delta G_{part}$  se rapporte au terme de désolvatation provenant du paramètre de lipophilie  $\log(P)$ ;  $\Delta G_{hb}$  est le terme correspondant à la liaison hydrogène,  $\Delta G_{\pi-\pi}$  est l'énergie d'interaction  $\pi-\pi$ , et  $\Delta G_{rot}$  exprime la perte d'énergie de rotation individuelle de liaison lorsqu'une petite molécule de substrat se lie au site actif du P450 [Lewis 2003b].

Selon l'équation (1),  $\Delta G_{bind} = a\Delta G_{part} + b$ ; où  $a$  est l'hydrophobie du site actif P450 et  $b$  représente les contributions combinées des liaisons hydrogène, des interactions  $\pi-\pi$  et d'autres composants de l'énergie de liaison [Lewis 2003b]. En utilisant l'équation (1), il est possible de calculer l'énergie de liaison globale en fonction du nombre d'interactions du site actif, calculés par des méthodes de modélisation moléculaire [Lewis 2003b]. Il est à noter que jusque-là, les interactions du substrat au P450 sont prises en compte au niveau du site actif seulement, alors que notre hypothèse sur les canaux suggère d'étudier les interactions substrat-enzyme durant toute la traversée depuis la surface de l'enzyme, et même depuis la membrane plasmique, ce qui paraît plus réaliste car l'accessibilité du substrat au site actif est un critère de sélectivité très important! Les affinités de liaison des substrats du CYP3A4 ont été calculées avec un bon accord avec des valeurs expérimentales [Lewis 2003b]. Le terme de désolvatation est calculé directement avec le  $\log(P)$  expérimental. Une corrélation avec les résultats expérimentaux a montré que la perte d'énergie de translation et de rotation du substrat après liaison au P450 est négligeable, et probablement due à un certain degré d'énergie de translation et de rotation évidente dans le P450 lui-même, car il est incorporé dans une double couche phospholipidique relativement fluide constituant la membrane réticulaire endoplasmique.

Plusieurs études pointent la pertinence du caractère lipophile du substrat, de l'inhibition et de la clairance métabolique [Al-Gailany *et al.* 1978, Hansch & Zhang 1993, Gao & Hansch 1996, Ramos-Nino *et al.* 1996, Ishigami *et al.* 2001, Riley *et al.* 2001, Austin *et al.* 2002]. Par exemple, pour une série de statines agissant comme inhibiteurs compétitifs du CYP3A4, il y a une bonne corrélation ( $R=0.96$ ) entre le  $\log(K_i)$  (où  $K_i$  est la constante d'inhibition) et le  $\log(D_{7.0})$ , qui mesure la lipophilie du composé comme étant la valeur du  $\log(P)$  ionisé corrigé au pH 7.0 [Ishigami *et al.* 2001]. Une relation semblable est apparue pour une série structurellement plus diverse d'inhibiteurs et de substrats du CYP3A4 [Riley *et al.* 2001]. Des études sur le CYP1A2 [Lewis *et al.* 1995] ont démontré que la longueur de la chaîne du substituant alkoxy augmente l'hydrophobie et affecte l'affinité pour l'enzyme 1A2 résultant de la gêne stérique dans le site de liaison [Lewis *et al.* 1995]. Des substrats du CYP2D6 offrent aussi une corrélation nettement linéaire entre la liaison à l'enzyme et la valeur de  $\log(D_{7.4})$  [Ferrari *et al.* 1991]. En outre, dans une étude portant sur 7-alkoxycoumarines [Vermeir *et al.* 1992a], le logarithme de la O-désalkylation catalysée par le CYP2B4 se révèle être directement proportionnel au  $\log(P)$ . Il apparaîtrait que des substrats basiques ont tendance à se lier aux CYP2B dans l'état ionisé [Vermeir *et al.* 1992b], par opposition à la situation rencontrée pour le CYP2D6. Suite à deux études [Hansch & Zhang 1993, Gao & Hansch 1996], il a été examiné les relations possibles entre lipophilie, au niveau des valeurs du  $\log(P)$  ou du  $\log(D_{7.4})$ , et liaison aux P450s humains [Lewis & Dickins 2002, Lewis & Dickins 2003]. Ces études ont montré que l'affinité de liaison au substrat, déduite du  $K_m$  ou du ( $K_D$ ), offre une relation linéaire simple entre  $-\log(K_m)$  (affinité) et  $\log(P)$  [Lewis *et al.* 2004].

CYP	% de médicaments ( <sup>a</sup> )	Gamme de valeurs de log(P)	Valeur moyenne log(P) [n]	Autres caractéristiques	Interactions déduites des relations de lipophilie des substrats [n] <sup>(b)</sup>
1A1	3	1.39 à 6.35	3.41 [16]	HAPs <sup>(c)</sup> et leurs diols	1 liaison hydrogène et 2 $\pi$ - $\pi$ [5]
1A2	10	0.08 à 3.61	2.01 [18]	Amines et amides	2 liaisons hydrogènes et 1 $\pi$ - $\pi$ [10]
1B1	1	1.40 à 6.35	3.73 [12]	HAP et leurs diols	2 liaisons hydrogènes et 2 $\pi$ - $\pi$ [7]
2A6	3	0.07 à 2.79	1.44 [18]	Molécules plutôt petites	1 liaison hydrogène et 1 $\pi$ - $\pi$ [5]
2B6	4	0.23 à 4.89	2.54 [16]	Basique, non ionisé	1 liaison hydrogène et 1 $\pi$ - $\pi$ [16]
2C8	?	0.06 à 6.98	3.38 [12]	Acide, ionisé	1 liaison hydrogène [5]
2C9	25	0.89 à 5.18	3.20 [18]	Acide, non ionisé	1 liaison hydrogène et 1 $\pi$ - $\pi$ [9]
2C19	?	1.49 à 4.42	2.56 [16]	Amines et amides	1 liaison hydrogène [6]
2D6	15	0.75 à 5.04	3.08 [16]	Basique, ionisé	1 paire d'ions [6]
2E1	3	-1.35 à 3.63	2.07 [20]	Petites molécules	1 liaison hydrogène ou 2 $\pi$ - $\pi$ [7]
3A4	36	0.97 à 7.54	3.10 [50]	Grandes molécules	Liaisons hydrogènes et/ou $\pi$ - $\pi$ [ $>40$ ]

**Table 1.1** : Intervalles et moyennes des valeurs de log(P) pour des substrats P450 humains des familles CYP1, CYP2 et CYP3, d'après [Lewis *et al.* 2004]. [n] : Les valeurs entre crochets sont les nombres de composés étudiés. (a) : Pourcentages d'oxydations totales de médicaments catalysées par un P450 donné pour 2112 composés [Rendic 2002]. (b) : D'après [Lewis & Dickins 2003]. (c) HAP : Hydrocarbure Aromatique Polycondensé.

En général, plus le log(P) d'un composé est grand, plus il se liera à un P450 donné, bien qu'au-delà d'un certain point, le composé deviendra trop hydrophobe et l'affinité de liaison (et probablement la clairance) commencera à diminuer avec l'augmentation du log(P). Il va probablement y avoir une valeur optimale du log(P) pour les substrats d'un P450 donné, avec en majorité des composés sélectifs se situant dans une certaine gamme de valeurs du log(P) (voir tableau 1.1). La nature lipophile précise du substrat déterminera donc, au moins dans une certaine mesure, l'isoforme P450 la plus probable le métabolisant, et une affinité de liaison approximative qui peut être évaluée à partir du log(P) du substrat. Bien qu'il soit important de souligner que d'autres facteurs pourraient aussi être impliqués, comme la polarité, le caractère donneur/accepteur des liaisons hydrogènes et le nombre de cycles aromatiques dans la molécule, de telles considérations pourraient faciliter le processus de sélection conduisant à la découverte de médicaments.

## 1.8.2 Les facteurs contributifs dans les relations de lipophilie

En plus du rôle de la lipophilie discuté auparavant, il y a plusieurs facteurs contributifs à l'énergie totale de liaison enzyme-substrat, induisant des relations statistiques avec la lipophilie. Un des aspects utiles des relations de lipophilie est l'indication, par l'observation de davantage de substrats P450, de la variation du nombre de liaisons hydrogène en interactions  $\pi$ - $\pi$  (voir table 1.1) pour un P450 donné, comme c'est le cas des substrats du CYP3A4, qui présentent des combinaisons différentes de liaisons hydrogène et d'interactions  $\pi$ - $\pi$ , contrairement aux autres isoformes qui présentent une combinaison bien définie de liaisons [Lewis & Dickins 2003]. En fait, la relation statistique établie sur la lipophilie fournit directement une mesure de la moyenne de la non désolvatation, qui elle-même est une combinaison linéaire simple des énergies des liaisons hydrogène et des interactions  $\pi$ - $\pi$  (rapportées à -2 kcal/mol et -0.9 kcal/mol, respectivement). En outre, les intervalles et les moyennes des valeurs du log(P) des substrats P450 indiqueraient directement leur sélectivité enzymatique (table 1.1).

Quelques points peuvent être établis :

- L'affinité de liaison est déterminée par le  $\log(P)$ , le nombre de donneurs et d'accepteurs de liaisons hydrogène, les interactions  $\pi$ - $\pi$  (N) et le nombre de liaisons en libre rotation [Lewis 2003b].
- Le site de métabolisme du substrat est déterminé par les donneurs accepteurs de liaison hydrogène dans la molécule [Lewis 2003b].
- Les valeurs du  $\log(P)$  de substrats typiques sont proches de la valeur moyenne pour le P450s en question (table 1.1).
- Des relations de lipophilie avec  $\Delta G_{bind}$  existent pour des substrats de tous les P450s examinés jusqu'à présent [Lewis & Dickins 2003]. Il est habituel qu'il y ait plus d'une relation linéaire de lipophilie pour chaque isoenzyme.

La réunion de ces informations peut s'avérer utile pour déterminer ce qui caractérise structurellement la liaison au P450 et la sélectivité; en plus de lipophilie, le rôle de la liaison hydrogène devrait être souligné [Abraham *et al.* 2002], car c'est une composante clé des interactions enzyme-substrat [Lewis 2004, Lewis & Broughton 2002]. La modélisation moléculaire des complexes P450-substrat complète l'analyse de lipophilie, bien que la dynamique de liaison au substrat puisse suggérer un processus de désolvatation gouverné par l'entropie, qui pourrait être lié à la lipophilie de composé [Winn *et al.* 2002], même si la modélisation moléculaire devrait concerner l'ensemble de l'interaction des deux partenaires du complexe, et non se restreindre au site actif. De plus, pour les P450s, la modélisation par pharmacophore et l'analyse QSAR [Elkins *et al.* 2001, de Groot & Elkins 2002] ont tendance à venir en support de la modélisation moléculaires [Smith *et al.* 1997b, Smith & van de Waterbeemd, 1999], où la similarité structurale des substrats est une autre caractéristique utile [Clark & Pickett 2000]. Donc, bien qu'importante, la lipophilie d'un composé est un facteur qui devrait être conforté par d'autres informations afin d'obtenir un profil complet de la sélectivité des P450s et du métabolisme d'un composé chimique.

### 1.8.3 Lipophilie et sélectivité du P450

La lipophilie du substrat potentiel est cruciale pour la métabolisation. En général, chez les mammifères, les substrats des P450s ont tendance à avoir des valeurs de  $\log(P)$  positives (peu ont un  $\log(P)$  négatif). Les raisons probables de cela sont au niveau du transport membranaire et de l'attachement du substrat au site actif du P450, où la désolvatation de l'environnement de la poche héminique hydrophobe représente en général la contribution majeure à l'affinité de liaison.

Les P450 humains affichent une certaine sélectivité du substrat, reflétée en partie par les intervalles et les valeurs moyennes du  $\log(P)$  du substrat. En effet, on peut expliquer que quelques composés pourraient agir comme substrats de plus d'un P450, en partie par leurs valeurs du  $\log(P)$  tombant dans les intervalles rencontrés pour ces enzymes. Les intervalles du  $\log(P)$  pour des P450s individuels peuvent être rationalisés en termes de variation d'hydrophobie et de constante diélectrique locale de chaque site actif P450, et la pente dans la relation de lipophilie pour les substrats d'un P450 donné peut fournir une mesure du degré d'hydrophobie au voisinage du groupement hème.

La valeur du  $\log(P)$  d'un substrat peut informer sur sa sélectivité, bien qu'un substrat soit parfois métabolisé par plus d'une enzyme, et fournit un moyen d'évaluer son affinité de liaison lorsque la relation lipophilie est connue, particulièrement pour des

séries homologues, et aussi à condition que les interactions avec les sites actifs potentiels soient bien caractérisées par modélisation moléculaire.

Le  $\log(P)$  d'un composé est un paramètre crucial pour évaluer de nouvelles entités chimiques; pour des substrats ionisables, le  $\log(D_{7.4})$  peut être plus pertinent, ce qui signifie qu'il est essentiel de considérer les valeurs des pKa. Au niveau d'un composé, les discriminants potentiels importants de la sélectivité incluent [Lewis *et al.* 2004] :

- La lipophilie:  $\log(P)$  ou  $\log(D_{7.4})$ .
- La taille (diamètre, longueur) de la molécule et la masse moléculaire.
- La planarité de la molécule.
- Le pKa (basicité/acidité).
- Le moment de dipolaire.

## 1.9 Ingénierie sur les substrats des isoenzymes pour le contrôle du métabolisme

### 1.9.1 Une conception rationnelle pour une approche *Early-ADME*

Les paramètres ADME (Absorption, Distribution, Métabolisme, Elimination) sont incontournables pour le contrôle de la biodisponibilité d'un médicament. Ces phénomènes sont liés entre eux : le métabolisme rend généralement les molécules filles plus solubles pour faciliter leur élimination et peut produire de nouvelles molécules qui possèdent de nouvelles propriétés et activités. Ils sont aussi très versatiles, dépendant de la variabilité génomique et des pathologies, dues à l'âge, au sexe, etc. Le contrôle de ces paramètres est complexe.

Pour concevoir un médicament, il ne suffit pas d'optimiser l'affinité, qui se traduit par l'activité ou de la puissance (celle-ci se réfère à la quantité de composé requis pour produire un effet biologique donné), il faut aussi contrôler la sélectivité (celle-ci a pour conséquence de minimiser ou d'éliminer les effets secondaires ou indésirables), maximiser l'efficacité (niveau maximal d'effet biologique qu'un composé peut produire), contrôler la durée d'action, etc. Ceci implique la prise en compte des paramètres physico-chimiques responsables du comportement et du devenir des xénobiotiques dans l'organisme. Dès la conception de nouvelles entités chimiques, il faut pouvoir contrôler la biodisponibilité car la meilleure activité *in vitro* ne pourra s'exprimer *in vivo* que si la molécule est disponible en quantité suffisante au niveau du site actif. Pour une administration per os, l'obtention d'une biodisponibilité orale suffisante implique de franchir de plusieurs étapes, dont une bonne absorption intestinale, le transit gastro-intestinal, la stabilité chimique dans le tractus et l'effet du premier passage hépatique dû au métabolisme (foie et paroi intestinale) [Testa 1995].

Beaucoup de candidats médicamenteux échouent dans des phases de développement coûteuses pour des raisons de comportement ADME inadéquat (taux d'attrition : clinique > 90%, préclinique  $\pm$  50%). Par exemple, 50% des échecs observés sur 319 nouvelles entités chimiques de 7 grandes entreprises pharmaceutiques, sur une période de 21 ans, sont dus à des problèmes de type ADME/Tox rencontrés lors des essais cliniques ! Seulement 10% des candidats proposés aux phases de développement vont jusqu'au marché (2% à partir de la recherche préclinique) [Lipper 1999].

Les causes sont :

- pK inadaptés : 41%.
- Efficacité insuffisante : 31%.
- Toxicité : 22%.
- Inadéquation avec le marché : 6%.

Jusqu'ici, la robotisation à haut débit (expérimentale) a été utilisée pour l'optimisation des activités recherchées, et plus particulièrement pour maximiser la puissance et la sélectivité. Mais cette méthode basée sur le hasard ne peut être efficace, surtout à notre époque où l'espace des entités chimiques est de plus en plus gigantesque. Une des stratégies des technologies à haut débit est d'augmenter la puissance par l'introduction des groupements hydrophobes. Même si cette approche est très avantageuse au niveau du métabolisme, c'est au détriment de l'absorption car elle diminue la solubilité : seule une molécule en solution et non ionisée peut passer une barrière (sauf s'il existe un transport actif) et augmenter la masse. Le but de cette stratégie est d'augmenter le volume de distribution donc le temps d'action, mais cela devient néfaste lorsqu'il s'agit de composés ayant des effets secondaires importants.

L'espace possible des "petites" molécules organiques est évalué à  $10^{62}$ . Or, le nombre de cibles connues est environ de 500 sur  $10^{26}$  estimées, ce qui veut dire que la probabilité de trouver par hasard une molécule active parmi  $10^{40}$  sur une des  $10^{26}$  cibles est de  $10^{14}$ . Cette analyse montre que si on veut trouver un ligand au niveau nanomolaire dans une chimiothèque, il faut que celle-ci ait au moins 24 millions de molécules [Lipinski 2000, Hann & Oprea 2004]. Le hasard doit alors laisser sa place à la rationalité.

Afin de surmonter les facteurs ADME/Tox, on a recours à des stratégies de manipulation dans la synthèse des molécules. Le but est de comprendre le plus tôt possible, afin anticiper et de prévoir les meilleurs candidats dans une phase de recherche, quand les facteurs financiers ne sont pas encore pesants et quand les pharmaco-modulations laissent encore beaucoup de possibilités [Sinko 1999]. C'est le concept de *fail fast, fail cheap* (un échec rapide est bon marché), qui a été nommé *Early-ADME*.

La robotisation de l'ADME est en plein développement, mais les très hauts débits sont impossibles car il s'agit surtout de tests *in vivo* et les paramètres à mesurer sont complexes [Caldwell 2000]. Ils nécessitent une grande fiabilité pour être utilisés efficacement sous peine de ne servir que de filtres, avec beaucoup de risques d'obtenir des faux positifs et des faux négatifs. D'où l'intérêt du *drug design*, ou conception rationnelle de molécules : il faut donc "piper les dés" pour donner un maximum de chance aux molécules synthétisées de devenir des médicaments. De nouveaux concepts ont donc été développés, comme la *druggability* [Wermuth 2004] (paramètres à éviter ou à maintenir), le *drug-like* (plus le nouveau produit ressemble à un médicament qui marche, plus on aura de chance que le nouveau marche, et inversement), et les criblages virtuels ou *in silico* (multiples filtres de sélection) [Nikolova & Jaworska 2003]. En tenant compte des contraintes physicochimiques de l'ADME, l'espace actuel des *drug-like* ne dépasse pas 10.000 molécules.

L'objectif est le développement de modèles prédictifs. Il est illusoire de penser établir une règle générale de prédiction car il s'agit de mécanismes trop complexes agissant sur trop de paramètres, qui touchent à la fois les aspects structuraux et les aspects physico-chimiques, et qui ne forment un ensemble cohérent qu'au sein d'une famille chimique homogène. Et tout cela dans des organismes qui présentent souvent de fortes variations physiologiques dues à différentes causes (pharmaco-génomique, pathologique,

vieillesse, sexe, etc.). On espère seulement prédire des classes, sans être plus précis. Pourtant le défi est à la hauteur des espoirs attendus : prédire le comportement (bon, moyen, mauvais) des nouvelles entités chimiques à partir des propriétés physico-chimiques et structurales calculées avant de les synthétiser (chimiothèques focalisées et criblage virtuel), se chiffre en temps et en argent. Comprendre les facteurs responsables, c'est pouvoir les utiliser en retour pour la conception rationnelle de nouvelles entités chimiques. C'est non seulement gagner du temps, mais aussi réduire les coûts de recherche et développement. Le concept de plate-forme *early-ADME* est de plus en plus mis en parallèle avec la chimie combinatoire et le criblage à haut débit, et non plus considéré comme la suite de l'étape d'optimisation de la puissance.

### 1.9.2 Stratégies de contrôle du métabolisme par les CYPs

Le développement des modèles prédictifs du métabolisme a recours à deux stratégies, selon le but visé : soit de contrôler le taux de métabolisme et d'inhibition (connaître l'affinité et la sélectivité), soit éviter le métabolisme par les CYPs (connaître la spécificité). Les propriétés physico-chimiques impliquées sont :

- Ionisation (pKa), dépendant du pH : souvent les bases sont plus métabolisables que les acides.
- Taille et forme de la molécule.
- Lipophilie (Log(P), Log(D)), solubilité.
- Polarité, polarisabilité.
- $\Delta\log(P)$  (entre octanol-eau et cyclohexane-eau), représentatif du potentiel de liaison hydrogène.
- Capacité de formation de liaisons hydrogène.
- Surface accessible au solvant.
- Aspects conformationnels.

Un seul paramètre ne suffit pas pour comprendre : il faut une approche multivariée. Des difficultés peuvent être rencontrées :

- Choix des molécules, des familles représentatives, des tests biologiques, de la dose (la plus proche possible de la dose thérapeutique?), coûts, confiance dans les résultats obtenus, conditions expérimentales, variabilité entre les différents laboratoires.
- Etude de problèmes spécifiques : exemple, lorsque la biodisponibilité dépend de la dose (pharmacocinétique non linéaire), pro-médicaments, métabolites actifs.
- Interactions médicamenteuses compliquant l'analyse et la compréhension (exemple, CYP induit ou inhibé).
- Représentativité des modèles animaux, *in vitro* vs. *in vivo*.

Nous présentons quelques exemples de traitements de molécules pour le contrôle du métabolisme :

- Introduire des fonctions "métabolisables" : alcool (conjuguable et modérément oxydable), hydroxylation aromatique, N-désalkylation, oxydation des thiols, réduction des cétones et des nitro aromatiques.
- Introduire certains groupes tels que phénol (conjuguable) ou sulfonamide (N-acétylable et conjuguable), diminuant la biodisponibilité en réduisant la vitesse d'absorption intestinale.

Le CYP3A4 étant spécifique des molécules très lipophiles, on évite les interactions avec le 3A4 par réduction des interactions hydrophobes en diminuant la lipophilie des substrats qui diminuent le métabolisme par le 3A4. Ainsi on diminue la lipophilie en incorporant une fonction plus polaire (ex. aryle ou aniline) loin du site de métabolisme initial, cette nouvelle fonction polaire ajoutée devenant le nouveau site de métabolisme. Le 3A4 peut accueillir des ligands dans de multiples orientations, et le métabolisme intervient souvent sur les sites labiles extrayant hydrogènes et électrons. Une autre façon de éviter le métabolisme du 3A4 est de supprimer ou d'altérer des fonctions afin de diminuer la labilité de ces sites d'extractions. Par exemple, pour atténuer la N-désalkylation, on peut réduire la fonctionnalité autour d'une amine tertiaire en le remplaçant par une amine secondaire ou primaire. Le 3A4 étant très sensible aux interactions hydrophobes (lipophilie), il est judicieux d'accélérer le passage pour diminuer le temps de contact si on veut éviter la métabolisation.

Le métabolisme par une isoenzyme polymorphe comme le CYP2D6 peut être problématique pour développer un médicament en raison de la large variation de pharmacocinétique au sein de la population des patients. Les molécules de type aryl-alkylamine sont des substrats probables du CYP2D6. La régiosélectivité du CYP2D6 envers son substrat est liée au fait que la liaison au substrat est induite par l'interaction entre la paire d'ions de l'azote protoné du groupement amine de l'acide aminé dans le site actif du 2D6, et par l'interaction l'hydrophobe ou  $\pi$ - $\pi$  de la fonction aryle. Les substrats doivent avoir une chaîne alkyle de dimensions suffisante pour permettre ces interactions clés. Des stratégies de synthèse pour contrôler la métabolisation par le CYP2D6 incluent la perturbation de l'interaction hydrophobe ou  $\pi$ - $\pi$ , en ajoutant une fonctionnalité au groupe aryle (altérant ainsi la séparation aryl-amine en changeant la longueur ou la rigidité de la chaîne alkyle), ou en supprimant l'interaction avec la paire d'ion en atténuant la basicité de l'azote (du 2D6). Le Betoxolol fournit un exemple d'altération du métabolisme par ajout d'une fonctionnalité au groupe aryle : ici, la position en para du groupement aryle d'un antagoniste bêta-adrénocepteur a été substituée par une fonction volumineuse métaboliquement stable pour obtenir une pharmacocinétique très améliorée et une absence de métabolisme polymorphe [Smith & Jones 1992]. Ces changements résultent d'affinités nettement inférieures, mesurés par la capacité d'inhiber le métabolisme d'autres substrats par l'enzyme.

Une autre interaction clé de la liaison avec les P450 apparaît avec les composés contenant des hétérocycles azotés saturés comme l'imidazole et la pyridine. Ces groupes fonctionnels sont ajoutés pour faciliter la solubilité, mais peuvent avoir une forte interaction avec l'hème via le doublet de l'azote. Les calculs suggèrent que la présence d'une telle fonctionnalité ajoute 6 kcal/mol d'énergie de liaison aux interactions existantes. Une telle affinité explique la grande puissance de certains inhibiteurs des P450s, comme le kétoconazole et le sulfaphénazole, respectivement pour le 3A4 et le 2C9. En outre, la liaison est si forte que les composés contenant ces fonctions peuvent devenir des inhibiteurs généraux des P450s au lieu d'être spécifiques à une ou deux isoenzymes. Là où ce type de liaison au P450 n'est pas essentiel pour l'activité, comme dans le cas des inhibiteurs des aromatasés ou de la C-14 deméthylase, une addition stérique autour de l'azote ou la substitution de la fonction avec un bioisostère peut résoudre le problème. Par exemple la cimétidine, qui possède un groupement imidazole, inhibe le métabolisme de divers médicaments, tels que phénytoïne, cyclosporine et propranolol, habituellement attribués

respectivement au CYP2C9, CYP3A4 et CYP2D6. A l'inverse, la ranitidine, dans lequel le cycle imidazole est remplacé par un furanne, offre peu de potentiel d'interaction.

La connaissance des canaux, de par leur localisation et leur structure chimique, permet ce contrôle rationnel de la sélectivité. Comprendre les relations structure/fonction des P450s, permettre de les modifier afin de métaboliser des substrats importants ou de produire des composés chimiques d'intérêt. Connaître la structure des isoformes et les relations structure-activité avec les ligands, depuis la bouche d'entrée au site actif en passant par les canaux, permet une conception plus rationnelle des médicaments pour un meilleur contrôle du métabolisme et des interactions médicamenteuses. C'est le défi lancé dans cette thèse d'avancer dans la conception rationnelle de molécules via les canaux.

## 1.10 Isoformes P450 humains : interactions spécifiques avec les substrats. Sélectivité

Le CYP est une grande famille de protéines similaires nommées isoenzymes ou isoformes possédant le même complexe porphyrine-hème au centre actif du côté distal, plutôt variable en séquences d'acides aminés (de 20% à 90% de similarité de séquence entre CYPs), mais de repliement globulaire identique. Les structures secondaires et tertiaires ont été largement conservées à travers l'évolution. L'analyse comparative des structures des P450s bactériens (procaryotes) [Hasemann *et al.* 1995] suggère que tous les P450s bactériens possèdent la même structure tertiaire, et que les P450BM-3 et CYP102 bactériens sont proches structuralement des P450s eucaryotes avec une certaine similarité de séquences. Les différences de séquences d'acides aminés ont un impact sur la topographie du site actif et des canaux d'accès au site actif. Cette propriété les différencie dans la sélectivité de substrats, et dans la catalyse d'une grande variété de réactions de monooxygénation par l'activation d'un oxygène moléculaire.

Pour la recherche, tant appliquée que fondamentale, comprendre les interactions des isoformes P450 humains avec les substrats, en particulier ceux présents dans le foie, est d'intérêt majeur. Les humains ont 57 gènes CYP et 33 pseudogènes, classés dans 18 familles et 42 sous-familles [Guengerich & MacDonald 1990], qui sont principalement exprimés dans le foie et le tube gastro-intestinal. Plusieurs CYPs humains ont été reconnus pour leur implication dans le métabolisme pharmaceutique humain : 1A2, 2D6, 2C9, 2C19, 2E1 et 3A4 [Smith & Jones 1992], suivis plus récemment par d'autres, tels que 3A5, 2A6, 2B6 et 2C8. La figure 1.6 illustre la distribution des xénobiotiques métabolisés par les isoformes humaines. On peut voir que les isoformes 3A et 1A jouent un rôle central dans le métabolisme humain, puisque, ensemble ils sont impliqués dans le métabolisme de 50% des xénobiotiques commercialisés, dont des composés non médicamenteux. Le CYP3A4 est un des membres les plus abondants de cette famille, qui métabolise près de 50% des médicaments. Au niveau des microsomes membranaires, les isoformes 3A4, 2D6, 2C9, 2C19, 2E1 et 1A2, réunies, sont responsables du métabolisme oxydatif de plus de 90% des médicaments commercialisés.

Dans la sous-famille 3A, un fait nous interpelle : bien que les isoformes 3A4 et 3A7 (isoforme fœtal) aient environ 80% d'identité de séquence, le 3A7 a très peu de substrats connus, essentiellement dérivant de stéroïdes, alors que le 3A4 peut métaboliser des milliers de composés, de poids moléculaires 200 à 1400 Da. En utilisant un modèle par homologie du 3A7, on peut se rendre compte que leur imposant site actif ne diffère

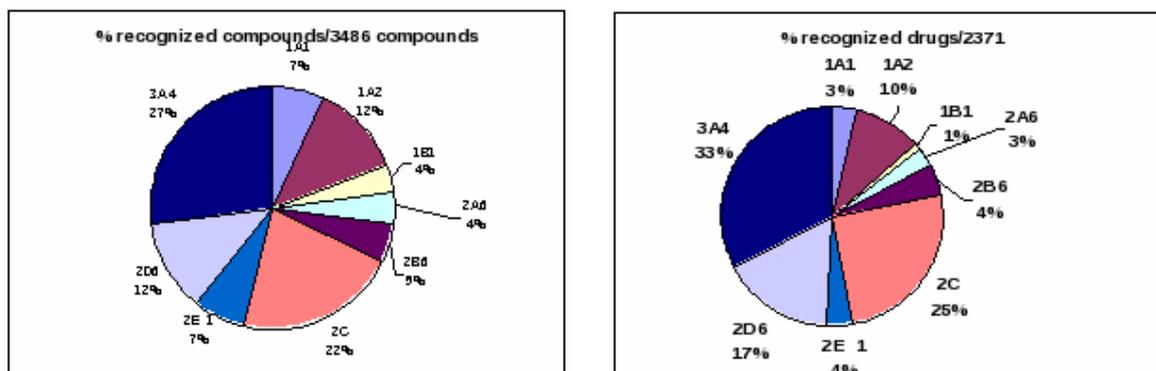


Figure 1.6 : Pourcentages de composés et de médicaments métabolisés [Rendic 2002].

que de 2 acides aminés situés en positions non cruciales pour l'orientation du substrat au niveau de l'hème. Les centaines de positions mutées sont localisées ailleurs dans la structure, dans des boucles et sur la surface et dans les canaux. Ceci révèle l'importance de l'environnement lors du trajet du substrat dans le CYP pour atteindre le site actif, et l'importance de la sélectivité, cette dernière n'opérant pas uniquement dans la poche catalytique contenant le groupement hème.

De nombreuses isoformes ont un site actif hydrophobe avec au moins un acide aminé aromatique proche de l'hème réactif et capable de présenter des interactions  $\pi$ - $\pi$  avec les substrats. La nomenclature des isoformes est basée sur l'homologie de séquence d'acides aminés, de telle sorte que tous les membres d'une famille ou d'une sous-famille aient au moins 40% d'identité de séquence. En fonction de la différence de séquence d'acides aminés, la topographie du site actif diffère selon les isoformes, ce qui induit une réactivité différente.

### 1.10.1 Diversité de formes et de fonctions. Modes d'interactions isoformes-substrats

La spécificité chez les CYPs est une question d'affinité structurale et physicochimique entre la protéine et la petite molécule, et cela dès le début de l'interaction, depuis le passage par la membrane suivi de l'accroche à la surface de la protéine, jusqu'à l'arrivée du ligand au centre réactionnel, en passant par les canaux. Les propriétés physicochimiques des ligands jouent un rôle clé dans l'interaction avec une isoforme déterminée. Comprendre la structure et la physicochimie de chacun des deux partenaires, les isoformes et les substrats/inhibiteurs, peut révéler la part que jouent les fonctions chimiques, la désolvatation, la lipophilie, etc., dans la reconnaissance, l'accessibilité et la liaison effective au site actif de ces composés.

Le mécanisme catalytique du CYP est identique pour toutes les isoformes. Le mécanisme chimique est régi par la capacité oxydante du complexe Fer-oxo formé au cours du cycle catalytique [Guengerich & MacDonald 1990]. Le métabolisme de chaque composé par une isoenzyme est déterminé par 4 facteurs :

- La topographie du site actif (dépendant du CYP).
- La topologie des canaux d'accès au site actif, bouche et corps des canaux (dépendante du CYP).

- La facilité à perdre des hydrogènes ou des électrons des carbones ou des hétéroatomes des substrats (dépendant du substrat).
- L'encombrement stérique et l'affinité physico-chimique, rendant compte de l'accessibilité d'un ligand au site actif, puis au site de métabolisation au niveau de l'entité réactive feroxo (dépendant du couple CYP-substrat).

Les substrats des CYPs appartiennent à des classes chimiques diverses, et certaines caractéristiques moléculaires (forme tridimensionnelle, dimensions, lipophilie), et des sous-structures (groupes fonctionnels, caractéristiques stériques et stéréochimiques), peuvent induire sa capacité à se lier à chaque isoenzyme, permettant de classer les substrats selon ces caractéristiques. Celles-ci sont à l'origine du concept de pharmacophore, qui a été défini comme "*un ensemble de caractéristiques structurales dans une molécule qui est reconnue par un site récepteur et est responsable de l'activité biologique de la molécule*" [Gund 1977]. Il a été montré que les pharmacophores sont très utiles pour prévoir les sites d'attaque et les réactions catalytiques effectuées sur eux, et ils sont souvent utilisés en analyse QSAR [Hansch *et al.* 2004].

Pour une molécule, les principes généraux de la chimie organique peuvent nous renseigner intuitivement sur les positions réactives au niveau du site de métabolisation par un CYP, et sur le type de transformations effectuées sur les groupements fonctionnels. Les sites probables d'attaque sont, soit des fragments riches en électrons (ex. oléfines, aromatiques, certains hétéroatomes), soit des positions allyliques ou benzyliques (où le radical peut être stabilisé par délocalisation) : voir Table 2 dans [Brown *et al.* 2008].

Les informations sur les structures des canaux sont rares dans la littérature, et nous proposons de les enrichir dans ce travail de thèse. Les premières études sur les CYPs ont été concentrées sur le site actif. En général les isoformes des CYPs possèdent des acides aminés aromatiques au niveau du site actif, capables d'interactions  $\pi$ - $\pi$  avec les ligands. D'autres acides aminés plus distants du site actifs participent aux interactions avec les ligands, telles que les interactions hydrophobes (3A4), les liaisons hydrogènes (1A2 et 2C9), ou la formation de paires d'ions (2D6, 2C9) [Brown *et al.* 2008]. Les réactions au niveau du site de métabolisation contrôlent la réactivité mais pas uniquement, car, au niveau des canaux il peut intervenir d'autres facteurs, agissant sur l'orientation des molécules avant leur arrivée voisinage de l'hème.

En général les substrats des P450 sont lipophiles ( $\log(D)>0$ ). Il a été suggéré que les interactions lipophiles sont à longue portée [Mancy *et al.* 1995], et donc peuvent être responsables de l'attraction des ligands depuis la membrane vers les P450. Une fois la molécule attirée vers le P450, ce sont ses propres propriétés structurales (géométriques) et physico-chimiques qui déterminent la sélectivité, à savoir quelle isoforme la métabolisera, et cela dès l'accroche à l'enzyme. Sauf pour le 2C9 et le 2D6, il n'y pas de préférence entre les substrats acides ou basiques pour les CYPs [Smith *et al.* 1997b].

Les règles générales de sélectivité basées sur des corrélations avec la structure stéréo-électronique des substrats, permettent de classer les ligands, puis de prévoir ceux qui sont potentiellement métabolisables par une isoforme déterminée. De plus il est possible de concevoir des molécules qui ne seront pas métabolisées par les CYPs.

Les interactions entre les acides aminés des CYPs et les ligands sont :

- Liaisons hydrogènes O–O et O–N [Lewis *et al.* 2004], mais on sait que des liaisons hydrogènes faibles peuvent aussi se former entre cycles aromatiques et donneurs de liaison hydrogène [Lewis *et al.* 2004].
- Ponts salins.

- Interactions électrostatiques.
- Interactions hydrophobes : cycles aromatiques plans, intercalations  $\pi$ - $\pi$  avec les cycles aromatiques des chaînes latérales.
- Interactions cation- $\pi$ s, interactions de paires d'ions.

Ces interactions peuvent régir l'orientation du ligand lors de son admission à l'entrée du canal, puis au sein du canal, et enfin dans le site actif. Cependant, il est important de se rappeler les valeurs suivantes :

- Pour les systèmes biologiques, la gamme d'énergies de liaisons hydrogène est comprise entre -1.5 et -2.5 kcal/mol, avec une valeur moyenne représentative de -2.0 kcal/mol.
- Les valeurs des énergies d'interaction  $\pi$ - $\pi$  vont de -0.8 to -1.0 kcal/mol, avec une valeur moyenne représentative -0.9 kcal/mol.

Dans une protéine, l'espace accessible est divisé en 3 types de régions, suivant l'affinité de liaison pour certains groupements de la molécule :

1. Régions hydrophobes : favorables par l'occupation de groupement hydrophobes de ligand.
2. Régions hydrophile : favorables par l'occupation de groupement hydrophiles de ligand.
3. Ni hydrophobe ni hydrophile : critère mixte.

Dans notre étude sur la sélectivité à travers les canaux, nous avons choisi d'étudier les isoformes 3A4 et 1A2, car ce sont les plus contrastées au niveau de leurs propriétés, de leurs ligands, et de leur sélectivité. En effet, le 3A4 présente une grosse poche, il est peu sélectif et métabolise un large spectre de molécules, contrairement au 1A2 qui possède une petite poche et s'associe à des molécules bien spécifiques et aplaties. Nous présenterons dans ce qui suit la structure de chacune de ces deux isoenzymes, ainsi que les particularités de leurs substrats et inhibiteurs intervenant dans des corrélations structure-activité.

## 1.10.2 Structure de l'isoenzyme CYP3A4 et de ses ligands

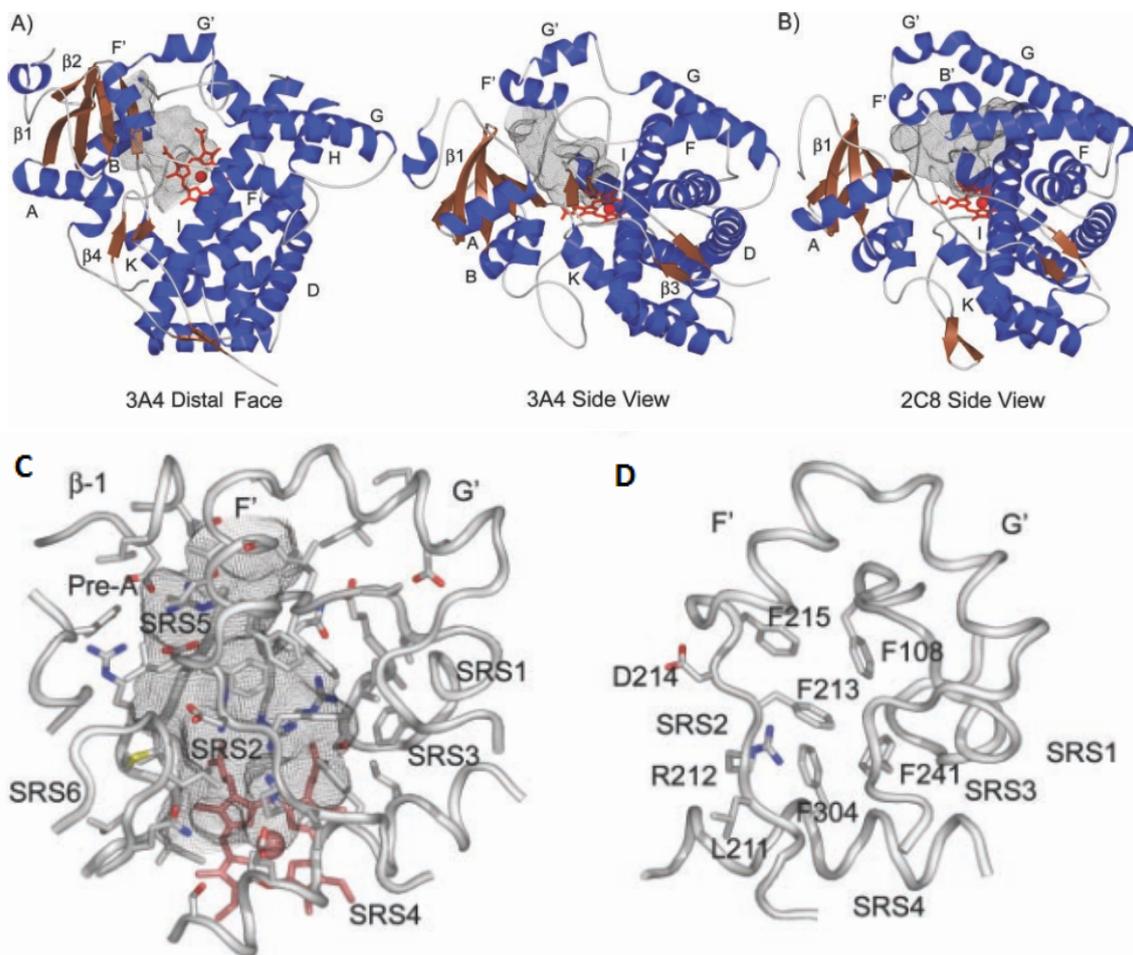
Parmi les 57 CYPs humains, l'isoforme 3A4 est la plus abondante et la plus importante parce qu'elle métabolise la majorité des médicaments administrés. Une caractéristique remarquable du 3A4 est sa multi-spécificité de substrats, et ses liaisons coopératives des substrats, qui accélèrent le rendement du métabolisme, comme elles peuvent mener à des interactions médicamenteuses indésirables et des effets secondaires toxiques. Du fait de son importance dans le développement de médicaments et de thérapies, le 3A4 a été le P450 le plus largement étudié chez le mammifère.

Bien que des efforts soient faits pour appréhender plus finement les effets coopératifs multi-substrats au niveau expérimental [Baas *et al.* 2004, Fernando *et al.* 2006], une description précise manque au niveau moléculaire. De plus en plus de structures sont aujourd'hui disponibles dans la base PDB (Protein Data Bank), qui pourraient clarifier certains points, mais leur compréhension en termes de comportement enzymologique non canonique du 3A4 est toujours difficile. Le caractère multi-substrats est cependant clairement prouvé par une structure cristalline du 3A4 liée à deux molécules de kétoconazole [Ekroos & Sjögren 2006], puis deux autres structures liées à deux molécules de désoxyritonavir ont suivi [Sevrioukova & Poulos 2013b]. En raison de la forte malléabilité du 3A4 et de sa capacité à reconnaître une grande variété de substrats,

il faudrait fournir d'énormes efforts en termes de cristallisation et de détermination de structure, pour d'établir des règles déterministes dans des interactions inter-substrats. Le site actif du 3A4 étant malléable et les substrats pouvant avoir un effet sur la déformation de ce site actif, celui-ci adopte une nouvelle conformation en fonction du substrat, ce qui peut contribuer à l'existence de la grande diversité de ligands. Ce phénomène a été démontré dans l'étude de l'effet des substrats dans la liaison du monoxyde de carbone au site actif du 3A4 [Koley *et al.* 1995].

Le site actif du 3A4 accueille de nombreuses classes chimiques de substrats [Smith & Jones 1992, Li *et al.* 1995, Guengerich 1995], des plus petites molécules telles que les stéroïdes (ex. la testostérone, de poids moléculaire PM=288 Da) aux molécules volumineuses telles que la cyclosporine (PM=1202 Da) et des macrolides (érythromycine, roxithromycine). Le site actif du 3A4 est défini comme étant large et ouvert, composé essentiellement d'acides aminés hydrophobes et de quelques acides aminés neutres avec un petit nombre de chaînes latérales polaires. Les interactions hydrophobes sont donc les plus importantes. Des chaînes latérales aromatiques sont aussi présentes, permettant les interactions  $\pi$ - $\pi$  avec les substrats aromatiques [Smith *et al.* 1997a]. La grande poche du 3A4 permet l'accueil de 2 molécules simultanément, identiques ou différentes. Dans certains cas l'administration d'un substrat au sein du 3A4 active le métabolisme d'un autre substrat [Guengerich 1995], comme c'est le cas de l' $\alpha$ -naphthoflavone, capable d'activer le métabolisme de la testostérone [Lee *et al.* 1994] et de l'aflatoxine B<sub>1</sub> [Buening *et al.* 1978]. Cela suppose que l'activation résulte de l'occupation simultanée du site actif par deux substrats, ce qui conforte l'hypothèse d'une grande dimension du site actif du 3A4, mais aussi de l'utilisation de chemins d'accès au site actif, et nous montrerons dans notre étude que l'accès d'une ou deux molécules au site actif peut se faire par un même canal ou par deux canaux différents, selon leurs propriétés. L'accueil des ligands dans le 3A4 peut se faire dans de multiples sites, dans le site actif ou à l'extérieur. Des études ont suggéré que le 3A4 possède un site effecteur non catalytique dans la cavité du site actif, capable de fonction modulatrice. Les molécules de substrat et d'effecteur occupent des positions séparées, et seule la molécule de substrat accède à l'oxygène réactif. La forme et la distribution de charges du site actif sont modulées par un second ligand, ce qui accentue l'aptitude de l'enzyme à accueillir divers ligands ("*promiscuity*"), et à les accueillir simultanément [Domanski *et al.* 2001].

La régiosélectivité de l'oxydation des substrats dans le 3A4 apparaît plus dépendante des groupements chimiques réactifs au sein de la molécule que des interactions spécifiques avec le site actif. Ceci reviendrait à dire que l'orientation du ligand face à l'hème n'est pas régie par les interactions favorables avec les acides aminés, qui engendreraient une rotation de la molécule devant le site actif. Mais l'orientation est plutôt régie par un groupement fonctionnel du ligand réagissant avec l'hème pour s'y lier, et les interactions avec les acides aminés voisins stabilisent la position en second lieu. Ceci est d'autant plus vrai que les molécules volumineuses admises par le 3A4 ne peuvent pas se réorienter devant le site actif comme le feraient de petites molécules. Cette prédilection pour l'orientation du groupement fonctionnel face à l'hème pourrait être définie dès l'arrivée du substrat à l'entrée des canaux, ce qui sera discuté ultérieurement dans ce manuscrit.



**Figure 1.7 :** Visualisations des structures secondaires et tertiaires du CYP3A4. Images extraites de [Yano *et al.* 2004]. A : Deux visualisations des structures secondaires et tertiaires du 3A4. B : Par comparaison, visualisation du 2C8 (code PDB 1PQ2). Les hélices sont en bleu, les enroulements en gris, et les feuillets  $\beta$  en marron. L'hème est représenté par des bâtonnets en rouge. La surface accessible au solvant de la cavité du site actif, calculée pour une sonde de 1.4Å avec VOIDOO [Kleywegt & Jones 1994], est représentée par un maillage. Le feuillet  $\beta_1$  et diverses hélices sont annotées. C : Vue élargie de la cavité du site actif du 3A4. Différentes parties de la protéine sont indiquées par un ruban gris, et les chaînes latérales en contact avec la cavité sont indiquées par des bâtonnets avec les carbones en gris. L'oxygène, l'azote et le soufre sont respectivement en rouge, bleu et jaune. Les régions SRS sont annotées. D : Vue détaillée des régions SRS-2 et SRS-3 formant le toit de la cavité du site actif au-dessus de l'hème.

## Topologie du CYP3A4

L'isoforme 3A4 a une séquence de 503 acides aminés par chaîne (monomère), dont une hélice transmembranaire N-terminale. La structure globale du 3A4 se conforme au repliement caractéristique de la superfamille P450, avec un petit domaine dominé par un feuillet  $\beta$  N-terminal, et un plus grand domaine  $\alpha$  C-terminal, qui contient l'hème et le site actif (figure 1A, [Williams *et al.* 2004]). Le fer hémique est lié par une cystéine

conservée (C442) du côté proximal, et les propionates de l'hème interagissent avec les chaînes latérales de R105, W126, R130, R375 et R440.

L'architecture du site actif pourrait rendre compte des propriétés cinétiques inhabituelles du 3A4. La surface accessible au solvant de la cavité du site actif est évaluée à  $1386\text{\AA}^2$ , lorsqu'elle est tronquée aux étranglements des canaux du solvant. La structure du 3A4 est beaucoup plus ouverte aux alentours du fer hémique, et le volume de la cavité est plus uniformément distribué que la cavité sinueuse des autres P450s. L'hème du CYP3A4 a une plus grande accessibilité grâce à la flexibilité de la région FG. Le plus grand volume du site actif aux alentours du fer hémique, comparé avec celui du 2C8, reflète les changements de conformation dans la région SRS-5 puisqu'il passe de l'hélice K au feuillet  $\beta_1$ -brin 4, aussi bien que les différences structurales des régions SRS-2 et SRS-3 qui passent au-dessus de l'hème (figure 1.7C, 1.7D). Les acides aminés 209 à 217 et 237 à 242 (boucles FF' et G'-G) qui connectent les hélices F et G aux hélices F' et G', forment le toit de la cavité du site actif; ils contiennent les sites SRS-2 et SRS-3, respectivement, et ne présentent pas de structure secondaire régulière (figure 1.7D). Contrairement aux structures des P450s de la famille 2, les hélices F et G ne sont pas au-dessus de la cavité du site actif du 3A4 (figure 1.7A, 1.7B) parce que ces hélices sont plus courtes.

Le 3A4 a plusieurs caractéristiques qui diffèrent des autres structures de P450s. Les plus significatives sont dans les régions de la longue boucle BC (absence du brin 5 du feuillet  $\beta_1$  et de l'hélice B' chez le 3A4), et dans les hélices D à H (figure 1.7A). Les P450s eucaryotes, dont le 3A4, ont des séquences plus longues entre les hélices F et G, qui présentent généralement deux hélices supplémentaires, F' et G'. Dans la famille des CYP2 les surfaces extérieures des hélices F' et G' sont hydrophobes. Une autre région hydrophobe est localisée autour de la boucle qui suit l'hélice A (figure 1A dans [Williams *et al.* 2004]), cette hélice étant non observée dans les autres structures cristallines des P450s. Cette dernière région, avec la région des hélices F' et G', ainsi que la boucle G'-G, toutes hydrophobes, sont supposées former un domaine d'interaction avec la membrane microsomale [Williams *et al.* 2000]. Une autre caractéristique inattendue de la structure de CYP3A4 est la région qui suit l'hélice F raccourcie (figures 1A et 1B dans [Williams *et al.* 2004]), et la boucle FF'. Cette hélice F est très courte comparée à d'autres structures P450s, ce qui a pour conséquence d'allonger la boucle FF' qui est très déstructurée. Cette région est localisée au-dessus, et perpendiculaire à l'hélice I, et contient un certain nombre d'acides aminés dont on a montré par mutagenèse dirigée qu'ils peuvent avoir un rôle direct ou indirect dans la reconnaissance des substrats et la vitesse catalytique du CYP3A4. La région de la boucle FF' qui contient quatre phénylalanines du cluster Phe (213, 215, 219 et 220), ce cluster étant défini par les acides aminés 57, 108, 213, 215, 219, 220, 241 et 304 (voir table 1.2). Le cluster Phe est environ à  $10\text{\AA}$  au-dessus de l'hème. Dans le 2C9 l'hélice F est plus longue et la boucle FF' est raccourcie, ce qui explique qu'il y ait moins de changement de formes dans 2C9 et de variation de volume du site actif car il n'y a pas de mouvement de clapet de la boucle FF'. La position du cluster Phe résulte d'une conformation fermée du site actif du CYP3A4, semblable à la conformation fermée adoptée par le P450 EryF [Cupp-Vickery & Poulos 1995], mais l'arrangement spatial de l'hélice F et de la boucle FG ne sont pas conservés. L'hélice F de CYP3A4 est notamment plus courte que celles de P450 BM3 ou P450 EryF. Le 3A4 a deux hélices F' et G', qui sont absentes dans le P450 BM3, le P450 EryF et dans d'autres P450s bactériens.

**Table 1.2** : Première partie.

Structure secondaire	Séquence des structures secondaires <sup>(b)</sup>			SRS	Séquence SRS	Site actif <sup>(c)</sup>	Cluster Phe
Hélice transmembranaire N <sub>ter</sub>	1-22			-	-	-	-
<b>Ancre-A</b>				-	-	-	-
<b>Domaine β :</b> N <sub>ter</sub> + Hélices A, A', B, K' + Feuillet β1 (brins 1-4) + Feuillet β2 (brins 1,2)	<b>25-94</b> <b>370-402</b>			SRS-5		F435, C442, I443, G444 boucle [K'- β1brin3]	
1 <sup>ère</sup> partie du domaine β (avant l'hélice C): N <sub>ter</sub> + Hélice A+ Hélice A' Feuillet β1(brins 1,2)+Hélice B	25-94			-	-	-	57 hélice A'
Feuillet β1 (brins 1-4)	70-84 brins 1,2	373-376 brin 4	393-396 brin 3	SRS-5 N <sub>ter</sub> -brin4	P368- V376	-	-
Feuillet β2 (brins 1,2)	381-388			-	-	-	-
Hélice K'	399-402						
Boucle B-C	95-122			SRS-1 <sup>(a)</sup>	V101- D123	R105, S119	108
<b>Région F-G</b>	<b>202-260</b>			SRS-2 <sup>(a)</sup> SRS-3 <sup>(a)</sup>			
Hélice F	202-209			SRS-2 <sup>(a)</sup> C <sub>ter</sub>	V204- L211		-
Boucle F-F'	209-219 4 aas chargés (K, R et 2 D) et les autres aas apolaires			SRS-2 <sup>(a)</sup>	V204- L211	R212	213, 215, 219
Hélice F'	219-225						220
Boucle F' -G'	226-229						-
Hélice G'	230-237						-
Boucle G' -G	237-243			SRS-3 <sup>(a)</sup>	237-242		241
Hélice G	244-260			SRS-3 <sup>(a)</sup> N <sub>ter</sub>	I238-E244		-
Hélice I	291- 323			SRS- 4	E294- S312  301-309 E308 et T309 conservés	I301, F302, F304, A305, E308, T309, T310	304

**Table 1.2** : Séquence d'acides aminés des structures secondaires, des SRS, et des acides aminés du site actif et cluster Phe dans le CYP3A4. (a) Au-dessus de l'hème (borde le site actif). (b) Les séquences ont été obtenues par le logiciel de visualisation Maestro (Schrödinger, LLC, New York, 2014). (c) Relevés manuellement ou sur le CSA (Catalytic Site Atlas, [Furnham *et al.* 2014]) ou par mutagenèse dirigée. Encadrement en gras du domaine β et de la région FG.

**Table 1.2** : Deuxième partie.

Boucle K- feuillet $\beta_1$ [brin4]	366-373	SRS-5	P368- V376	I369, A370, R372, E374	-
Boucle C-terminale	467-489	SRS-6	K476- P485	-	-

**Table 1.2** : Suite et fin.

Le site actif du 3A4 est défini par :

- Selon le CSA (Catalytic Site Atlas, [Furnham *et al.* 2014]) : [E308; T309; T310] (hélice I); [F435; C442; I443; G444] (boucle K'L); N451 (hélice L).
- Manuellement : feuillet  $\beta_1$ ; [R105, R106, S119] (boucle BC); [R212] (boucle FF'); [I301, F302, F304, A305, T309] (hélice I); [I369, A370; R372, E374] (boucle K-feuillet  $\beta_1$ -brin4).
- Selon la mutagenèse dirigée [Yano *et al.* 2004] : acides aminés impliqués dans le site actif : S119, I301, F304, A305, I369, A370 et E374.
- Poche polaire : bordée par les chaînes latérales de R106, R372, et E374.
- Le site contient 2 poches hydrophobes : la poche 1 (bordée par I369 et A370), et la poche 2 (bordée par L210, F241, I301 et F304).
- Le ligand qui occupe le site actif établit des contacts hydrophobes non-liés et des contacts polaires non spécifiques.

La région des canaux 2 (voir section 1.10.2), qui contient les canaux 2a à 2f, est située entre les structures secondaires suivantes : le domaine N-terminal, les hélices A/A', le feuillet  $\beta_1$ , la boucle BC, la région FG, l'hélice I et la boucle C-terminale. Les hélices A, A', B, F, G et leurs boucles sont des régions variables associées à la liaison du substrat. Les trois régions du feuillet  $\beta_1$ , de la boucle BC et de la région FG sont fortement variables en séquence et en structure parmi les P450s. Le domaine des canaux 2 se situant entre ces trois régions, est reconnu pour son importance dans la spécificité du substrat à travers les contacts avec le substrat. A partir des analyses des séquences de la famille CYP2, un ensemble de 6 SRS a été identifié contenu dans cette zone des canaux 2 (cf. Table 1.2), et il a été proposé que ces 6 SRS soient impliqués dans la liaison des substrats de par le grand nombre de substitutions non synonymes [Gotoh 1992]. La région FG fait partie du canal d'accès et du plafond du site actif (SRS-2 et SRS-3). La plupart des SRS sont dans la poche hémiquine, impliqués dans l'orientation de substrat dans la poche.

Les SRS sont les régions les plus variables des P450s. Ces régions définissent géométriquement le site actif et les canaux [Yano *et al.* 2004, Williams *et al.* 2004]. Les acides aminés des SRS influencent la liaison au substrat, la coopérativité et/ou la régiosélectivité. Un seul acide aminé dans une seule région SRS peut profondément altérer la fonction catalytique des enzymes et par conséquent les voies métaboliques des molécules [Xue *et al.* 2001].

## Substrats et inhibiteurs du CYP3A4 : corrélations et mécanisme de liaison au site actif

Le métabolisme du 3A4 couvre une large gamme de molécules thérapeutiques, lipophiles, neutres ou basiques. Les réactions principales du 3A4 avec le substrat sont l'élimination d'hydrogène dans le cas de l'hydroxylation de carbone, ou l'élimination d'électrons dans le cas de N-désalkylation [Smith & Jones 1992]. Le site de métabolisation est souvent un groupement allylique ou bien un azote dans le cas d'une N-désalkylation. Le 3A4 métabolise principalement les substrats lipophiles avec prédominance des interactions hydrophobes. L'accès d'un ligand au site de métabolisation et son exposition à l'hème ne sont pas forcément dus à sa fonction labile mais à des raisons statistiques. Un groupement tertibutyle serait par exemple plus facilement oxydable qu'une autre fonction plus labile, du fait de son équivalence avec un maximum de carbones primaires [Smith *et al.* 1997b].

Le site actif et les canaux des CYPs sont supposés contenir des molécules d'eau au repos [Wade 1990]. Dans le 3A4, la liaison aux substrats est due principalement aux forces lipophiles, qui résultent de l'expulsion de l'eau par le ligand, dans les canaux et au niveau du site actif. Cette expulsion de l'eau fournit la force motrice entropique après la liaison du substrat, qui induit le changement de l'état de spin et la formation de l'entité catalytique (FeO)<sup>3+</sup>. De plus, les molécules d'eau associées aux ligands doivent être retirées avant liaison au site actif. Nous supposons que c'est aussi le cas au niveau des canaux, car on n'observe pas de coexistence du ligand avec des molécules d'eau dans un canal, alors que des molécules d'eau occupent les parties non occupées par le ligand. Lorsque les ligands ont des groupements fonctionnels associant fortement les molécules d'eau, cela produit une couche aqueuse enveloppant partiellement la molécule, comme dans le cas de la sulfonylurée du tolbutamide pour le 2C9. Pour permettre la liaison au site actif composé d'acides aminés hydrophobes, cette enveloppe aqueuse doit être dénudée. L'énergie requise par ce processus peut être trop grande par rapport à celle due à la compensation par les interactions hydrophobes. Cependant, si le site actif était capable former des liaisons hydrogène avec des molécules d'eau, ces dernières pourraient jouer un rôle important dans la liaison du ligand à l'enzyme [Jones *et al.* 1996]. Dans un tel processus, la région lipophile de chaque ligand serait encore capable de déplacer l'eau et les effets entropiques de liaisons résultantes mèneraient au changement d'état de spin pour la formation de l'entité (FeO)<sup>3+</sup>. Les substrats du 3A4 ont des valeurs de log(D) (coefficient de partage avec l'octanol ou le cyclohexane; pas de liaison hydrogène), qui sont supérieures à celles du 2C9 [Walker *et al.* 1995]. Ceci indique que l'environnement des ligands dans le 3A4 est dénué d'eau, alors que d'autres substrats sont plus hydratés chez d'autres CYPs, tel que le 2C9.

Les fonctions lipophiles des substrats changent de conformation selon leur environnement. Elles sont cachées dans un environnement aqueux et plus étendues dans un environnement lipophile. Ce phénomène a été confirmé par une étude du CYPBM-3 (CYP102) bactérien soluble. Cette étude porte sur des acides gras qui adoptent une forme globulaire dans une solution aqueuse. Une fois le ligand entré dans le canal hydrophobe, celui-ci s'ouvre et adopte une conformation élargie, avec le groupement de tête de la molécule dirigé vers l'hème et la fonction acide polaire dirigée vers le solvant. Ceci soulève la question de savoir si le contact de la molécule avec la bouche du canal (lipophile) s'est effectué par la tête lipophile. Ce principe peut être étendu aux substrats du 3A4 comme la cyclosporine [Smith & Jones 1992]. Le groupement allylique, site de métabolisme primaire

de la cyclosporine, est protégé du solvant aqueux, alors que la conformation s'ouvre dans un solvant aprotique lipophile retirant le site de métabolisme primaire de la partie centrale de la molécule. Ce positionnement des ligands le long des canaux hydrophobes, initialement gêné stériquement, comme indiqué par les structures cristallisés ou en solution aqueuse, peut aider à comprendre la sélectivité du 3A4. Ceci sera discuté dans le chapitre 3 de ce manuscrit.

Les interactions de liaison dans le 3A4 sont également hydrophiles et spécifiques, comme les liaisons hydrogènes et les interactions des paires d'ions. Ces interactions spécifiques permettent le mouvement du substrat dans le site actif, où le substrat peut adopter plusieurs conformations et s'orienter différemment vis-à-vis du  $(\text{FeO})^{3+}$ , ce qui induit des produits d'oxydation différents. Ceci est illustré par la métabolisation du midalozam, qui produit 2 métabolites : le 4-hydroxymidazolam et le 1'-hydroxymidazolam [Gorski *et al.* 1994]. Parfois deux molécules différentes sont métabolisées simultanément par le 3A4, comme c'est le cas pour le phenanthrène et la 7,8-benzoflavone [Shou *et al.* 1994], qui ont donc accès simultanément à l'entité ferroxo.

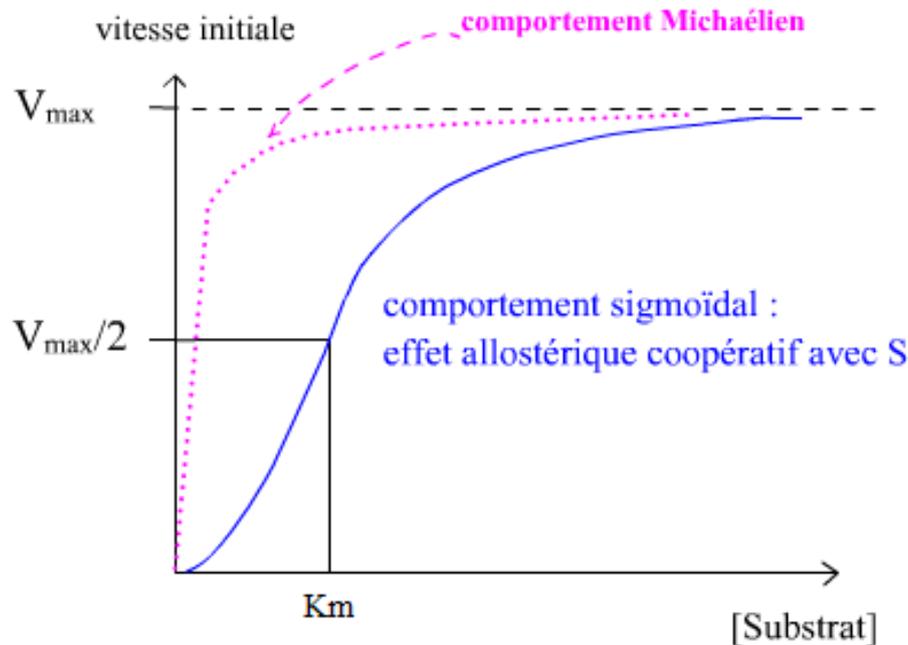
Pour les substrats modérément hydrophiles (valeurs négatives de  $\log(D)$  ou de  $\log(P)$  dans l'octanol et le cyclohexane), les interactions avec le 3A4 se font par liaisons hydrogène. Le fait que le 3A4 lie des substrats aux propriétés différentes avec des interactions différentes (hydrophiles/hydrophobes), suggère que le 3A4 interagit avec les ligands par des voies différentes, bien qu'il existe une voie privilégiée induisant une relative sélectivité plutôt qu'à une spécificité [Smith *et al.* 1997b].

Beaucoup de médicaments peuvent agir comme des inhibiteurs modérés ou puissants du 3A4, mais non nécessairement par la liaison à l'azote. Par exemple, le 3A4 peut être inhibé par des antibiotiques macrolides, des antidépresseurs, des stéroïdes, et par plusieurs autres types de médicaments via la liaison réversible ou irréversible, et par le mécanisme d'inhibition [Zhou 2008]. L'inactivation du 3A4 cause fréquemment des interactions médicamenteuses défavorables et pourrait mener à un changement de propriétés du médicament, en efficacité ou en toxicité. Nous étudierons l'interaction structurale du 3A4 avec des inhibiteurs dont les deux plus puissants, le kétoconazole, le ritonavir, et des analogues du ritonavir comme le désoxyritonavir.

## **Liaison au substrat/inhibiteur et coopérativité dans le CYP3A4**

La fixation de la molécule effectrice sur un site allostérique induit un changement de conformation spatiale de l'enzyme allostérique, ayant pour conséquence soit une augmentation (effecteur positif), soit une diminution (effecteur négatif) de l'activité enzymatique. Pour l'allostérie, cela a pour conséquence de modifier le site de liaison de l'un au moins des substrats. Dans le modèle Monod-Wyman-Changeux, les enzymes allostériques existent sous deux conformations différentes : l'une appelée T (pour tendue), désignant conventionnellement la forme de faible affinité pour le substrat, l'autre appelée R (pour relâchée), de forte affinité pour le substrat. Il existe une allostérie dite positive, où l'on parle de fixation coopérative, telle que la fixation d'un effecteur augmente l'affinité de liaison du ligand. La molécule effectrice peut être le ligand lui-même, qui modifie dans ce cas l'affinité des autres sites de fixation (effet homotrope, par opposition à l'effet hétérotrope qui concerne des molécules de nature différente). Il s'agit du système K (K est le symbole de la constante d'affinité), dont la courbe représentant la vitesse de réaction en fonction de la concentration en substrat, est de forme sigmoïde. Le cas inverse existe

aussi : dans une allostérie négative la fixation de l'effecteur diminue l'affinité du ligand. C'est le cas du système V (pour vitesse), dont la courbe représentative est telle que sur la figure 1.8.



**Figure 1.8** : Evolution des vitesses de réaction  $V_i$  en fonction de la concentration en substrat. En rose : comportement Michaélien de la fonction. En bleu : comportement sigmoïdal de la fonction, effet allostérique coopératif (ou homotrope) positif de type K avec le substrat.

Par extension, la notion d'allostérie s'applique au cas des canaux ioniques, bien qu'il n'y ait pas de réaction chimique catalysée mais plutôt un transfert de matière (ions) à travers une membrane. Dans ce cas, la fixation d'une molécule effectrice positive favorisera l'ouverture du canal, quel que soit le mode d'ouverture de ce canal, et inversement, pour un effecteur négatif, le processus d'ouverture sera défavorisé. On parle respectivement d'agoniste et d'antagoniste des récepteurs allostériques. On peut imaginer le même processus pour les canaux des CYPs, où une molécule effectrice favoriserait l'ouverture non spontanée du canal du bloc 2, pour le passage puis la réactivité du second substrat (voir section 3). Les protéines allostériques sont des oligomères (elles présentent donc une structure quaternaire), comme l'est le CYP3A4, et les propriétés cinétiques ou de saturation ne sont plus toujours d'allure Michaélienne. En effet, la liaison d'un ligand sur un site à l'état T induit des changements conformationnels propagés (trans-conformation plus ou moins globale) et réversibles vers l'état R par déplacement des structures secondaires, ce qui modifie les propriétés de liaison des autres sites. Concernant, l'effet coopératif (ou homotrope) positif, les différences de propriétés enzymatiques entre les états T et R dans l'effet coopératif de type K sont au niveau de l'affinité pour le substrat et non au niveau de la vitesse de réaction. Ceci explique l'allure sigmoïde des fonctions  $v_i=f([S])$  (cinétique des enzymes allostériques) (figure 1.8).

L'effecteur allostérique, dans le cas des systèmes K (appelés activateurs et inhibiteurs allostériques, au sens strict), remplit 3 conditions :

- C'est une molécule autre que le substrat, ayant un ou plusieurs sites de fixation sur l'enzyme.
- Le(s) site(s) en question est (sont) distinct(s) du site catalytique.
- La liaison de l'effecteur allostérique entraîne des changements conformationnels propagés à l'échelle de l'enzyme entière, et à l'origine d'une activation ou d'une inhibition.

Il faut remarquer que l'effecteur allostérique se lie à un site totalement distinct et éloigné du site catalytique, et que la liaison de cet effecteur modifie (par changement conformationnel induit) les propriétés du site catalytique pour son substrat : cet effet est qualifié d'effet hétérotrope. Ceci est à comparer avec l'effet coopératif, homotrope. Dans les systèmes allostériques K, la protéine passe d'une forme à l'autre sous l'influence de la liaison d'un ligand qui peut être le substrat (sur le site actif), ou un effecteur (site différent du site actif). Les différentes formes ne présentent pas la même affinité pour le substrat. Par exemple, pour qualifier le fait que la liaison à un activateur A favorise la liaison à d'autres molécules de A, on peut parler d'effet homotrope positif de A sur sa propre liaison. Pour qualifier le fait que l'activateur A provoque un déplacement vers le(s) état(s) à forte affinité pour le substrat S, on parlera d'effet hétérotrope positif de A.

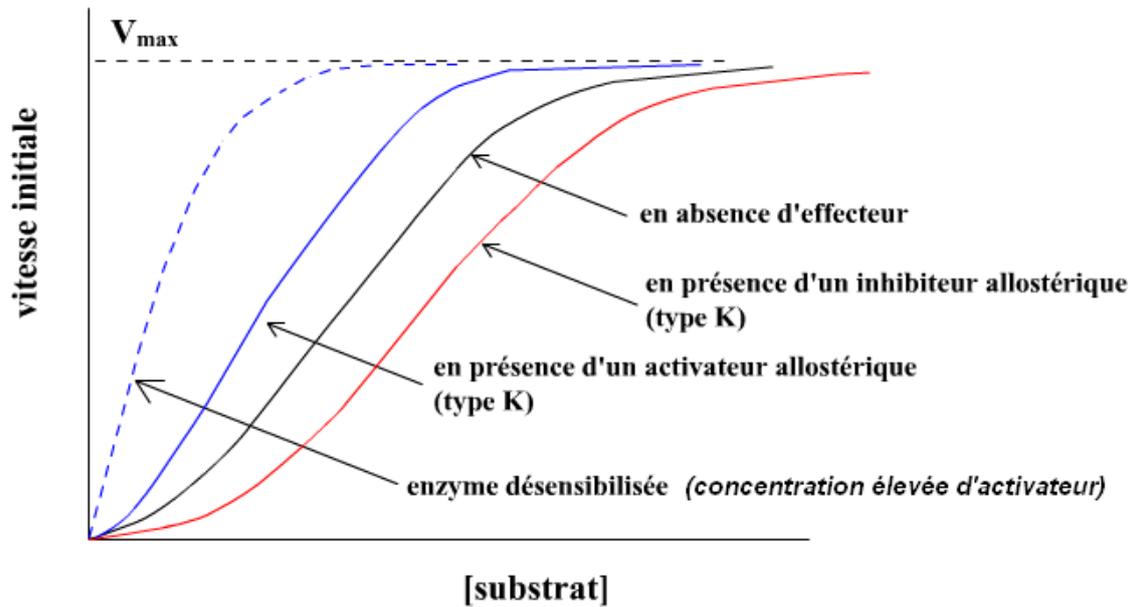
Dans ce qu'on appelle les systèmes K (les plus classiques), les changements conformationnels dus à la liaison avec l'effecteur allostérique modifient l'affinité des différentes sous-unités pour le substrat catalysé, et non pas le coefficient catalytique ( $k_0$ ,  $V_{max}$ ) :

- La liaison d'un inhibiteur allostérique (sur ses sites spécifiques) va favoriser l'état T à faible affinité pour le substrat. On peut évidemment raisonner en terme de déplacement d'équilibre.
- La liaison d'un activateur allostérique (sur ses sites spécifiques) déplace l'enzyme vers l'état R à forte affinité pour le substrat. A concentration saturante en activateur, toutes les sous-unités de l'enzyme sont déplacées sous la forme R : le comportement de l'enzyme devient Michaélien (il ne peut plus y avoir de comportement sigmoïdal puisque tout effet coopératif du substrat est aboli, car l'enzyme n'existe plus que sous la même forme, R). On parle parfois de "désensibilisation de l'enzyme".

La figure 1.9 illustre le cas d'une enzyme allostérique présentant un effet coopératif du substrat, avec un effecteur activateur et un effecteur inhibiteur allostérique, pour un système K.

Dans les systèmes allostériques V, la protéine passe d'une forme à l'autre sans changer d'affinité pour ses ligands, mais c'est la vitesse de l'acte catalytique qui varie, c'est à dire le coefficient catalytique  $k_0$ , donc la  $V_{max}$ . Une allure cinétique qui semble à première vue Michaélienne peut cacher de l'allostérie ! Il existe des cas plus complexes, où affinité et  $k_0$  sont modifiés par des mouvements transconformationnels allostériques et homotrophiques. Certains auteurs parlent alors d'allostérie K-V.

Le CYP3A4 est un complexe enzymatique héminique qui présente des cinétiques atypiques non Michaéliennes, et de coopérativité hétérotrophe envers plusieurs substrats [Guengerich 1999, Hutzler & Tracy 2002]. Ainsi, le CYP3A4 offre un comportement cinétique atypique *in vitro* avec l'aflatoxine B1, l'amitriptyline, la cabamazépine, la progestérone et le diazépam [Schwab *et al.* 1988, Ueng *et al.* 1997]. Une des toutes premières observations sur le CYP3A4 était des cinétiques d'états d'équilibre atypiques avec la testostérone, la progestérone, l' $\alpha$ -naphthoflavone, et quelques autres substrats,



**Figure 1.9 :** Evolution des vitesses de réaction  $V_i$  en fonction de la concentration en substrat pour une enzyme allostérique de type K à effet coopératif du substrat. En noir : en l'absence d'effecteur. En bleu : en présence d'un effecteur activateur. En rouge : en présence d'un effecteur inhibiteur allostérique.

manifestés par une sigmoïde ou un graphe biphasé de la vitesse de réaction en fonction de la concentration du substrat [Guengerich 1999]. Un tel comportement allostérique est dû la liaison simultanée de plus d'une molécule de substrat dans, ou près du site actif du CYP3A4, qui peut soit stimuler soit inhiber le renouvellement (*turnover*).

La liaison du substrat au CYP3A4 a été largement étudiée par une variété de techniques [Isin & Guengerich 2008], menant aux conclusions suivantes :

1. Les substrats ne sont pas bloqués dans le site actif du CYP3A4, et peuvent se dissocier et se relier de nouveau pendant les étapes différentes du cycle catalytique.
2. A la suite d'un libre ajustement, l'affinité de liaison et le renouvellement d'un substrat (*turnover*) peuvent être modulés par une autre molécule présente dans le site actif.
3. L'association au substrat peut impliquer une étape n'induisant pas de changement dans le spectre Soret de l'hème, et ne pouvant pas ainsi être détectée par spectroscopie d'absorbance ("absorbance silencieuse").
4. Avant le déplacement dans la cavité du site actif, les ligands du CYP3A4 peuvent interagir avec un site de liaison périphérique.

On ne sait toujours pas si la raison principale de la cinétique atypique est la liaison multiple au substrat [Harlow & Halpert 1998, He *et al.* 2003], ou si c'est la préexistence de conformères cinétiquement différents [Koley *et al.* 1997, Atkins *et al.* 2001, Davydov *et al.* 2003]. Le nombre de molécules simultanément liées à une enzyme et présentant une coopérativité positive ou négative, est traditionnellement évalué en se basant sur le coefficient de Hill ( $n_H$ ), dérivé de la cinétique de l'état stationnaire ou des titrages d'équilibre de liaison au substrat. Les valeurs de  $n_H$  rapportées pour divers substrats du

CYP3A4 n'excèdent jamais 2.0 [Guengerich 1999, Farooq & Roberts 2010]. Cependant, certaines données expérimentales suggèrent que trois molécules de testostérone puissent simultanément se lier au CYP3A4 [He *et al.* 2003, Baas *et al.* 2004]. Plus récemment, ceci a été confirmé par l'analyse globale de liaison au substrat à l'équilibre, et par les cinétiques d'état stationnaire de consommation du NADPH pendant l'hydroxylation de la testostérone par le CYP3A4 [Denisov *et al.* 2007]. Cette étude a conclu que seule la liaison à la deuxième molécule de testostérone mène à la formation significative du produit, tandis que la troisième étape de liaison améliore l'efficacité de couplage entre la consommation électronique et la formation de métabolite. Au contraire, l'analyse par RPE (résonance paramagnétique électronique) de l'équilibre d'état du spin ferrique et du ligand dans le CYP3A4 [Roberts *et al.* 2005], indique une stoechiométrie de la testostérone limitée à 2:1, et indique que la première molécule de testostérone se lie avec une plus grande affinité, mais a un effet négligeable sur l'équilibre de spin (un mode non productif), et rend la liaison à la deuxième testostérone moins énergétiquement favorable. La deuxième étape de liaison, cependant, augmente la probabilité de produire l'état haut du spin (coopérativité positive), bien que le fort déplacement de spin ne soit pas complet. Il est intéressant de noter que le couplage des énergies libres défavorise fortement, non seulement la deuxième liaison à la testostérone, mais aussi liaison de la molécule d'eau à l'hème.

### 1.10.3 Résumé des problèmes posés par l'isoforme 3A4

Le CYP3A4 métabolise plus de médicaments que toutes les autres isoformes ensemble. L'enzyme fournit aussi une barrière significative à la biodisponibilité de nouveaux candidats médicaments, contribuant au ralentissement des projets de développement. Des interactions entre substrats et inhibiteurs de l'enzyme peuvent profondément affecter la sécurité ou l'efficacité du traitement médicamenteux [Wilkinson 1996, Tanaka 1998]. Une caractéristique remarquable du 3A4 est sa capacité à accueillir et à métaboliser des composés divers en taille et en structure chimique, ainsi que sa liaison coopérative au substrat, qui mène souvent aux interactions médicamenteuses et aux effets secondaires toxiques [Guengerich 1999]. En raison de la dégradation rapide des produits médicamenteux, le 3A4 peut baisser leur biodisponibilité et leur efficacité thérapeutique, et vice versa, les concentrations de médicaments dans le plasma peuvent être augmentées si le 3A4 est inhibé. De plus, pour la détoxification et la clairance de xénobiotiques, le 3A4 exerce des effets génotoxiques via l'activation de procarcinogènes [Rendic & Guengerich 2012], et contribue aux interactions médicamenteuses défavorables en raison de la capacité à accueillir dans le site actif deux ou plus de deux molécules, du même type ou de types différents (coopérativité). Ceci permet au 3A4 d'oxyder plus de la moitié de tous les médicaments administrés. Le 3A4 catalyse la clairance métabolique d'un grand nombre de médicaments cliniquement utilisés, et un certain nombre d'interactions médicamenteuses défavorables reflètent l'inhibition ou l'induction de l'enzyme.

Les structures des CYPs mammaliens adoptent toutes le repliement général décrit initialement en 1987, lorsque la structure du CYP101 a été élucidée par cristallographie X [Poulos *et al.* 1987]. Le 3A4 présente une relativement grande cavité de liaison au substrat (ou inhibiteur), compatible avec sa capacité d'oxydation des substrats volumineux comme de la cyclosporine (PM 1203 Da), la bromocriptine (PM 655 Da), des statines, des taxanes, et des macrolides antibiotiques. Bien que le volume du site actif soit grand et que la cavité du site actif soit importante au voisinage du fer hémique, le peu de contraintes sur les

mouvements de petits substrats près du site d'activation de l'oxygène peut diminuer l'efficacité d'oxydation du substrat, et la présence d'une deuxième molécule de substrat peut, à son tour, par des contraintes spatiales, améliorer l'efficacité d'oxydation.

Plusieurs études indiquent que la cinétique de liaison au ligand du 3A4 n'est pas toujours Michaelienne, et il a été suggéré que deux ou plus de deux ligands peuvent être liés simultanément au 3A4 [Harlow & Halpert 1998, Kenworthy *et al.* 2001, Davydov *et al.* 2003, Atkins 2005]. Les autres membres de la famille 3A présentent aussi des caractéristiques cinétiques inhabituelles, qui suggèrent l'occupation simultanée de la cavité du site actif par de petits substrats. Cependant, le mécanisme de la cinétique atypique d'oxydation du substrat par le 3A4 fait toujours débat. Actuellement, il y a une évidence expérimentale soutenant tant le mécanisme de coopérativité ligand induit que le mécanisme de sélection conformationnel. Le site de liaison effecteur peut être périphérique (en surface), dans le canal, ou dans la cavité catalytique. Alors que l'attention se porte actuellement sur le ou les sites de liaison périphériques du ligand, on ne sait pas si ce sont des sites de liaison effecteurs ou des zones d'amarrage du ligand non spécifiques, d'où des molécules hydrophobes pourraient être transportées vers le site actif en vue d'une éventuelle métabolisation. L'emplacement précis de ces sites de liaison d'effecteur dans la cavité catalytique reste également inconnu, et les structures du co-substrat activateur du 3A4 restent à déterminer. Une autre question concerne le rôle relatif des interactions hydrophiles et polaires dans la liaison du ligand. Bien que la domination des forces hydrophobes dans l'association 3A4-substrat-inhibiteur ait été bien documentée et soit largement acceptée, les études suggèrent que des contacts polaires fournis par S119, R212 et T224 pourraient être également cruciaux [Sevrioukova & Poulos 2013a]. Ces nouvelles pistes doivent être explorées pour comprendre les mécanismes contrôlant la liaison du 3A4 au substrat et à l'inhibiteur, qui mènent à une rationalisation et à une prédiction plus précise des interactions protéine-médicament et médicament-médicament.

Il y a en 2016 24 structures du 3A4 cristallisées dans la base PDB. L'étude des structures du 3A4 facilite la compréhension de la sélectivité de cette isoenzyme au substrat. Nous utiliserons les données récemment publiées pour mieux comprendre cette isoenzyme (section 3), et en particulier nous analyserons les caractéristiques structurales du 3A4 et l'implication des acides aminés bordant les canaux dans la déformation du 3A4 pour l'adoption de ses 3 conformations majeures (voir section 3.4), et sa relation avec la coopérativité (voir section 3.8.2).

#### 1.10.4 Structure de l'isoenzyme CYP1A2 et de ses ligands

Le cas des isoenzymes CYP1A est aussi extrêmement illustratif. Les isoenzymes 1A sont responsables de la désactivation métabolique de nombreux médicaments et de composés endogènes [Kim & Guengerich 2005], et en même temps, de l'activation métabolique de composés chimiques procarcinogéniques connus [Nebert & Dalton 2006]. Puisque les isoformes 1A1 et 1A2 sont fortement similaires en séquence, le nombre d'éléments de séquences potentiels qui expliqueraient n'importe quelle différence de propriétés fonctionnelles devrait être plutôt limité. En réalité, de légères différences dans les séquences peuvent affecter profondément la palette de substrats reconnus. Les canaux des 1A ont été étudiés récemment en utilisant des chimères : il apparaît des corrélations entre les variantes des enzymes et leur capacité à métaboliser des substrats stéroïdiques [Urban *et al.* 2014]. Il a été émis l'hypothèse que les canaux différencient les

petits substrats des substrats volumineux grâce aux acides aminés situés à la surface de la protéine [Urban *et al.* 2015].

Avant la résolution de la structure cristallisée, le site actif du 1A2 avait été modélisé par [Lewis & Lake 1996]. Le site actif du 1A2 a été caractérisé par sa dimension et son hydrophobie, avec des liaisons hydrogène et des interactions  $\pi$ - $\pi$ . La cavité de liaison contient plusieurs chaînes latérales aromatiques, incluant les cycles plans de Ph181 et de Tyr437, qui restreignent la taille et la forme de la cavité uniquement pour les structures planaires (environ 6Å de largeur). Cette caractéristique du site de liaison peut expliquer la préférence des CYP1A2 pour les composés aromatiques planaires et hydrophobes, et même à avoir des interactions  $\pi$ - $\pi$ . Deux autres acides aminés aromatiques, Phe88 et Phe266, influencent la dimension de la cavité, et se situent aux extrémités de chaque côté de la cavité. Certaines chaînes latérales avec groupements donneurs de liaisons hydrogène sont stratégiquement placés, évoquant une interaction avec les substrats : Thr78, Thr87 et Thr184.

### **Substrats du CYP1A2 : corrélations et mécanisme de liaison au site actif**

Le CYP1A2 est capable d'accueillir des molécules neutres et/ou basiques. Il a été très tôt déterminé que les ligands du 1A2 sont aplatis, aromatiques et lipophiles, comme le benzo[a]pyrene [Lewis 1986], avec au moins un site supposé donner des liaisons hydrogène. Le CYP1A2 lie des ligands par don de liaisons hydrogène à des atomes du ligand à potentiel électrostatique négatif (accepteurs de liaison hydrogène, comme l'oxygène du carbonyle). Les interactions  $\pi$ - $\pi$  entre les régions aromatiques des substrats et les cycles des acides aminés aromatiques sont les plus importantes dans l'interaction avec le CYP1A2 [Lewis & Lake 1996]. Cette interaction est différente de l'interaction lipophile car elle requiert le nuage électronique des systèmes aromatiques.

## 2 METHODES THEORIQUES D'EXPLORATION DES CAVITES ET CANAUX DANS LES PROTEINES

### 2.1 Définitions, terminologie

Nous présentons ci-dessous les termes mathématiques essentiels à la compréhension des méthodes de calcul des canaux.

- **Graphe, sous-graphe** : Un graphe est un ensemble de nœuds (ou sommets) reliés par des arêtes, orientées ou non (voir exemple figure 2.1). L'ensemble des sommets est noté  $S$ , l'ensemble des arêtes est noté  $A$ , et le graphe est noté  $(S,A)$ .

Un graphe est orienté lorsque les arêtes ont un sens; sinon, le graphe est non orienté : ce sera le cas de tous les graphes considérés par la suite.

Remarque : dans la définition d'un graphe, les nœuds n'ont pas de coordonnées.

Un sous-graphe d'un graphe  $(S,A)$  est un graphe dont les sommets sont dans  $S$  et dont les arêtes sont dans  $A$ .

Exemple de graphe : en chimie, on associe souvent à la formule développée d'un composé un *graphe moléculaire*, où les nœuds sont les atomes et les arêtes sont les liaisons chimiques; dans cet exemple, on néglige la nature des atomes et la nature des liaisons.

- **Chemin, composante connexe** : Un chemin dans un graphe est une suite d'arêtes joignant deux nœuds.

Soit un graphe  $(S,A)$ . Une composante connexe de  $(S,A)$  est un sous-graphe  $(S',A')$  de  $(S,A)$  satisfaisant aux trois conditions suivantes : (1) tout couple de nœuds de  $S'$  est relié par au moins un chemin, (2) pour tout couple de nœuds  $x$  dans  $S'$  et  $y$  dans  $S-S'$  il n'existe aucun chemin joignant  $x$  et  $y$ , (3) toute arête de  $A$  joignant deux nœuds de  $S'$  est une arête de  $A'$ .

Un graphe est connexe lorsqu'il ne comporte qu'une seule composante connexe.

Exemple : le graphe moléculaire du nitrate d'ammonium est un graphe non connexe comportant deux composantes connexes, qui sont la composante associée à l'anion nitrate et la composante associée au cation ammonium.

- **Coûts, chemin de coût minimal** : Soit un graphe connexe  $(S,A)$  pour lequel on associe à chaque nœud de  $S$  un nombre réel strictement positif, appelé coût. Le coût d'un chemin joignant deux nœuds de  $S$  est la somme des coûts des nœuds situés sur ce chemin.

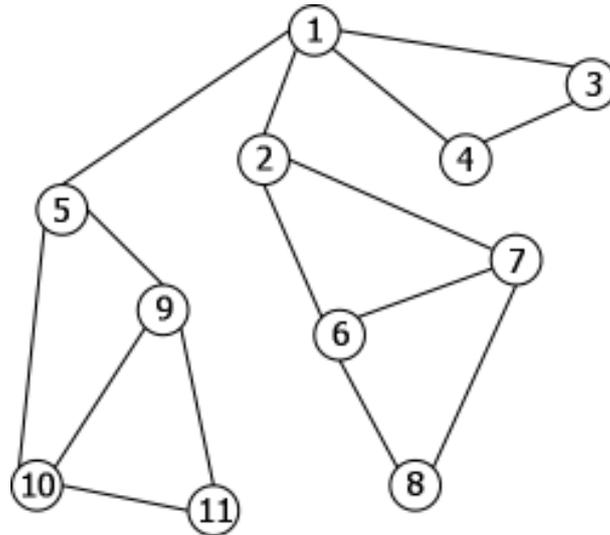
Pour un couple de nœuds donné, il existe un chemin de coût minimal, non nécessairement unique.

Soit maintenant un graphe connexe  $(S,A)$  pour lequel on associe à chaque arête de  $A$  un nombre réel strictement positif, appelé coût. Le coût d'un chemin joignant deux nœuds de  $S$  est la somme des coûts des arêtes situées sur ce chemin.

Pour un couple de nœuds donné, il existe un chemin de coût minimal, non nécessairement unique. Celui-ci sera par la suite noté **MCP** (*Minimal Cost Path*).

Le coût associé à un MCP est une distance, au sens mathématique du terme (notamment, l'inégalité triangulaire est valide).

Lorsque les coûts sur les arêtes ne sont pas spécifiés, on leur attribue implicitement la valeur 1, et ainsi le coût associé à un chemin n'est autre que le nombre d'arêtes de ce



**Figure 2.1** : Exemple de graphe non orienté comportant une seule composante connexe. Les nœuds sont numérotés. Il n’y a pas de coûts associés aux nœuds ni aux arêtes. Le plus court chemin joignant les nœuds 8 et 11 a pour longueur 6.

chemin (exemple figure 2.1).

Remarque : par la suite, il ne faudra pas confondre les deux types de chemins de coûts minimaux.

- **Ensemble convexe** : Soit un espace vectoriel  $E$ . Un sous-ensemble  $K$  de  $E$  est un convexe si, pour tout couple de points  $x$  et  $y$  de  $K$ , le segment  $[x,y]$  est inclus dans  $K$ .

Exemples : un cube, une sphère, un disque, sont des convexes. Un cercle n’est pas convexe. En pratique, la notion de convexe traduit l’absence de cavité.

- **Enveloppe convexe** : L’enveloppe convexe d’un sous-ensemble  $X$  d’un espace vectoriel  $E$  est l’intersection de tous les convexes de  $E$  contenant  $X$ .

Dans le cas particulier où  $X$  est constitué de  $n$  points en position quelconque dans l’espace tridimensionnel, l’enveloppe convexe est le plus petit domaine polyédrique convexe englobant ces  $n$  points, et les faces de ce polyèdre sont triangulaires.

Pour une protéine, l’enveloppe convexe des points dont les coordonnées sont celles des atomes protéiques fournit une première approximation du domaine protéique, mais sans aucune cavité.

- **Triangulation** : Soit un ensemble  $X_n$  de  $n$  points dans un espace Euclidien (espace vectoriel muni d’un produit scalaire), de dimension  $d$  finie.

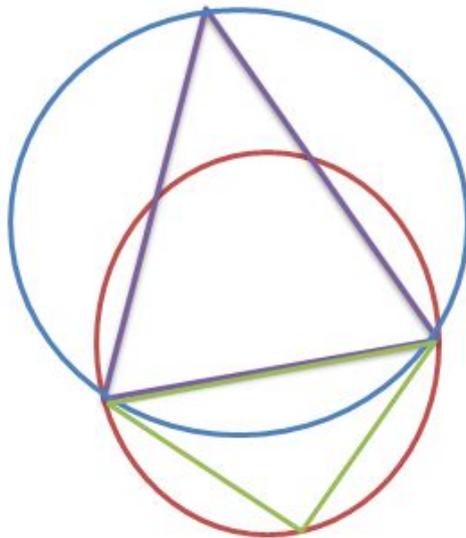
Afin de simplifier notre exposé, nous supposons par la suite que les distances dérivent de la norme induite par le produit scalaire usuel.

Une triangulation de  $X_n$  est une subdivision de l’enveloppe convexe  $K$  de  $X_n$  en  $d$ -simplexes ayant chacun pour sommets  $d + 1$  points distincts de  $X_n$ , telle que l’union de ces simplexes soit égale à  $K$ , et telle que l’intersection des intérieurs respectifs de deux quelconques de ces simplexes soit vide.

Dans le cas de  $n$  points de l’espace 3D, ces simplexes sont des tétraèdres adjacents (accolés deux à deux par une face triangulaire), chaque tétraèdre pouvant avoir au maximum 4 voisins. Cette relation de voisinage définit une structure de graphe non

orienté, que nous appelons **graphe facial**, et que nous utiliserons par la suite. Donc, dans le graphe facial, chaque nœud correspond à un tétraèdre, et chaque arête correspond à un triangle séparant deux tétraèdres.

Afin d'éviter toute confusion avec le graphe facial, nous appelons **graphe nodal** le graphe non orienté dont les  $n$  nœuds sont les sommets définis par les points de  $X_n$ , et dont les arêtes sont celles définies par les arêtes des tétraèdres de la triangulation. Par convention ces arêtes sont simples (une seule arête entre deux nœuds).



**Figure 2.2** : Cas le plus simple d'une triangulation de Delaunay dans le plan (avec les cercles circonscrits).

- **Triangulation de Delaunay** : Une triangulation de Delaunay d'un ensemble  $X_n$  de  $n$  points est une triangulation de  $X_n$  telle que l'intérieur de chaque  $d$ -sphère circonscrite aux  $d+1$  sommets de tout  $d$ -simplexe de la triangulation ne contienne aucun des  $n-d-1$  autres points de  $X_n$ .

Des exemples de triangulation de Delaunay dans le plan sont donnés figures 2.2 et 2.3. Les différences entre triangulations de Delaunay 2D et 3D sont rappelées dans le tableau 2.1.

Sauf cas particulier où plus de  $d+1$  points sont à la surface de l'une de ces  $d$ -sphères circonscrites, la triangulation de Delaunay est unique.

Lorsqu'on n'est pas dans le cas particulier précédent, on dit que les  $n$  points sont en *position générale*.

La triangulation de Delaunay est connue expérimentalement pour tendre à éviter les simplexes quasi-dégénérés, en particulier dans le cas où  $d=3$ , ainsi que dans le cas où  $d=2$  où l'on dispose d'un théorème : le plus petit angle de tous les triangles de la triangulation est maximisé [Sibson *et al.* 1978] (ce théorème n'a pas d'équivalent connu pour  $d>2$ ).

La triangulation de Delaunay a été introduite par le mathématicien Boris Delaunay (ou Delone), qui lui a donné son nom [Delaunay 1934].

	Triangulation 2D	Triangulation 3D
Elément de subdivision de l'enveloppe convexe		
Zone de séparation entre deux éléments		
cercle ou sphère circonscrit(e) à l'élément		

Table 2.1 : Comparaison entre les triangulations de Delaunay 2D et 3D.

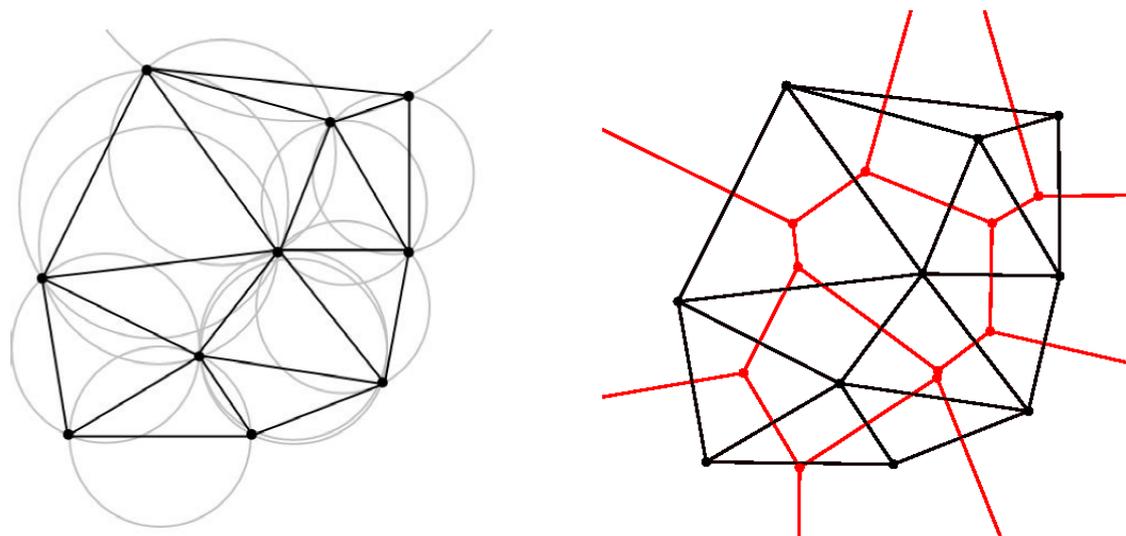
• **Diagramme de Voronoï** : La définition mathématique générale de ce diagramme, due au mathématicien Georges Voronoï [Voronoï 1908], est hors du cadre de ce mémoire, et donc nous en donnons une description simplifiée à partir de la triangulation de Delaunay dans le cas particulier d'un ensemble  $X_n$  de  $n$  points en position générale dans l'espace tridimensionnel.

Pour  $d=3$ , chaque point de  $X_n$  est dans une cellule polyédrique unique, convexe, éventuellement ouverte. L'union de ces cellules couvre tout l'espace. Lorsque deux points  $x$  et  $y$  de  $X_n$  sont reliés par une arête appartenant à au moins l'un des tétraèdres de la triangulation de Delaunay de  $X_n$ , l'intersection de la cellule contenant  $x$  avec la cellule contenant  $y$  est une région polygonale convexe, éventuellement ouverte, appartenant au plan médiateur du segment  $[x,y]$  : autrement dit, ces deux cellules ont une paroi commune. Les intersections des cellules sont vides dans les autres cas. Tout point intérieur à la cellule contenant  $x$  est plus proche de  $x$  que de tout autre point de  $X_n$ , et ceci est vrai quel que soit le point  $x$  choisi dans  $X_n$ .

La structure définie par le diagramme de Voronoï existe et est unique.

Exemple, dans le cas le plus simple on a  $n=4$ , la triangulation de Delaunay n'a qu'un seul tétraèdre, les 6 plans médiateurs des arêtes se coupent en un point unique qui est le centre de la sphère circonscrite au 4 points (de même que dans le plan, les 3 médiatrices d'un triangle se coupent en un point unique qui est le centre du cercle circonscrit au triangle). Ce point est à l'origine de 4 cellules polyédriques contenant chacune un sommet du tétraèdre, et les faces polygonales de ces cellules sont dans les plans médiateurs.

Pour un espace de dimension quelconque, le diagramme de Voronoï peut être défini comme précédemment, à partir du réseau des  $(d-1)$ -hyperplans médiateurs des arêtes des



**Figure 2.3** : Exemple de triangulation de Delaunay et de diagramme de Voronoï dans le plan. L'intérieur des cercles circonscrits aux triangles ne contient aucun sommet. Les parois des cellules du diagramme de Voronoï sont les traits en rouge. Les deux images sont extraites de [https://en.wikipedia.org/wiki/Delaunay\\_triangulation](https://en.wikipedia.org/wiki/Delaunay_triangulation)

d-simplexes issus de la triangulation de Delaunay.

Les structures de données respectives de la triangulation de Delaunay et du diagramme de Voronoï sont duales l'une de l'autre. Il existe diverses variantes de ces deux concepts [Edelsbrunner 1987].

Un exemple de diagramme de Voronoï dans le plan est donné figure 2.3.

- **$\alpha$ -shape** : Ce concept a été introduit en 1992 [Edelsbrunner & Mücke 1992]. L'énoncé de la définition mathématique complète étant très long et abstrait [Edelsbrunner & Mücke 1994], nous en résumons (en simplifiant) les points essentiels dans le cadre du calcul de canaux. L'ensemble  $X_n$  est celui des  $n$  points dont les coordonnées cartésiennes sont celles des centres des atomes lourds. Après avoir effectué la triangulation de Delaunay de  $X_n$ , on définit une sonde sphérique de rayon arbitraire  $\alpha$  censée représenter un ligand potentiel, et on retire du domaine convexe triangulé les tétraèdres dont la sphère circonscrite possède un rayon supérieur ou égal à  $\alpha$ . Ce qui reste du domaine définit alors la forme de la protéine, dite  $\alpha$ -shape. Le complément de l' $\alpha$ -shape à l'enveloppe convexe de  $X_n$  est considéré comme étant accessible au ligand sphérique, mais cela ne décrit pas des canaux.

Ce concept d' $\alpha$ -shape a donné lieu à de nombreuses variantes (voir une revue dans [Benkaidali *et al.* 2014]).

## 2.2 Etat de l'art : approches existantes, méthodologie et algorithmes.

Les méthodes basées sur des calculs d'énergie sont principalement des simulations de dynamique moléculaire. Elles ont été traitées section 1.7. Les techniques d'apprentissage et les méthodes basées sur des corrélations structure-activité ne sont pas considérées ici,

car elles ne fournissent pas d'information sur les canaux : voir par exemple l'étude QSAR de [Terfloth *et al.* 2007] portant sur 303 descripteurs structuraux pour classer les substrats des isoformes 3A4, 2D6, et 2C9. Il existe une revue récente sur la manière dont les canaux sont utilisés pour modifier la spécificité au substrat ou le degré de catalyse [Prokop *et al.* 2012]. Les premières tentatives d'utilisation des données sur les canaux pour concevoir de nouveaux médicaments sont en cours [Stsiapanava *et al.* 2014]. Tout comme l'arrimage (*docking*) d'une petite molécule comprend une recherche de positionnement et un calcul du score de ce positionnement, prédire un canal comprend une phase d'identification et une phase d'évaluation [Sousa *et al.* 2006]. Les méthodes de calcul de canaux sont souvent scindées en deux catégories : celles qui se basent uniquement sur la structure protéique, et celles qui incluent le ligand ou une molécule sonde. Dans la première catégorie, on peut trouver des algorithmes de calculs de chemins de coûts minimaux basés sur la structure de graphe associée au diagramme de Voronoï ou à sa structure duale, la triangulation de Delaunay [Edelsbrunner 1987], ces chemins reliant la surface de la protéine au site actif. Ces méthodes sont parfois appelées *méthodes de flux minimal* [Petřek *et al.* 2007, Yaffe *et al.* 2008a]. Dans la seconde catégorie, on utilise un ligand ou une sonde pour tester des trajets possibles d'entrée ou de sortie avec des techniques de type  $\alpha$ -shape ([Edelsbrunner & Mücke 1992, Edelsbrunner & Mücke 1994]), ou des simulations de dynamique moléculaire [Lüdemann *et al.* 2000a]. Nous référons ces méthodes comme étant à *ligand explicite*. La différence clé entre ces deux catégories de méthodes est que les méthodes de flux minimal identifient les canaux sans considérer le ligand, alors que les méthodes à ligand explicite évaluent directement un ou plusieurs canaux pertinents pour un ligand donné. Deux revues récentes résument les méthodes de calculs de canaux [Pérot *et al.* 2010, Benkaidali *et al.* 2014].

Nous ne considérons ci-dessous que les méthodes géométriques. Celles-ci sont avantageuses dans la mesure où leurs coûts calcul sont souvent minimales. De plus, beaucoup d'entre elles peuvent générer des informations sur les dimensions des canaux et sur les atomes qui les bordent.

Une première classe d'approches géométriques est les méthodes de type RPM (*Robot Planning Motion*), destinées à simuler les mouvements des molécules. Les méthodes RPM sont utilisées en robotique pour prévoir les mouvements réalisables de robots articulés dans un espace de travail encombré d'obstacles. Elles sont basées sur un échantillonnage Monte-Carlo de l'espace conformationnel, partant d'une conformation initiale. Un mouvement réalisable est effectué lorsqu'aucune collision n'est détectée lors d'un déplacement aléatoire, et ainsi on espère que le ligand atteigne le site actif. De tels outils ont effectivement été utilisés dans un contexte enzyme-ligand [Cortès *et al.* 2005, Kirillova *et al.* 2008]. Les techniques RPM restent inadéquates en terme de coût calcul pour retrouver rapidement le réseau complet des canaux recherchés.

La seconde classe est constituée d'approches géométriques destinées à exhiber les cavités et canaux avec des algorithmes déterministes, et ce sont de loin les approches plus populaires. La méthode développée dans le cadre de notre projet, et implémentée dans le logiciel CCCPP [Benkaidali *et al.* 2014], tombe dans cette catégorie.

Plusieurs algorithmes géométriques, basés soit sur des grilles rectangulaires soit sur la triangulation de Delaunay, ont été développés pour identifier des cavités ou canaux protéiques. Il y a deux principales familles d'algorithmes basées sur la triangulation de Delaunay : algorithmes à ligand explicite, ou algorithmes de flux minimal.

### 2.2.1 Méthodes géométriques à ligand explicite

Ce sont principalement les méthodes de type  $\alpha$ -shape et leurs variantes (voir section 2.1). Notamment, une variante améliorée tenant compte des sphères van der Waals des atomes protéiques a été proposée, et a donné lieu au logiciel CAST [Liang *et al.* 1998]. Mais la sonde sphérique de rayon  $\alpha$  reste supposée modéliser le ligand, alors que ce modèle est éloigné de la réalité pour les ligands des CYPs, ces ligands étant bien mieux modélisés par des cylindres [Benkaidali *et al.* 2014]. De plus, les méthodes de type  $\alpha$ -shape modélisent l'intérieur de la protéine plutôt que les canaux eux-mêmes.

### 2.2.2 Méthodes géométriques à flux minimal

Certaines méthodes, comme dans la première version de CAVER [Petřek *et al.* 2006], placent la protéine dans une grille rectangulaire de pas arbitraire (figure 2.4), et, en chaque point de la grille on calcule la sphère de rayon maximal n'ayant pas d'intersection avec les sphères van der Waals des atomes protéiques. Il existe une variante de la méthode avec grille, destinée à évaluer le volume d'une poche [Durrant *et al.* 2011]. D'autres méthodes, comme MOLE, destinée à pallier aux inconvénients de la grille définie dans CAVER [Petřek *et al.* 2007], utilisent la triangulation de Delaunay ou le diagramme de Voronoï [Yaffe *et al.* 2008a, Yaffe *et al.* 2008b, Chovancova *et al.* 2012]. Pour celles-ci, les atomes protéiques servent de guide pour définir une suite d'arêtes décrivant l'espace accessible dans la protéine, représentant des canaux (figure 5, [Kingsley & Lill 2015]).

Une partie des auteurs sont communs à MOLE et à CAVER, les deux méthodes étant semblables si l'on excepte la grille, qui n'est plus rectangulaire dans MOLE. Le point commun entre les deux méthodes réside dans la fonction de coût du passage en chaque nœud du réseau (hors des sphères van der Waals des atomes protéiques), qui est inversement proportionnel au carré du rayon de la plus grande sphère centrée sur le nœud et sans intersection avec les sphères van der Waals des atomes protéiques. Ce rayon est d'autre part augmenté, avant ou après élévation au carré, d'une quantité  $\epsilon$  arbitrairement petite destinée à éviter une instabilité numérique. L'utilisateur choisit alors un nœud extérieur au domaine protéique et un nœud au voisinage du site actif. Le coût associé à un chemin reliant ces deux nœuds est la somme des coûts associés à chaque nœud du chemin, et on calcule le chemin de coût minimal avec l'algorithme de Dijkstra [Dijkstra 1959]. Les chemins de coûts minimaux sont classés par coûts croissants et appelés "canaux", ce qui impose à l'utilisateur de fixer un nombre maximum de chemins s'il veut éviter d'en avoir un nombre trop élevé. Concrètement, plus les sphères centrées sur les nœuds du chemin auront de grands rayons (donc éloignées des sphères van der Waals des atomes protéiques), plus le coût du chemin sera faible, et mieux le canal sera classé. Les auteurs de MOLE n'ont pas trouvé de différences significatives entre les résultats donnés par CAVER et ceux donnés par MOLE, si ce n'est que MOLE est plus rapide. Les longueurs et largeurs des canaux sont les paramètres les plus communément utilisés pour évaluer ces canaux. Récemment est apparue une variante fournissant des données électrostatiques et d'hydrophobie sur les canaux, basée sur la localisation des acides aminés bordant ces canaux [Sehna *et al.* 2013].

Ni le ligand ni son accessibilité n'interviennent dans les calculs effectués par ces méthodes à flux minimal. L'hypothèse sous-jacente est que, plus les canaux sont courts et larges, plus ils ont de chances de favoriser la passage du ligand. Les interactions potentielles

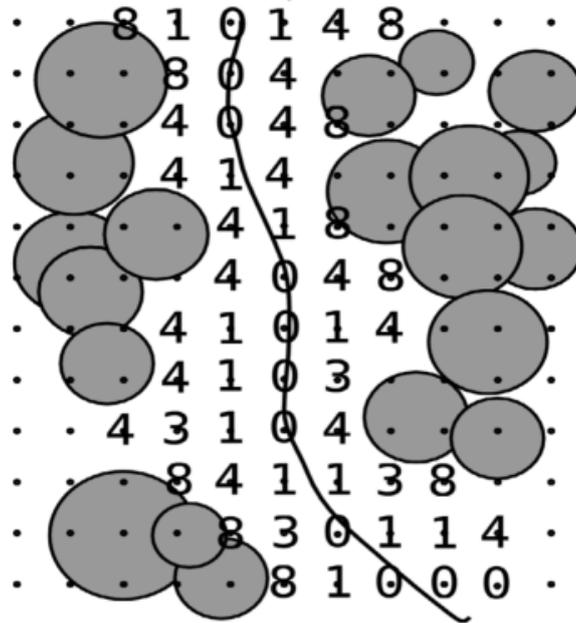


Figure 2.4 : Grille CAVER, selon [Petřek *et al.* 2006].

du ligand avec le canal ne sont pas explicitement évaluées. Les méthodes à flux minimal semblent actuellement être les plus utilisées pour identifier les canaux.

Bien qu'un grand nombre de méthodes aient été proposées, savoir quel canal est préféré pour l'entrée ou la sortie d'un ligand reste un défi majeur. Les défauts des méthodes à flux minimal se résument ainsi :

- On ne sait pas compter le nombre de canaux possibles, donc l'utilisateur doit en fixer le nombre maximum, alors qu'il préférerait qu'on lui dise combien il y en a.
- La fonction de coût a une expression arbitraire et complexe.
- Le ligand n'est pas considéré, alors que le nombre des canaux qui lui sont accessibles devrait dépendre de sa taille et de sa forme.
- Seule la protéine est prise en compte lorsqu'un chemin est déclaré optimal et classé comme étant un canal, alors que le rayon de la plus grande sphère passant dans le canal peut être petit (cas du goulot d'étranglement).
- Comme dans le cas de l'approche  $\alpha$ -shape, la géométrie des canaux s'interprète par rapport à des ligands sphériques, ou au mieux en termes de sphères englobant les ligands, ce qui n'est pas un modèle réaliste, et en général il est difficile de conclure si un ligand accède ou non au site actif.

Les méthodes géométriques ont le désavantage de ne pas considérer la flexibilité de la protéine, alors que l'importance de cette flexibilité est reconnue [Teague 2003]. L'utilisation d'ensembles de conformations protéiques est de plus en plus répandue pour prédire les canaux [Chovancova *et al.* 2012]. Malheureusement, l'utilisation de méthodes basées sur le diagramme de Voronoï produit souvent des canaux quasi-identiques pour de multiples structures. Sachant que jusqu'à une trentaine de canaux peuvent être trouvés par CAVER pour 5 conformations de la même protéine, il faut utiliser une procédure de *clustering* (classification). Il existe de nombreux critères de clustering, mais l'évaluation du nombre de classes reste polémique. Peu d'études ont considéré l'influence du clustering sur

la qualité de prédiction des canaux [Kingsley & Lill 2014c]. De plus, les méthodes basées sur des ensembles de conformations ne considèrent pas l'impact du ligand sur la flexibilité de la protéine, alors que cet impact est souvent critique pour les systèmes enzymatiques ayant des canaux [Liew *et al.* 2010].

A ce jour il n'y a pas d'étude systématique comparant les méthodes d'évaluation des canaux, et il n'y a pas de consensus sur le "meilleur" moyen de les évaluer quantitativement. Il a été rapporté que des approches de type RMD donnent des résultats en accord avec ceux issus de calculs par SMD [Sgrignani *et al.* 2012] et ceux obtenus par CAVER [Chovancova *et al.* 2012]. Mais des méthodes plus précises restent nécessaires pour identifier les changements conformationnels et quantifier le coût énergétique de la migration du ligand dans les canaux. En raison de ses défauts et malgré ce qui a été publié récemment, l'approche à flux minimal est pire que l'approche  $\alpha$ -shape.

Plusieurs verrous technologiques sont donc identifiés, et nos principaux défis sont :

- Calculer la totalité du réseau des canaux spécifique à un couple donné enzyme-ligand potentiel, et pas seulement à une enzyme donnée. A partir de ce challenge découlent les trois suivants.
- Il faut considérer les substrats non sphériques, mais les modèles non sphériques de ligands induisent deux difficultés : prouver mathématiquement que le test de passage d'un ligand au travers d'un triangle dans l'espace est défini sans ambiguïté, et produire un algorithme réalisant ce test.
- Prendre en compte la flexibilité du ligand via une procédure adéquate de clustering opérant dans l'espace conformationnel, c'est à dire qui ne soit pas basée sur un nombre prédéfini de conformères rigides (ex. 50 ou 100).
- Prendre en compte la flexibilité de l'enzyme par l'étude de plusieurs de ses structures, permettant un bon échantillonnage de l'espace conformationnel.

## 2.3 Caractéristiques originales du calcul des canaux dans CCCPP

CCCPP est un logiciel implémentant un algorithme géométrique de calcul de canaux présentant plusieurs caractéristiques originales :

- Contrairement aux méthodes usuelles, la taille et la forme du ligand interviennent dans le calcul des canaux qui lui sont stériquement accessibles.
- Contrairement au modèle sphérique rencontré partout dans la littérature, la forme du ligand est modélisée par le cylindre d'épaisseur minimale qui l'englobe. La pertinence de ce modèle a été montrée : voir table 2 dans [Benkaidali *et al.* 2014]. Cet article, que nous avons publié, est reproduit dans ce mémoire à la fin de ce chapitre.
- Connaissant la localisation du site actif, CCCPP répond à la question de savoir si oui non le ligand peut être déclaré capable d'accéder au site actif en partant de l'extérieur du domaine protéique. Ceci peut servir de filtre pour effectuer le criblage virtuel d'une chimiothèque, en éliminant les substrats potentiels stériquement improbables. Par contre, les composés pouvant stériquement accéder au site actif ne sont pas, à ce stade, prouvés être des substrats.
- Dans une seconde version, CCCPP propose en plus des chemins préférentiels pour accéder au site actif. Ces chemins sont caractérisés de manière précise, à l'aide des

atomes lourds délimitant ces chemins et donc des acides aminés qui les bordent (dans la littérature, seules les structures secondaires à la sortie des canaux sont mentionnées).

- Dans cette seconde version, les parois des canaux et des trajets préférentiels sont définies, et permettent des calculs de volumes et de surfaces.

Le source du logiciel CCCPP (en fortran 77) a été produit par le Dr. Michel Petitjean. L'effet de la flexibilité conformationnelle des ligands sur leur forme a été pris en le compte dans des études préliminaires [Meslamani *et al.* 2009, Benkaidali *et al.* 2012], et sera examiné section 2.4. Les données sur chemins proposés par CCCPP peuvent à leur tour être utilisées en entrée des méthodes basées sur des calculs d'énergie, telles que les simulations de dynamique moléculaire, dont on sait que l'étape d'initialisation est cruciale.

### 2.3.1 Modélisation géométrique des protéines avec CCCPP

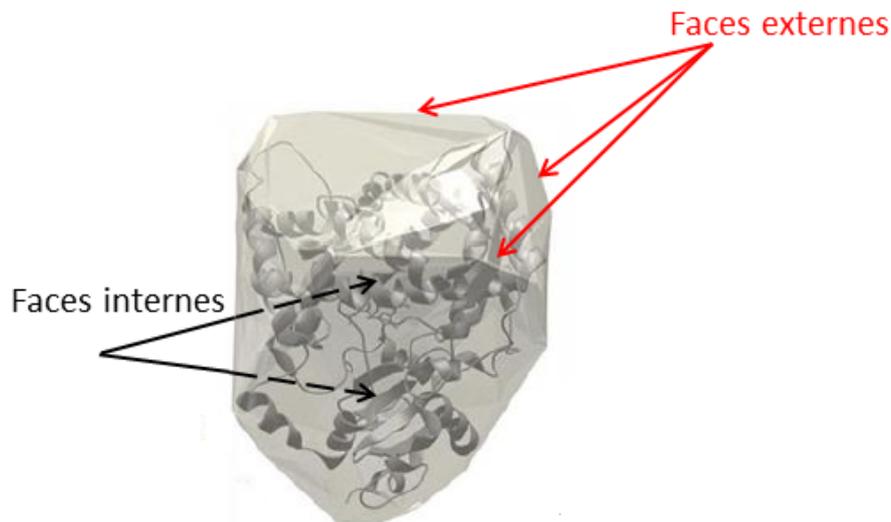
#### Première version de CCCPP : définition du canal

Cette première version était conçue à l'origine en grande partie pour servir de filtre stérique en vue du criblage virtuel de chimiothèques. Cette approche repose sur l'hypothèse que le site actif enfoui peut accueillir plusieurs types de substrats et peut être accédé de l'extérieur de la protéine via des canaux. Pour simplifier notre présentation, nous considérons d'abord un modèle rigide pour la protéine et le ligand.

Nous soulignons aussi que la terminologie habituelle traitant des vides à l'intérieur des protéines ne fait pas encore consensus : canaux, cavités, pores, poches, etc. Ici nous appelons des canaux les concavités reliant l'extérieur de la protéine à son site actif enfoui. Dans le cas d'une protéine avec un site actif à sa surface, nous dirons que la concavité est une poche, tandis que des concavités superficielles sans aucun site actif sont aussi souvent appelées des poches. Une concavité dans la protéine et reliant son extérieur à deux endroits peut être appelée un pore, sans tenir compte de l'existence ou non d'un site actif. Nous insistons sur le fait que ces définitions intuitives sont présentées pour des raisons de clarté, mais ne sont pas mathématiquement rigoureuses. En revanche, dans le cadre de notre travail, nos objets structuraux sont rigoureusement définis, et sont traités avec des outils basés sur la théorie des graphes (voir section 2.1).

- Triangulation du domaine protéique

Dans un premier temps, le domaine protéique est assimilé à l'enveloppe convexe des atomes lourds de la protéine. Ce domaine est partitionné en un ensemble de tétraèdres adjacents, c'est à dire que deux tétraèdres voisins partagent une face triangulaire. Dans le cas des CYPs, on obtient environ 25000 tétraèdres et 50000 faces triangulaires. Cette opération est faite dans CCCPP à l'aide de la triangulation de Delaunay (voir section 2.1), par la méthode du paraboloïde [Edelsbrunner 1987]. Il faut noter que la plupart des calculs récents de cavités s'appuient sur la triangulation de Delaunay. A cette étape du calcul, il n'y a ni cavités ni canaux : le domaine protéique est convexe, limité par un polyèdre à faces triangulaires (ce sont les faces des tétraèdres situés au bord du domaine). Par contre, la triangulation de Delaunay induit une structure appelée graphe facial (voir section 2.1). Ce graphe est tel que chaque tétraèdre est un nœud du graphe, et tel que chaque face triangulaire séparant 2 tétraèdres adjacents soit une arête du graphe. Une face triangulaire au bord du polyèdre englobant la protéine n'appartient qu'à un seul



**Figure 2.5** : Faces externes et internes de l'enveloppe convexe du CYP3A4 triangulé.

tétraèdre. Une telle face est qualifiée de **face externe**, et un tel tétraèdre est qualifié de **tétraèdre externe**.

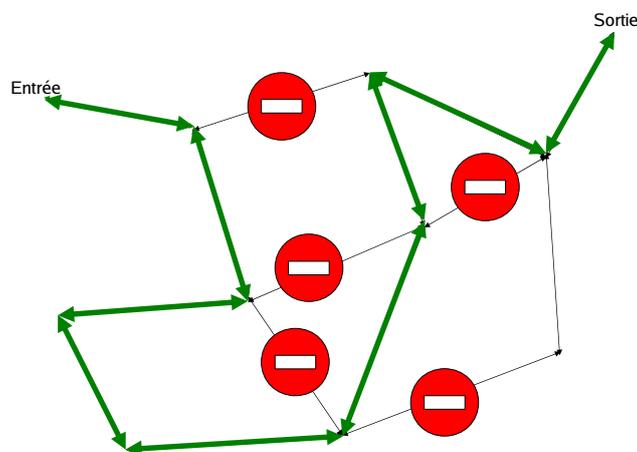
Par commodité, à chaque face externe on ajoute un nœud nommé **nœud extérieur**, et une arête reliant ce nœud supplémentaire au tétraèdre externe présentant cette face externe. Ces nœuds extérieurs, qui contrairement aux autres nœuds ne sont pas des tétraèdres, sont par définition tous accessibles au ligand puisqu'ils représentent l'extérieur de la protéine.

CCCPP doit ensuite calculer les canaux dans les protéines sous forme d'un réseau de tétraèdres accessibles au ligand.

- Définition et calcul des canaux

Pour savoir si un tétraèdre est accessible, on regarde si le ligand peut y entrer par l'une de ses faces triangulaires. Le test d'accessibilité s'effectue en deux étapes : on modélise la forme géométrique du ligand (voir plus loin), puis CCCPP teste si le passage au travers la face triangulaire est possible (voir plus loin les tests de passage). Lorsque le test de passage d'une face triangulaire échoue, c'est à dire que le ligand ne passe pas par la "porte" triangulaire, l'arête du graphe facial associée à ce triangle (externe ou non), est supprimée. Une telle arête est dite "infranchissable". Les autres sont dites "franchissables". Il suffit donc d'effectuer une boucle sur la liste des faces triangulaires pour mettre à jour le graphe facial : pas besoin d'énumérations combinatoires ni d'échantillonnage par Monte-Carlo. Par extension de langage, ce qui reste du graphe facial après élimination des arêtes infranchissables sera encore appelé graphe facial.

CCCPP demande ensuite à l'utilisateur de choisir un atome cible au niveau du site actif. Pour les CYPs, le choix de l'atome de fer s'impose. Tous les tétraèdres ayant cet atome pour sommet sont des nœuds du graphe facial voisins du site actif. CCCPP compte dans le graphe facial (celui qui reste après élimination des arêtes infranchissables), les chemins (suites d'arêtes franchissables), reliant chaque nœud voisin du site actif à au moins un nœud extérieur. Si au moins un de ces chemins existe, CCCPP déclare que le ligand est stériquement capable d'accéder au site actif (voir figure 2.6).



**Figure 2.6** : Recherche d'un chemin dans le graphe facial. Les arêtes franchissables sont en vert, et les arêtes infranchissables sont indiquées par des sens interdits.

Après élimination des arêtes infranchissables, le graphe facial n'est plus obligatoirement connexe : il peut se décomposer en plusieurs composantes connexes (voir section 2.1). Ces composantes connexes permettent de préciser les définitions de plusieurs concepts, dont les canaux.

**Cavité** : composante connexe n'incluant aucun nœud extérieur.

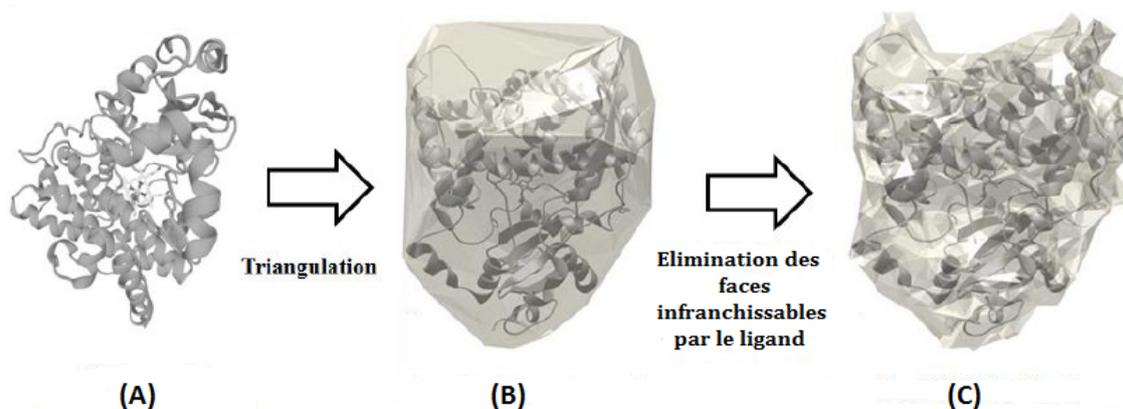
**Poche** : composante connexe incluant au moins un nœud extérieur. Cette définition peut masquer des situations plus complexes, telles que les pores, pour lesquelles le mot "poche" n'est pas intuitivement approprié.

**Canal** : poche incluant au moins un nœud (tétraèdre) voisin du site actif. Cette définition n'a cependant de sens pratique que pour un site enfoui. De plus, l'enfouissement lui-même n'est pas défini.

Il peut donc y avoir plusieurs canaux, et par conséquent on peut parler d'un "réseau" de canaux.

Le nombre de canaux de ce réseau est calculé, contrairement aux approches de type CAVER et MOLE dans lesquelles l'utilisateur en fixe implicitement le nombre (voir section 2.2.2). Il est aussi rappelé que les approches de type CAVER et MOLE ignorent le ligand, contrairement à CCCPP et aux approches de type  $\alpha$ -shape. La dépendance au ligand ne devrait pas être choquante : exemple, l'espace disponible dans la protéine pour une petite molécule comme l'eau ne devrait pas être identique à l'espace disponible pour un ligand volumineux comme la cyclosporine ou érythromycine.

Une différence fondamentale apparaît entre l'approche CCCPP et les approches de type  $\alpha$ -shape. En effet, ces dernières modélisent uniquement la forme et la structure géométrique de la protéine, alors que CCCPP modélise prioritairement les "vides" (dont les canaux) constituant le complément de la zone protéique à son enveloppe convexe (figure 2.7). Une autre différence est liée au modèle géométrique du ligand, sphérique pour les



**Figure 2.7** : Modélisation géométrique par CCCPP du CYP3A4 (code PDB : 1TQN). (A) Structure du CYP3A4. (B) Enveloppe convexe du CYP3A4 triangulé. (C) Réseau des canaux du CYP3A4.

$\alpha$ -shapes, cylindrique pour CCCPP, même si CCCPP prévoit que l'utilisateur peut choisir optionnellement le modèle sphérique.

- Test de passage

Le test de passage d'un objet dans un triangle est l'étape la plus délicate, car cela nécessite une définition rigoureuse. Même dans un cas simple comme celui d'un objet convexe, nous ne disposons pas d'une telle définition, et si l'on en avait une, il resterait à trouver un algorithme qui décide si le passage est possible ou non.

Il a été montré sur une famille de 70 substrats du CYP3A4 que la forme géométrique du ligand était beaucoup mieux modélisée par un cylindre englobant d'épaisseur (hauteur) minimale plutôt que par une sphère englobante minimale : voir table 2 dans [Benkaidali *et al.* 2014] (cet article est reproduit section 4). Très grossièrement, cela revient à considérer que les ligands sont majoritairement plutôt plats (cylindre englobant de hauteur minimale), mais moins souvent globulaires (sphère minimale englobante) ou allongés (cylindre englobant de rayon minimal).

Il reste à s'assurer qu'un cylindre d'épaisseur connue "passe" dans un triangle donné. L'hypothèse simplificatrice suivante est émise : on considère que la hauteur du cylindre englobant (sous-entendu de hauteur minimale) est le paramètre critique pour le test de passage. Ce paramètre est appelé **CV** (*Critical Value*) dans CCCPP. En d'autres termes, on assimile le ligand à un objet occupant tout l'espace compris entre deux plaques parallèles infinies séparées par une distance égale à CV. Si l'objet peut passer entre deux des trois sommets du triangle sans collision avec le troisième sommet (il y a une solution simple à ce problème de géométrie), le passage est réussi. Par contre, ce test de passage ne tient pas compte des collisions avec d'autres sommets que celui du triangle.

Concernant le ligand, le calcul de son épaisseur minimale, du cylindre de hauteur minimale englobant, du cylindre de rayon minimal englobant ([Petitjean 2012]), de la sphère minimale englobante, du plus petit polyèdre englobant, ainsi que de divers autres paramètres géométriques, sont effectués par le logiciel RADI, en libre accès sur la même page web que CCCPP : <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>.

Il existe une valeur de  $CV$  en-dessous de laquelle au moins un canal existe, et au-dessus de laquelle il n'y a plus de canal : cette valeur est appelé valeur critique limite, et est notée  $CV_{lim}$ . Un ligand d'épaisseur supérieure à  $CV_{lim}$  ne peut pas accéder au site actif en raison des contraintes stériques.

Dans notre modèle, on ignore l'encombrement stérique dû aux liaisons entre atomes protéiques et l'encombrement stérique dû aux sphères atomiques, sans oublier les hydrogènes, qui sont normalement absents des fichiers PDB. Déclarer que le ligand est stériquement capable de passer est un point de vue optimiste. D'un autre côté, la protéine est déformable, tant au niveau du squelette que des chaînes latérales des acides aminés. De cet autre point de vue, déclarer que le ligand est stériquement incapable de passer est un point de vue pessimiste. Une approche globale simplifiée ignore ces deux facteurs contradictoires, dont on espère qu'ils se compensent.

A ce stade, CCCPP est supposé produire un réseau de canaux sous l'hypothèse du modèle rigide du ligand. En première approximation, on espère que ce modèle marche pour les ligands non flexibles. Le cas des ligands flexibles est traité section 2.4.

Malgré les nombreuses hypothèses simplificatrices, le test de passage effectué dans CCCPP est le seul qui a permis de rendre compte de la capacité d'accès au site actif du CYP3A4 d'un substrat volumineux tel que la cyclosporine.

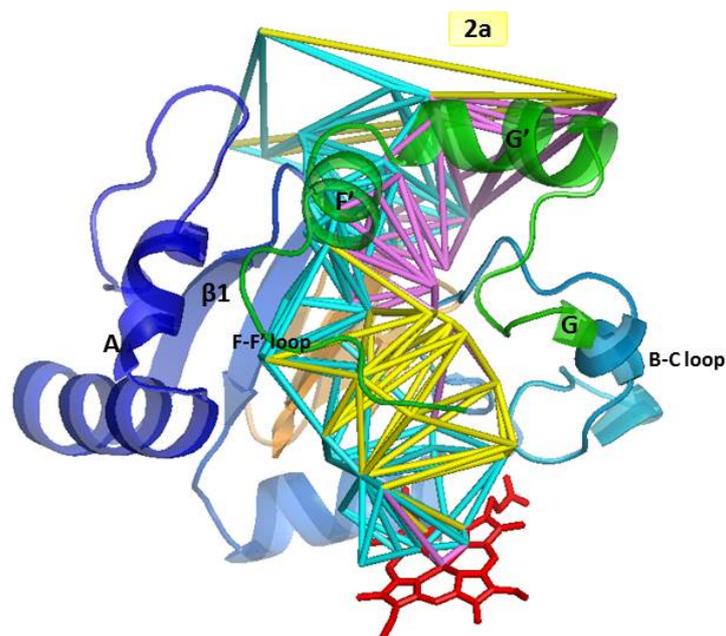
## Seconde version de CCCPP : chemins de coûts minimaux

Conclure à l'existence ou non de canaux d'accès au site actif ne suffit pas. Il faut en plus caractériser ces canaux, et proposer des trajets préférentiels potentiels pour le ligand à l'intérieur des canaux, dont la forme géométrique, dans le cas des CYPs, s'apparente à une sorte d'entonnoir. C'est ce qui est fait dans la seconde version de CCCPP.

Les canaux sont des composantes connexes du graphe facial (après élimination des arêtes infranchissables). La liste des tétraèdres composant chaque canal, ainsi que la liste des triangles infranchissables sont calculées. Ceci permet de calculer le volume du canal (somme des volumes des tétraèdres du canal), et la surface de sa paroi (somme des surfaces des triangles infranchissables).

Pour un canal donné, choisi par l'utilisateur, CCCPP calcule des chemins de coûts minimaux appelés MCP (voir section 2.1), qui peuvent être interprétés par l'utilisateur comme des trajets préférentiels pour le ligand. Le coût total d'un chemin est la somme des coûts de chacune des arêtes qui le composent. Le coût associé à une arête est le rapport  $CV/CV_{max}$ , où  $CV$  est la valeur critique pour le ligand courant (entrée par l'utilisateur), et  $CV_{max}$  est l'épaisseur maximale permise pour qu'un ligand quelconque franchisse l'arête (autrement dit la face triangulaire). Ce coût par arête est donc compris entre 0 et 1. Pour chaque tétraèdre (nœud du graphe facial) voisin du site actif, CCCPP calcule le chemin de coût minimal (MCP) le reliant à un nœud extérieur, avec l'algorithme de Dijkstra [Dijkstra 1959].

Dans le cas des CYPs, les MCPs (lorsqu'ils existent) pour chaque tétraèdre voisin du site actif sont quasiment identiques (tronc commun), et ne divergent qu'au voisinage du site actif. On peut donc considérer que ces MCPs sont redondants, et qu'ils correspondent à un seul trajet préférentiel potentiel pour le ligand. CCCPP retire alors du graphe facial les arêtes du MCP de plus petit coût, et effectue une nouvelle passe de recherche de MCPs. Si la passe est réussie, c'est à dire qu'au moins un MCP est trouvé, CCCPP effectue une



**Figure 2.8** : Représentation du trajet 2a à l'aide de trois MCPs. Les tétraèdres des MCPs sont en jaune, turquoise et mauve. Les MCPs ont des tétraèdres communs.

nouvelle passe, jusqu'à ce que plus aucun MCP ne soit trouvé. Il y aura donc, au plan stérique, autant de propositions de trajets préférentiels que de passes réussies.

Comme pour le canal entier, le volume et la surface de la paroi de chaque MCP sont calculés. De plus, pour chaque MCP, CCCPP fournit la liste des tétraèdres et des faces triangulaires, et donc la liste des atomes bordant le MCP, et par conséquent la liste des acides aminés bordant le MCP. On peut en déduire quelles structures secondaires bordent les canaux. Ces informations sont inaccessibles dans la littérature, où les approches de type CAVER et MOLE ont seulement permis d'établir manuellement quelles étaient les structures secondaires en sortie des canaux [Cojocaru *et al.* 2007], et c'est malgré tout cette assignation manuelle qui est à la base de la nomenclature des canaux couramment utilisée encore aujourd'hui. Au niveau terminologique, remarquons que les canaux proposés par CAVER et MOLE, qui sont aussi des chemins de coûts minimaux (avec un critère de coût différent de celui de CCCPP, et dont aucune paroi n'est définie), s'apparentent plus aux trajets préférentiels proposés par CCCPP, plutôt qu'aux canaux calculés par CCCPP. Il resterait à examiner si les canaux de la littérature peuvent être représentés par des MCPs (voir exemple du canal 2a, figure 2.8).

Un autre avantage de la seconde version de CCCPP est de permettre la comparaison des canaux et des MCPs de différentes structures protéiques, grâce aux structures de données détaillées qui les décrivent. Les deux versions de CCCPP sont disponibles à la page <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>.

### 2.3.2 Les deux modes de visualisations des cavités

Les cavités calculées par CCCPP sont les vides dans la protéine. Ces vides représentent les passages possibles des ligands et qui peuvent être visualisés de deux manières : soit à l'aide du graphe facial, soit à l'aide du graphe nodal (voir section 2.1).

La visualisation à l'aide du graphe facial consiste à attribuer arbitrairement à chaque nœud de ce graphe un point dont les coordonnées sont celles du barycentre du tétraèdre associé à ce nœud, puis à tracer les arêtes du graphe. En pratique, il suffit de générer une molécule virtuelle dont les atomes ont pour coordonnées celles des nœuds du graphe facial, et dont les liaisons chimiques sont les arêtes de ce graphe. N'importe quel utilitaire de visualisation de molécules pourra ainsi afficher les canaux.

La visualisation à l'aide du graphe nodal s'applique principalement aux MCPs dans la seconde version de CCCPP. On génère une molécule virtuelle dont les atomes sont précisément ceux des sommets des tétraèdres composant un MCP donné, et dont les arêtes sont les côtés de ces tétraèdres. Si les atomes sont bien réels (ils sont parmi ceux de la protéine), les arêtes visualisées sont virtuelles, et ne représentent en aucun cas des liaisons chimiques au sein de la protéine. A nouveau, un utilitaire de visualisation de molécules permet d'afficher les MCPs.

Pour les deux modes de visualisation, il faut se rappeler que ce qui s'affiche à l'écran indique des vides dans la protéine, et non la protéine elle-même, ce qui n'empêche pas, par exemple, de superposer graphiquement les canaux avec les structures secondaires de la protéine. Certaines des figures de ce manuscrit sont basées sur une combinaison des deux modes de visualisation avec des plans de coupure appropriés, parfois associés avec l'hème et le ligand. Toutes les figures de ce manuscrit affichant des MCPs ou des canaux calculés par CCCPP ont été générées avec PyMOL<sup>TM</sup> (<https://www.pymol.org/>).

## 2.4 Echantillonnage conformationnel des ligands

La flexibilité du ligand est prise en compte en générant un nombre calculé de conformères dits "représentatifs" des conformations possibles du ligand. Celles-ci sont en nombre infini, mais on émet l'hypothèse que seules quelques-unes bien choisies suffisent pour la plupart des applications, dont la nôtre. Les approches classiques qui consistent à générer systématiquement, par exemple 50 ou 100 conformères, sont irréalistes. Par exemple, on admet que le cyclohexane adopte deux conformations préférentielles uniquement : la forme chaise et la forme bateau. Le nombre de conformères représentatifs n'est pas défini, ni même ce que signifie "représentatif".

L'approche de [Meslamani *et al.* 2009, Benkaidali *et al.* 2012] s'effectue en deux étapes. D'abord on utilise un générateur aléatoire de structures tel que Frog [Bohme *et al.* 2007] ou Frog2 [Miteva *et al.* 2010] pour obtenir un excès de conformères. Ensuite, on effectue une classification ascendante hiérarchique incluant un critère d'arrêt automatique, ce qui a pour effet de fixer le nombre de classes de conformères et la composition de ces classes. Dans chaque classe, on sait définir un "conformère moyen", et c'est ce dernier qui sera retenu comme représentant de sa classe [Meslamani *et al.* 2009]. Le ligand sera déclaré capable d'atteindre le site actif si au moins l'un de ses conformères représentatifs est capable d'atteindre le site actif.

## 2.5 Préparation des données

Toutes les structures P450 étudiées ont été déterminées à partir de protéines modifiées où les hélices transmembranaires N-terminales ont été tronquées. Les fichiers structuraux des CYPs présents dans la base PDB ne contiennent généralement pas les atomes de la partie N-terminale transmembranaire, car les CYPs recombinant sont tronqués d'environ 30 acides aminés pour la cristallisation. En dehors de la partie N-terminale, certaines structures de CYPs, dont celles du 3A4, sont incomplètes. Dans certains cas, des mutations ont été introduites pour augmenter la solubilité.

Certaines boucles inter-hélices trop flexibles sont non résolues (zones de densité floue), et la position des acides aminés N- et C-terminaux (présence ou non d'un marquage par l'histidine) peut aussi varier selon le cristal. Dans toutes les structures des CYPs, ce sont les deux boucles liées à l'hélice H, les boucles GH et HI, très flexibles, qui ne sont pas cristallisées. Ces deux boucles ne sont pas dans la zone des canaux 2 et ne bordent aucun canal important déterminé par CCCPP pour le passage des ligands dans le 3A4.

En revanche, sur les 18 structures que nous avons étudiées, quatre d'entre elles (3TJS, 4K9X, 2J0D et 4K9U) ont des acides aminés manquants sur la boucle FF', dont une, 4K9X, a quelques acides aminés manquants en plus dans la boucle F (table 2.2).

Pour plus de précisions dans notre étude, notamment sur les structures ouvertes, nous avons reconstruit les acides aminés manquants. Ceci modifie à priori la triangulation, mais n'a pas de conséquences sur l'identification des canaux car les tétraèdres absents sont situés en périphérie de la protéine (boucles flexibles) [Benkaidali *et al.* 2014].

Nous sommes dépendants de la fiabilité de la reconstruction. Les analyses des résultats de calculs sur les structures reconstruites seront faites avec réserve quant aux positions et aux conformations des acides aminés restaurés au niveau de la boucle FF'.

Nous avons reconstruit les structures en deux étapes :

1. Ajout des acides aminés manquants avec le logiciel en ligne Swiss Model (<http://swissmodel.expasy.org/>).
2. Raffinement des boucles avec le logiciel en ligne Modeller (<https://salilab.org/modeller/>).

La reconstruction des protéines n'inclut pas les ligands et les molécules d'eau. Il est nécessaire de réintroduire ces molécules dans le fichier reconstruit pour la suite du traitement.

Les coordonnées des structures cristallisées ne sont pas dans un même référentiel. Afin de comparer les structures et les canaux calculés dans un référentiel commun, nous prenons comme référence celui de la structure 1TQN. Les alignements ont été faits sur le squelette à l'aide de la méthode des moindres carrés, implémentée dans le logiciel PyMOL<sup>TM</sup>. Cette méthode, couramment utilisée et implémentée dans de nombreux logiciels, consiste à minimiser la distance quadratique moyenne, généralement notée **RMSD** (*Root Mean Square Deviation*), entre deux groupes de N points dont on connaît la correspondance par paires, en calculant la translation et la rotation optimale qui minimisent cette RMSD, ou, ce qui revient au même, qui minimisent la somme des carrés des N distances entre ces deux groupes de points. La solution de ce problème de superposition optimale est connue analytiquement (voir appendice dans [Petitjean 1999], et appendice A dans [Petitjean 2002] pour la solution du problème généralisé). Un

alignement sur le squelette plus l'hème a donné les mêmes valeurs de RMSD. Nous avons alors retenu l'alignement le plus simple, celui effectué avec le squelette uniquement.

Code PDB	Structure secondaire contenant les acides aminés manquants					
	Boucle GH		Boucle HI		Boucle FF'	
	Nombre	Séquence	Nombre	Séquence	Nombre	Séquence
1TQN	-	-	4	282–285	-	-
1W0E	9	262–270	13	278–290	-	-
1W0F	7	262–268	8	281–288	-	-
1W0G	7	263–269	12	277–288	-	-
3UA1	9	259–268	7	280–286	-	-
4K9V	4	265–268	9	280–288	-	-
4I3Q	-	-	5	282–286	-	-
4NY4	5	266–270	6	280–285	-	-
<b>3TJS</b>	4	263–266	8	281–288	7	212–218
3NXU	3	265–267	8	281–288	-	-
4I4G	3	265–267	8	281–288	-	-
4I4H	3	265–267	8	281–288	-	-
4K9W	3	265–267	7	281–287	-	-
<b>4K9X</b>	3	265–267	8	281–288	(6)	(200–205 hélice F)
					2	212–213
<b>2J0D*</b>	10	261–270	5	282–286	5	214–218
2V0M	5	264–268	9	280–288	-	-
4K9T	5	265–269	9	280–288	-	-
<b>4K9U</b>	6	264–269	8	281–288	6	210–215

**Table 2.2 :** Acides aminés manquants dans les structures secondaires (rayons X) du 3A. Les structures marquées en rouge sont celles présentant des RM sur la région FG.

\*boucle EF : 196–197 (2 RM).

Remarque : dans plusieurs structures, il y a des acides aminés loin de la bouche du canal 2a appartenant à la boucle HI reconstruite (281–285)

## 3 RESULTATS ET DISCUSSION

L'importance du CYP3A4 et sa complexité mécanistique et structurale sont rappelés section 1.10.3. Nous présentons ci-dessous une analyse et une interprétation biologique du réseau des canaux d'accès au site actif calculé par CCCPP pour le 3A4. Il s'agit de passer d'une liste des chemins présentée sous forme d'une suite de tétraèdres et de faces franchies, en un résultat interprétable pour les biologistes. Il s'agit aussi de prendre en compte les interactions moléculaires avec le ligand le long des canaux potentiels, afin de prédire si les trajectoires au sein des canaux sont énergétiquement favorables ou non. Nous comparerons les résultats obtenus sur les structures du CYP3A4 avec et sans ligand. La mesure des surfaces et volumes de la composante connexe observée pour les substrats à différents CV permet de quantifier le degré de malléabilité du CYP3A4 et donc sa capacité à se déformer pour reconnaître et métaboliser des xénobiotiques différents en taille et en structure chimique, ce qui est un sujet récurrent dans la littérature sur les CYPs, en particulier sur l'isoforme 3A4 qui est le paradigme de la multispécificité chez les CYPs.

### 3.1 Génération des canaux et des chemins de coûts minimaux du CYP3A4

Afin d'analyser le comportement du 3A4 vis-à-vis des différentes molécules en interaction, nous avons utilisé CCCPP sur les 18 structures du 3A4 trouvées en 2013 dans la base PDB (3 apo et 15 holo), puis nous avons calculé les graphes de ces structures, les canaux et les chemins. Les étapes et les résultats obtenus avec CCCPP sur ces structures sont présentés dans cette section. Nous relierons les résultats obtenus par CCCPP au comportement biologique du 3A4, et à l'analyse de la sélectivité lors de la prise en charge de molécules.

Du fait de la déformabilité du 3A4, les structures étudiées adoptent différentes formes selon leur état enzymatique. La triangulation et l'accessibilité du substrat (ligand virtuel utilisé pour le test de passage) sont sensibles à la position des atomes. La liste des tétraèdres résultants (accessibles et inaccessibles) est forcément différente, même pour des structures très semblables structurellement. La triangulation (dépendante de la structure) est donc fonction de l'état de l'enzyme, et les résultats obtenus avec CCCPP permettent d'interpréter la réponse biologique de l'enzyme.

L'analyse des structures du 3A4 nécessite 7 étapes :

1. Préparation des données.
2. Exécution de CCCPP.
3. Génération du graphe de l'enveloppe convexe.
4. Génération du réseau complet des canaux.
5. Extraction et marquage des composantes connexes.
6. Inventaire des chemins d'accès au site actif.
7. Interprétation biologique de ces chemins.

Les 6 premières étapes sont décrites tout au long de cette section 3.1. La dernière étape est présentée dans les sections suivantes, pour différents ligands co-cristallisés.

### 3.1.1 Préparation des données

Cette étape consiste à modifier le fichier original extrait de la base PDB, afin que les résultats calculés par CCCPP soient exploitables.

#### Reconstruction des acides aminés manquants

Cette étape a été nécessaire sur les 18 structures du 3A4, plus particulièrement sur les 4 structures présentant des acides aminés manquants dans la région FG, reconnue pour son rôle fonctionnel (voir section 2.5 et données table 2.2). Certaines boucles non résolues peuvent être impliquées dans le contrôle d'entrée des substrats. Il est donc indispensable de les reconstruire, au travers d'une modélisation prédictive.

Néanmoins, les deux boucles GH et HI sont très flexibles et non résolues pour pratiquement toutes les structures. Ces deux boucles ne bordent aucun des canaux défini dans la nomenclature de [Cojocar *et al.* 2007] et n'ont pas été relevées dans la littérature comme ayant un rôle important dans les CYPs. Afin de préserver l'originalité des structures cristallisées, nous avons pris le choix d'ignorer la reconstruction de ces boucles. Ces deux boucles ont été uniquement reconstruites dans les structures contenant des acides aminés manquants dans d'autres régions de la protéine.

#### Préparation du fichier PDB

Pour la triangulation, CCCPP prend en compte les atomes de la chaîne protéique, en repérant les lignes "ATOM" dans le fichier PDB. Les lignes marquées "HETATM" sont ignorées par CCCPP, ce qui élimine les molécules d'eau et de ligand. Garder les molécules d'eau aurait faussé le calcul des canaux, car ces molécules d'eau sont déplaçables. Afin de tenir compte de l'hème dans la triangulation, les 43 atomes lourds de l'hème sont renommés "ATOM". Le nombre d'atomes lus par CCCPP (ATOM + HETATM de l'hème) pour la structure apo du 3A4 passe de 4000 atomes lourds à 3844 dans le cas de la structure 1TQN.

#### Alignement dans un référentiel commun

Les structures ne sont pas dans un même référentiel. Lors de l'analyse, afin de comparer toutes les structures et leurs graphes dans un même référentiel, nous prenons comme référentiel commun celui de 1TQN. L'alignement spatial se fait sur le squelette. L'algorithme d'alignement est décrit à la fin de la section 2.5. Un alignement sur le squelette plus l'hème a donné les mêmes valeurs de RMSD. Nous avons alors retenu l'alignement le plus simple, celui avec le squelette uniquement. Les alignements ont été effectués avec PyMOL<sup>TM</sup> (<https://www.pymol.org/>).

#### Arrivée dans le site actif

Rappelons qu'un trajet est un ensemble de MCPs (voir section 2.1) possédant un tronc commun, et permettant à un ligand de traverser la protéine depuis l'extérieur jusqu'au site actif et inversement (voir à la fin de la section 2.3.1). Le départ d'un trajet depuis l'extérieur se fait par un tétraèdre externe, qui est un tétraèdre ayant au moins une face externe (voir figure 2.5).

Une condition nécessaire pour considérer que l'arrivée se fasse bien dans le site actif, est que l'un des MCPs se termine par un tétraèdre dont l'un des sommets (atome) appartient à l'hème. Cibler uniquement l'atome de Fer étant trop restrictif, nous considérons les autres atomes de l'hème. Les atomes extrêmes de l'hème sont trop éloignés du site actif. Nous considérons alors que, parmi les atomes de l'hème, les atomes voisins du Fer (c'est à dire liés chimiquement au fer), et les atomes voisins suivants constituent un choix acceptable pour définir l'arrivée au site actif. Ceci correspond à un total de 13 atomes hémiques. Donc, tout tétraèdre ayant au moins l'un de ces 13 atomes pour sommet est considéré comme étant un tétraèdre d'arrivée. Se limiter aux premiers voisins du Fer aurait été insuffisant, alors que l'inclusion de la troisième couche de voisins aurait été excessive. Ce choix de deux couches d'atomes voisins (appelés environnements *concentriques*) est considéré comme optimal dans de nombreuses autres situations [Dubois *et al.* 1987, Attias & Petitjean 1993].

### 3.1.2 Exécution de CCCPP

Une exécution de CCCPP sur un CYP (environ 3800 atomes lourds), prend environ 4 minutes. Lorsque toutes les options d'écriture des résultats sont activées, le fichier de résultats (hors visualisation) contient typiquement dans les 150000 lignes : l'exploitation des résultats nécessite parfois l'écriture d'un ou plusieurs scripts Unix.

Lors du calcul des canaux, la suppression des arêtes qui suit la triangulation de l'enveloppe convexe CCCPP génère des noeuds isolés, c'est à dire des tétraèdres connectés à aucun tétraèdre voisin (le ligand ne peut y entrer par aucune de ses 4 faces). CCCPP supprimant uniquement les arêtes, les noeuds isolés demeurent dans le fichier d'écriture des résultats. On obtient de l'ordre de 17000 tétraèdres isolés sur environ 25000. Ils sont supprimés avant l'étape de visualisation, afin de ne pas surcharger inutilement les images.

Nous calculons d'abord les canaux d'une structure donnée à  $CV_{lim}$  (ce paramètre a été défini avec les tests de passage à la fin de la section 2.3.1). Pour le 3A4, la valeur obtenue de  $CV_{lim}$  la plus grande est de 8.75Å pour la structure 2J0D, et en moyenne d'environ 6Å sur l'ensemble des structures. Le  $CV_{min}$  est constant (à 0.25Å près) à CV 5.25Å. Nous relevons alors les canaux et les MCPs. Pour toutes les structures étudiées du 3A4, on constate qu'un seul trajet existe à  $CV_{lim}$  (voir ci-dessus la définition du trajet, dans le paragraphe consacré à l'arrivée dans le site actif).

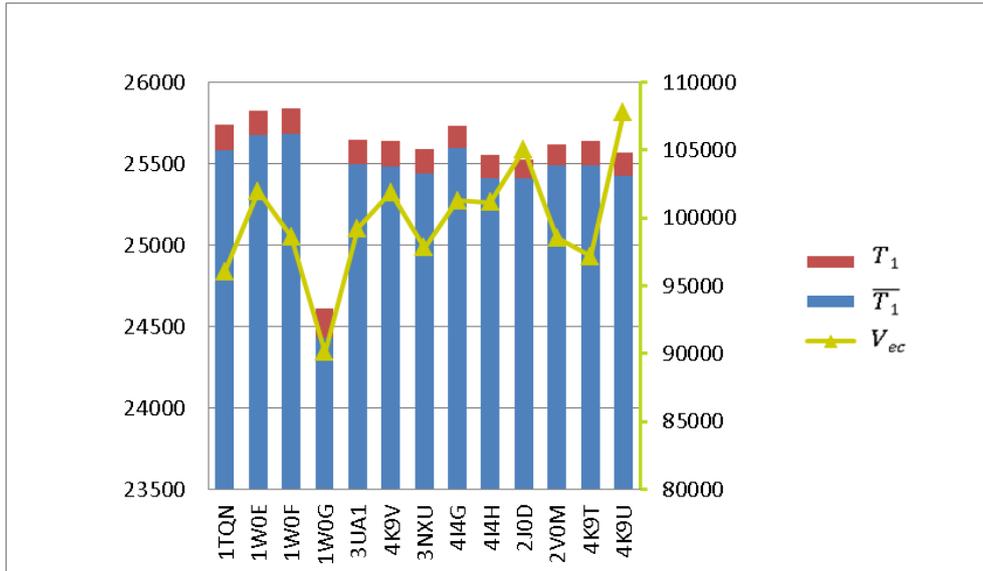
Les valeurs de CV qui génèrent des MCPs aboutissant à la face proximale de l'hème ne sont pas prises en compte. Les canaux proximaux ont été proposés dans la littérature comme participant à la circulation des autres partenaires de la réaction, molécules d'eau, protons ou électrons, via le partenaire rédox situé du côté proximal. La valeur de CV immédiatement supérieure à de telles CV est dénommé  $CV_{min}$ .

Nous diminuons la valeur critique CV depuis  $CV_{lim}$  jusqu'à  $CV_{min}$  par pas de 0.25Å afin de retenir les nouveaux MCPs et donc les nouveaux trajets.

Remarque : dans les MCPs, on retrouve souvent des acides aminés à la sortie du canal (en surface de la protéine) que l'on peut considérer comme ne bordant pas véritablement le canal car ils délimitent de très grands triangles dans des tétraèdres de surface très aplatis. Il a été retrouvé dans plusieurs cas des acides aminés de ce type, appartenant à la boucle HI reconstruite (281-285; voir table 2.2).

Code PDB	Forme	$V_{ec}$	$S_{ec}$	$R_{ec}$	$H_{ec}$	N	T	$t_1$
1TQN	apo	96011	10965	37.99	42.92	3844	25740	0.63
1W0E	apo	101966	11429	38.54	43.81	3860	25824	0.58
1W0F	holo	98613	11207	38.18	42.86	3860	25838	0.60
1W0G	holo	90081	10464	36.37	41.78	3695	24609	0.72
3UA1	holo	99229	11125	36.12	44.25	3825	25646	0.58
4K9V	holo	101853	11350	37.48	44.12	3825	25637	0.60
3NXU	holo	97807	11071	31.52	43.54	3815	25587	0.57
4I4G	holo	101250	11268	37.70	43.65	3832	25731	0.53
4I4H	holo	101161	11291	37.05	44.45	3815	25553	0.56
2J0D	holo	105060	11521	37.68	46.40	3815	25523	0.45
2V0M	holo	98517	11076	37.74	42.99	3821	25618	0.50
4K9T	holo	97175	11080	37.21	43.03	3823	25638	0.59
4K9U	holo	107748	11729	38.00	48.38	3832	25566	0.55

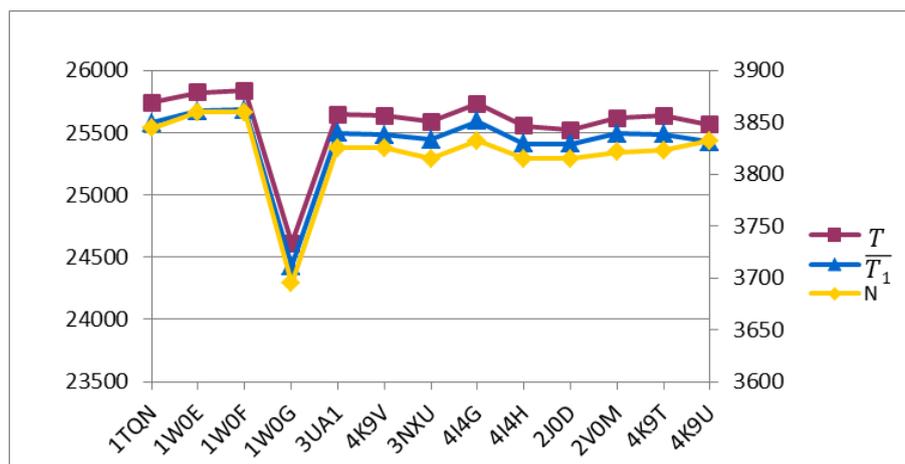
**Table 3.1 :** Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4. Paramètres de l'enveloppe convexe triangulée :  $V_{ec}$  (volume en  $\text{\AA}^3$ ),  $S_{ec}$  (surface en  $\text{\AA}^2$ ),  $R_{ec}$  (rayon en  $\text{\AA}$ ),  $H_{ec}$  (épaisseur en  $\text{\AA}$ ), N (nombre d'atomes), T (nombre de tétraèdres),  $t_1$  (pourcentage de tétraèdres externes). Le pourcentage  $\bar{t}_1$  de tétraèdres internes est supérieur à 99% ( $t_1 + \bar{t}_1 = 100\%$ ).



**Figure 3.1 :** Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4 : 1<sup>ere</sup> partie. En rouge, le nombre de tétraèdres externes  $T_1$ . En bleu, le nombre de tétraèdres internes  $\bar{T}_1$ . En vert, le volume de l'enveloppe convexe  $V_{ec}$  ( $\text{\AA}^3$ ). Axe vertical à gauche :  $T_1$  et  $\bar{T}_1$ . Axe vertical à droite :  $V_{ec}$ .

### 3.1.3 Génération du graphe de l'enveloppe convexe

La table 3.1 et les figures 3.1 et 3.2 présentent les caractéristiques de l'enveloppe convexe triangulée de 13 des 18 structures du 3A4 étudiées, avec ou sans substrat ou inhibiteur co-cristallisé. Le CYP3A4 est déformable si l'on se réfère à différents états structuraux, et la triangulation varie en fonction de l'état de l'enzyme car elle dépend de la position des atomes. La liste des tétraèdres varie donc aussi, même pour des structures très proches mais non superposables.



**Figure 3.2 :** Résultats de la triangulation de Delaunay de l'enveloppe convexe pour diverses structures du CYP3A4 : 2<sup>ème</sup> partie. En rouge, le nombre total de tétraèdres. En bleu, le nombre de tétraèdres internes  $\bar{T}_1$ . En jaune, le nombre d'atomes N (axe vertical à droite).

Les résultats de la table 3.1 sont cohérents : le nombre total de tétraèdres varie légèrement suivant les différentes structures de l'isoforme 3A4 car le nombre d'atomes résolus varie. De plus, une perturbation des coordonnées, même de faible ampleur, peut provoquer une augmentation ou une diminution du nombre de tétraèdres. La structure 1W0G a moins de tétraèdres, ce qui peut s'expliquer par un nombre plus élevé d'acides aminés manquants non reconstruits (voir table 2.2). Les tétraèdres externes sont aplatis et de faible volume. L'écrasante majorité des tétraèdres sont internes à l'enveloppe : il n'y a environ que 0.5% des tétraèdres en contact avec la surface. Ainsi le nombre total de tétraèdres et le nombre de tétraèdres internes suivent pratiquement la même courbe (figure 3.2). Le nombre total de tétraèdres dépend du nombre d'atomes résolus de la protéine. Le volume de l'enveloppe convexe est indépendant du nombre de tétraèdres, et ce dernier dépend du nombre d'atomes (figure 3.1). Les paramètres de la triangulation varient peu d'une conformation à l'autre, ce qui montre la robustesse des résultats fournis par CCCPP.

### 3.1.4 Génération du réseau complet des canaux

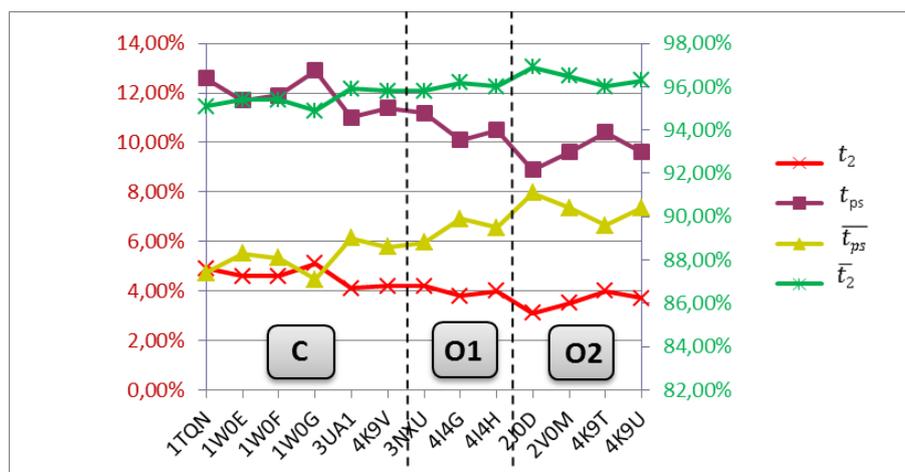
CCCPP calcule les canaux dans les protéines sous forme d'un ensemble de tétraèdres accessibles au ligand. Le réseau de canaux obtenu est modélisé par un graphe où chaque sommet représente un tétraèdre et chaque arête représente une face franchissable par le ligand. Nous présentons dans la table 3.2 et dans les figures 3.3 et 3.4 le réseau complet des canaux pour les structures du 3A4.

On constate que ce réseau contient la plupart des tétraèdres à la surface de l’enveloppe convexe ( $t_{ee} > 84\%$ ), c’est-à-dire que la surface de la protéine est majoritairement composée de poches. Nous observons que les pourcentages de tétraèdres externes dans les canaux ( $t_2$ ) sont faibles. Conventionnellement, nous appellerons **poches de surface** l’ensemble des tétraèdres accessibles externes et les tétraèdres accessibles voisins des tétraèdres accessibles externes. Ces poches de surface enveloppent toute la surface de la protéine. Elles ne sont pas réellement significatives, et il est intéressant d’analyser les canaux en les retirant (voir table 3.2) et figures 3.3 et 3.4). Les taux des tétraèdres externes dans les canaux ( $t_2$ ) sont plus petits que ceux des poches de surface ( $t_{ps}$ ), et sont faibles par rapport au pourcentage de tétraèdres internes dans les canaux ( $\bar{t}_2 = 100\% - t_2$ ). Les tétraèdres externes sont donc insignifiants pour l’analyse des canaux.

Code PDB	Forme	$T_a$	$t_a$	$t_{ii}$	$t_{ee}$	$t_{ps}$	$t_2$
1TQN	apo	3083	12.0	11.5	93.2	12.6	4.9
1W0E	apo	3085	11.9	11.5	94.0	11.7	4.6
1W0F	holo	3187	12.3	11.8	94.2	11.9	4.6
1W0G	holo	2957	12.0	11.5	84.3	12.9	5.1
3UA1	holo	3237	12.6	12.2	88.7	11.0	4.1
4K9V	holo	3185	12.4	12.0	87.7	11.4	4.2
3NXU	holo	3195	12.5	12.0	92.4	11.2	4.2
4I4G	holo	3214	12.5	12.1	89.7	10.1	3.8
4I4H	holo	3342	13.1	12.6	94.4	10.5	4.0
2J0D	holo	3408	13.4	13.0	93.0	8.9	3.1
2V0M	holo	3343	13.0	12.7	91.4	9.6	3.5
4K9T	holo	3418	13.3	12.9	89.5	10.4	4.0
4K9U	holo	3484	13.6	13.2	90.8	9.6	3.7

**Table 3.2 :** Réseau complet des canaux du CYP3A4 à  $CV=6\text{\AA}$ .  $T_a$  : nombre total de tétraèdres accessibles.  $t_a$  : pourcentage de tétraèdres accessibles par rapport au total des tétraèdres.  $t_{ii}$  : pourcentage de tétraèdres internes accessibles par rapport au total des tétraèdres internes.  $t_{ee}$  : pourcentage de tétraèdres externes accessibles par rapport au total des tétraèdres externes.  $t_{ps}$  : pourcentage de tétraèdres accessibles composant les poches de surface.  $t_2$  : pourcentage de tétraèdres externes dans les canaux.

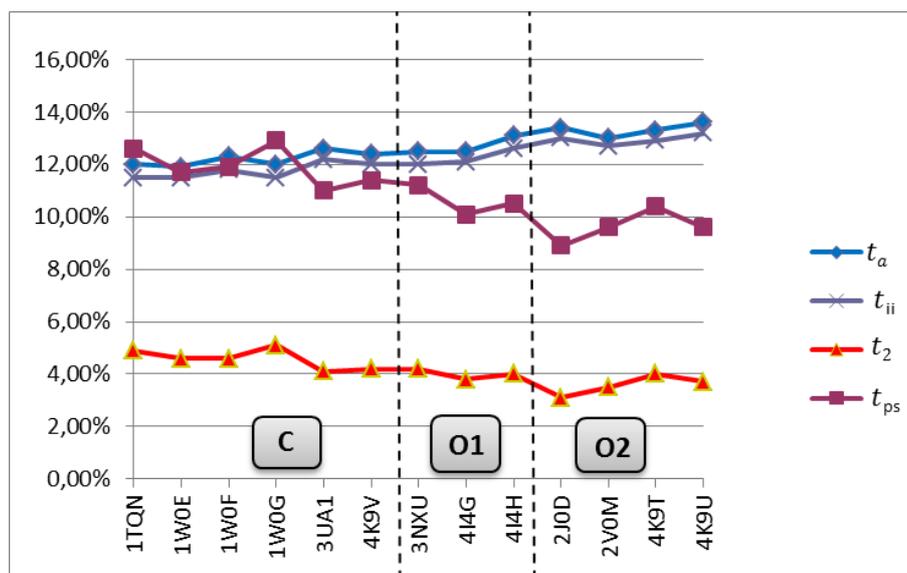
Nous observons sur la figure 3.4 que les pourcentages des tétraèdres accessibles  $t_a$  et des tétraèdres accessibles internes  $t_{ii}$  diffèrent peu, ceci car les tétraèdres accessibles externes sont peu nombreux. Le pourcentage  $t_{ps}$  de poches de surface est plus grand que celui des tétraèdres externes dans les canaux ( $t_2$ ), mais reste néanmoins petit par rapport à celui des canaux internes ( $\bar{t}_{ps}$ ). Le taux  $t_a$  de tétraèdres accessibles évolue suivant les conformations comme celui des tétraèdres internes accessibles ( $t_{ii}$ ), ce qui suggère de ne s’intéresser qu’aux canaux internes pour analyser le réseau. Les valeurs de  $t_{ps}$  et de  $t_2$  sont plus faibles pour les structures O1 puis O2, respectivement. Cela signifie qu’il y’a plus de canaux dans les structures O1 et O2, respectivement, que dans les structures C. Ceci peut s’expliquer par l’élargissement et l’ouverture des canaux des formes ouvertes par rapport aux formes fermées (voir section 3.4).



**Figure 3.3** : Répartition des tétraèdres internes et externes dans les canaux pour diverses structures du CYP3A4 : 1<sup>ère</sup> partie.  $t_1$  : pourcentage de tétraèdres externes (ordonnées : échelle de gauche).  $\bar{t}_1$  : pourcentage de tétraèdres internes (ordonnées : échelle de droite).  $t_{ps}$  : pourcentage de tétraèdres accessibles composant les poches de surface (ordonnées : échelle de gauche).  $\bar{t}_{ps}$  : pourcentage de tétraèdres accessibles dans les canaux, hors poches de surface (ordonnées : échelle de droite).

Les tétraèdres internes constituant le canal principal menant au site actif représentent une faible proportion ( $t_{ii}$ ) du nombre total des tétraèdres. C'est-à-dire que, à l'inverse des poches de surface, il n'y a qu'une faible proportion du domaine protéique accessible au ligand qui mène au site actif. Cette partie du domaine accessible au ligand menant au site actif définit un canal principal en forme d'entonnoir. On observe que, parmi les tétraèdres accessibles, ceux représentant les poches de surface sont minoritaires ( $t_{ps}$ ), et ceux représentant les canaux hors des poches de surface, sont majoritaires ( $\bar{t}_{ps}$ ). Ce n'est pas surprenant car les CYPs sont des protéines globulaires : la forme sphérique a pour propriété d'offrir la plus petite surface possible pour un volume donné. En revanche il faut souligner que les tétraèdres accessibles sont très majoritairement dans des poches de surface ( $t_{ee}$ ), et qu'une minorité des tétraèdres internes sont accessibles ( $t_{ii}$ ) : voir figure 3.12, section 3.2.1.

Après analyse des résultats pour plusieurs valeurs de CV, il est conclu que les canaux les plus pertinents pour la traversée des composés sont ceux des réseaux associés aux plus grandes valeurs de CV et correspondant aux plus larges passages. La pertinence de ces canaux s'apprécie par leur connexion de la surface du CYP à l'hème du côté distal, et par leur contenance de molécules co-cristallisées avec les structures, même pour des molécules de faible épaisseur (faible valeur de CV). Nous concluons que les voies d'accueil des molécules sont indépendantes de la dimension du ligand, mais plutôt liées à l'ouverture d'un canal. Rappelons que la plus grande valeur de CV révélant au moins un canal est nommée  $CV_{lim}$  (voir section 2.3.1). Chaque conformation du 3A4 aura sa valeur de  $CV_{lim}$  associée. Pour chaque isoforme nous diminuons la valeur de CV de  $CV_{lim}$  à  $CV_{min}$  avec un pas donné (0.25Å pour le 3A4), jusqu'à tomber sur des MCPs proximaux, c'est à dire reliant la surface l'hème du côté proximal (à l'opposé du site actif), et ne pouvant donc pas être une voie empruntée par les composés. Pour toutes les structures du 3A4 étudiées,



**Figure 3.4** : Répartition des tétraèdres internes et externes dans les canaux pour diverses structures du CYP3A4 : 2<sup>ème</sup> partie.  $t_a$  : pourcentage de tétraèdres accessibles par rapport au total des tétraèdres.  $t_{ii}$  : pourcentage de tétraèdres internes accessibles par rapport au total des tétraèdres internes.  $t_{ps}$  : pourcentage de tétraèdres accessibles composant les poches de surface.  $t_2$  : pourcentage de tétraèdres externes dans les canaux.

la valeur de  $CV_{min}$  est constante (à  $\Delta 0.25\text{\AA}$  près) et limitée par l'apparition d'un même canal (canal 2e : voir section 3.2.4).

Par contre, il apparaît une différence assez marquée entre les deux isoformes 1TQN et 1W0E qui sont pourtant très semblables ( $RMSD=0.77\text{\AA}$  sur 439 carbones  $\alpha$  du squelette) :  $CV_{lim}$  passe de  $6\text{\AA}$  (1TQN) à  $5.5\text{\AA}$  (1W0E). Une explication possible est que cette différence soit due à la position de certains acides aminés clés à longue chaîne situés dans le site actif (comme Arg212), qui ont été placés dans des conformations très différentes par les deux équipes de cristallographes, et qui peuvent alternativement obstruer ou libérer un canal. Il faut noter malgré tout que, pour une taille de ligand identique, le nombre de tétraèdres franchissables ( $T_a$ ) dans 1TQN et 1W0E est quasiment constant.

### 3.1.5 Extraction et marquage des composantes connexes

Le graphe facial du réseau complet des canaux présente plusieurs composantes connexes. Le nombre de composantes connexes augmente lorsque  $CV$  augmente. En effet, lorsque  $CV$  augmente, le nombre de faces infranchissables (arêtes du graphe facial) augmente, et plus il y a d'arêtes retirées (faces infranchissables), plus il apparaîtra de composantes connexes (donc non reliées entre elles par des arêtes), et en particulier il apparaît de plus en plus de composantes ne contenant qu'un seul noeud (tétraèdre isolés).

Le marquage des composantes consiste à sélectionner les composantes connexes reliant le site actif à l'extérieur. Ce marquage doit vérifier les contraintes suivantes :

- La composante sélectionnée doit contenir des tétraèdres faisant partie du site actif du CYP, c'est à dire qu'un des noeuds de ces tétraèdres a pour sommet l'atome

de Fer, et au moins l'un des 3 autres sommets du tétraèdre est un atome de la poche catalytique.

- La composante doit contenir des noeuds externes, c'est à dire des tétraèdres ayant au moins une face triangulaire externe franchissable.

Dans le cas des CYPs, la composante connexe marquée de plus grande taille (au sens du nombre de tétraèdres), est celle qui nous intéresse, car elle connecte l'extérieur à l'hème, et elle définit le canal principal du CYP pour le substrat donné (défini par son paramètre CV) : voir table 3.3.

Code PDB	Forme	$V_{pc}$	$S_{pc}$	$T_{pc}$	$t_3$	$t_{pceerc}$	$t_{pcec}$	$t_{pciiec}$	$t_{pceec}$
1TQN	apo	$42.4 \cdot 10^3$	$41.8 \cdot 10^3$	3073	4.7	95.3	11.9	11.5	88.8
1W0E	apo	$42.5 \cdot 10^3$	$40.5 \cdot 10^3$	3075	3.7	94.4	11.9	11.5	88.7
1W0F	holo	$44.1 \cdot 10^3$	$44.1 \cdot 10^3$	3177	4.4	96.6	12.3	11.8	91.0
1W0G	holo	$37.4 \cdot 10^3$	$40.8 \cdot 10^3$	2938	5.7	93.3	11.9	11.5	78.7
3UA1	holo	$43.5 \cdot 10^3$	$43.3 \cdot 10^3$	3082	5.1	95.5	12.0	11.6	84.7
4K9V	holo	$47.2 \cdot 10^3$	$43.6 \cdot 10^3$	3177	4.9	97.0	12.4	12.0	85.1
3NXU	holo	$44.7 \cdot 10^3$	$43.7 \cdot 10^3$	3185	4.4	95.5	12.4	12.0	88.3
4I4G	holo	$47.9 \cdot 10^3$	$43.9 \cdot 10^3$	3207	3.6	95.9	12.5	12.1	86.0
4I4H	holo	$46.9 \cdot 10^3$	$44.1 \cdot 10^3$	3333	4.2	95.6	13.0	12.6	90.2
2J0D	holo	$50.9 \cdot 10^3$	$44.8 \cdot 10^3$	3406	7.0	98.1	13.3	13.0	91.2
2V0M	holo	$46.7 \cdot 10^3$	$46.3 \cdot 10^3$	3337	4.7	96.6	13.0	12.6	88.3
4K9T	holo	$47.4 \cdot 10^3$	$45.0 \cdot 10^3$	3410	4.8	97.1	13.3	12.9	86.8
4K9U	holo	$49.9 \cdot 10^3$	$47.4 \cdot 10^3$	3480	4.6	98.4	13.6	13.2	89.4

**Table 3.3 :** Caractéristiques de la principale composante connexe, définissant le principal canal du CYP3A4, à CV=6Å.  $V_{pc}$  : volume du canal en Å<sup>3</sup>.  $S_{pc}$  : surface de la paroi du canal en Å<sup>2</sup> (somme des surfaces des triangles infranchissables du canal).  $T_{pc}$  : nombre de tétraèdres du canal.  $t_3$  : pourcentage de tétraèdres externes au sein du canal.  $t_{pceerc}$  : pourcentage de tétraèdres externes du canal relatif au nombre de tétraèdres externes du réseau complet des canaux.  $t_{pcec}$  : pourcentage de tétraèdres du canal relatif au total des tétraèdres de l'enveloppe convexe.  $t_{pciiec}$  : pourcentage de tétraèdres internes du canal relatif au nombre de tétraèdres internes de l'enveloppe convexe.  $t_{pceec}$  : pourcentage de tétraèdres externes du canal relatif au nombre de tétraèdres externes de l'enveloppe convexe. Le pourcentage de tétraèdres du canal relatif au nombre de tétraèdres du réseau complet des canaux, dépasse 99%, sauf pour 3UA1 où il n'atteint que 95.2%. De même, le pourcentage de tétraèdres internes du canal relatif au nombre de tétraèdres internes du réseau complet des canaux, dépasse 99%, sauf pour 3UA1 où il n'atteint que 95.2%.

Cette principale composante est largement majoritaire de par son nombre de tétraèdres. Elle définit un canal principal en forme d'entonnoir enveloppant la protéine à sa surface, avec une faible proportion de l'espace accessible localisée au niveau du site actif via la partie du canal principal situé du côté distal. La composante connexe associée au canal comporte plusieurs chemins de passage et trajets possibles convergeant vers la même zone du site actif, mais sortent en revanche dans des zones différentes à la surface. Les MCP sont calculés à partir des différents tétraèdres ayant l'atome de Fer comme sommet.

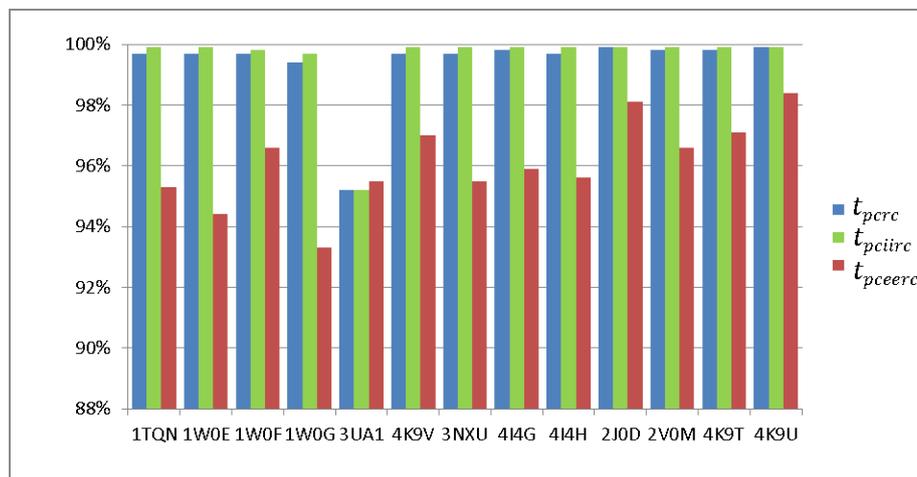
code PDB	$CV_{lim}$ (Å)	$V_{pc}$	$S_{pc}$	$T_{pc}$	$t_{pcec}$	$t_{pciiec}$	$t_{pceec}$
1TQN	6.00	$42.4 \cdot 10^3$	$41.8 \cdot 10^3$	3073	11.9	11.5	88.8
1W0E	5.75	$46.7 \cdot 10^3$	$44.1 \cdot 10^3$	3648	14.1	13.7	90.1
1W0F	6.00	$44.1 \cdot 10^3$	$44.1 \cdot 10^3$	3177	12.3	11.8	91.0
1W0G	6.50	$34.3 \cdot 10^3$	$39.6 \cdot 10^3$	2421	9.8	9.3	77.0
3UA1	6.75	$38.3 \cdot 10^3$	$40.4 \cdot 10^3$	2399	9.4	8.9	81.3
4K9V	6.50	$44.0 \cdot 10^3$	$42.8 \cdot 10^3$	2677	10.4	10.0	84.4
3NXU	6.25	$42.5 \cdot 10^3$	$42.2 \cdot 10^3$	2847	11.1	10.7	86.9
4I4G	6.00	$47.9 \cdot 10^3$	$43.9 \cdot 10^3$	3207	12.5	12.1	86.0
4I4H	6.25	$45.4 \cdot 10^3$	$44.3 \cdot 10^3$	3064	12.0	11.6	90.2
2J0D	8.75	$35.5 \cdot 10^3$	$39.4 \cdot 10^3$	1382	5.4	5.1	85.1
2V0M	7.00	$39.9 \cdot 10^3$	$42.7 \cdot 10^3$	2317	9.0	8.7	85.9
4K9T	6.75	$42.1 \cdot 10^3$	$42.5 \cdot 10^3$	2578	10.1	9.6	82.2
4K9U	6.75	$44.3 \cdot 10^3$	$44.6 \cdot 10^3$	2631	10.3	9.9	86.5

**Table 3.4 :** Caractéristiques de la composante connexe définissant le canal principal du CYP3A4 à  $CV_{lim}$ .  $V_{pc}$  : volume du canal en  $\text{Å}^3$ .  $S_{pc}$  : surface de la paroi du canal en  $\text{Å}^2$  (somme des surfaces des triangles infranchissables du canal).  $T_{pc}$  : nombre de tétraèdres du canal.  $t_{pcec}$  : pourcentage de tétraèdres du canal relatif au total des tétraèdres de l’enveloppe convexe.  $t_{pciiec}$  : pourcentage de tétraèdres internes du canal relatif au nombre de tétraèdres internes de l’enveloppe convexe.  $t_{pceec}$  : pourcentage de tétraèdres externes du canal relatif au nombre de tétraèdres externes de l’enveloppe convexe. Le pourcentage de tétraèdres du canal relatif au nombre de tétraèdres du réseau complet des canaux, dépasse 99%.

Dans le cas du 1TQN, le graphe facial possède 16 tétraèdres contenant l’atome de Fer, mais après marquage des composantes, la composante connexe sélectionnée ne contient plus que 9 tétraèdres sur 16. Ceci est attendu car seuls ces 9 tétraèdres sont du bon côté de la protoporphyrine : c’est le côté distal de l’hème du CYP3A4 où a lieu l’oxydation du ligand. Ces 9 tétraèdres ont donc bien l’atome du Fer comme sommet, mais pas l’atome de soufre attaché à la cystéine liée par covalence au Fer du côté proximal (côté opposé au site actif). Cela signifie que notre composante connexe marquée a du sens, car elle n’inclut pas de passage par le côté proximal, qui est la face de l’interaction supposée avec le partenaire P450-réductase (donneur d’électrons). Ceci est valable pour les autres structures du 3A4 mais avec un nombre différent de tétraèdres contenant l’atome de Fer et différent de ceux situés du côté distal. Cette différence révèle des variations entre les conformations des structures du 3A4, et de ce fait de la position des atomes, et donc une triangulation différente.

Le calcul de composantes a été effectué sur le 3A4 code 1TQN en-dessous du  $CV_{lim}$  jusqu’à  $CV=3.5\text{Å}$ . Il est intéressant de noter qu’à cette valeur de CV très basse il n’y a toujours pas de seconde composante marquée. Il n’existe qu’une composante qui permette le passage de l’extérieur vers le centre de la protéine (Fer). L’augmentation de CV jusqu’à  $CV_{lim}$  permet d’affiner la composante marquée. Lorsque CV augmente, le nombre de noeuds et d’arêtes de la composante marquée diminuent. Ceci est observé pour toutes les structures du 3A4. Cela signifierait que pour des petits ( $3.5\text{Å}$ ) comme pour des gros substrats du 3A4, il n’y a qu’un seul canal possible pour que le substrat puisse atteindre le site actif.

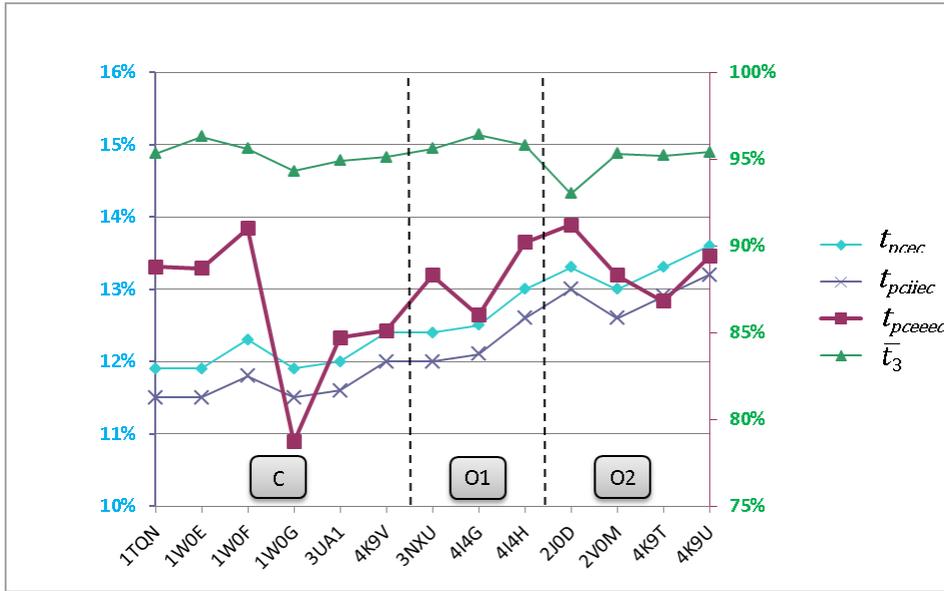
Lorsque CV diminue (substrat plus petit), le nombre de noeuds et d'arêtes de la composante marquée augmente, ce qui rend l'exploration et l'extraction des canaux d'accès au site actif plus complexe.



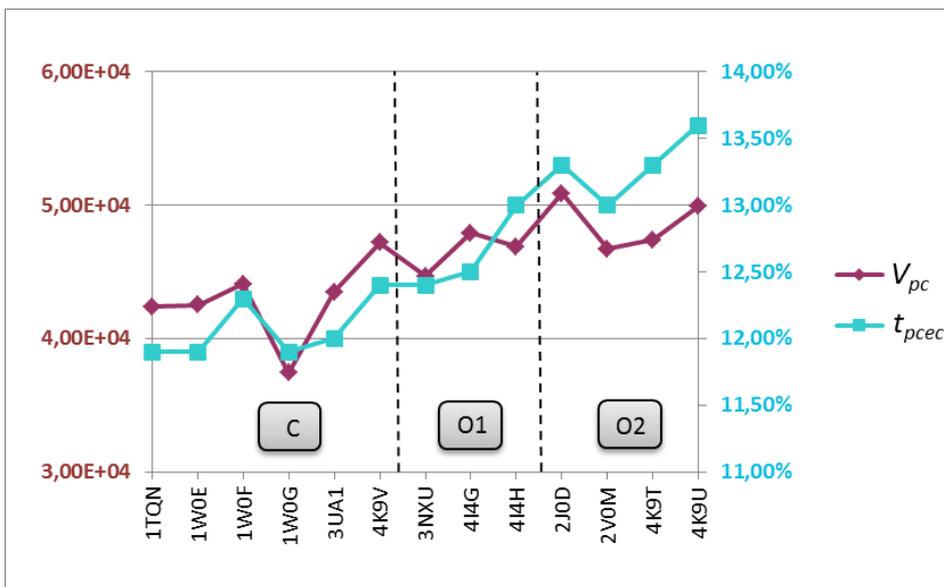
**Figure 3.5** : Répartition des tétraèdres dans le canal principal du CYP3A4 à  $CV=6\text{\AA}$ . En bleu,  $t_{perc}$  : pourcentage de tétraèdres du canal relatif au nombre de tétraèdres du réseau complet des canaux. En vert,  $t_{pciirc}$  : pourcentage de tétraèdres internes du canal relatif au nombre de tétraèdres internes du réseau complet des canaux. En rouge,  $t_{pceerc}$  : pourcentage de tétraèdres externes du canal relatif au nombre de tétraèdres externes du réseau complet des canaux.

A  $CV=6\text{\AA}$ , qui est la valeur moyenne de  $CV_{lim}$  pour l'ouverture du canal principal 2a du 3A4, la principale composante est quasiment identique au réseau de canaux internes (plus de 99.7%, sauf pour le 3UA1, à 95.2%). Cette identité est moins nette en surface mais reste quand même très importante (plus de 93%). Ceci s'explique par l'existence de nombreuses composantes connexes de 1 ou 2 noeuds au voisinage de la surface. A l'intérieur, le réseau comporte moins de composantes connexes de 1 ou 2 noeuds. Sur la figure 3.5, les compléments à 100% (non représentés), constituent les pourcentages des composantes connexes de 1 ou 2 noeuds ne connectant pas l'hème à l'extérieur.

La principale composante représente une faible proportion par rapport à l'enveloppe convexe (entre 11.9% et 13.6%). Il est intéressant de noter que la couche de surface est dans sa très grande majorité constituée de vides (environ 87%), quel que soit l'état du 3A4, fermé ou ouvert, comme dans le cas du réseau complet des canaux (voir 3.1.4). Cette couche de surface est constituée de tétraèdres externes très aplatis, avec une face externe très étendue, constituant une zone dont l'appartenance au domaine protéique peut être contestée. Il est donc plus utile de ne considérer que les couches internes pour analyser les canaux. De même que pour le réseau de canaux, la principale composante est plus grande respectivement pour les formes O1 et O2, par rapport à la forme C. Ceci se justifie par l'élargissement et l'ouverture des canaux (figure 3.6). La couche externe de la principale composante reste très majoritairement composée de vide (environ 87%). Les tétraèdres de la principale composante sont en grande majorité des tétraèdres internes (environ 95%).



**Figure 3.6** : Importance du canal principal du CYP3A4 au sein de l'enveloppe convexe : 1<sup>ere</sup> partie. En bleu,  $t_{pcec}$ , pourcentage de tétraèdres du canal relatif au total des tétraèdres de l'enveloppe convexe (ordonnées : échelle de gauche). En violet,  $t_{pciiec}$ , pourcentage de tétraèdres internes du canal relatif au nombre de tétraèdres internes de l'enveloppe convexe (ordonnées : échelle de gauche). En rouge,  $t_{pceec}$ , pourcentage de tétraèdres externes du canal relatif au nombre de tétraèdres externes de l'enveloppe convexe (ordonnées : échelle de droite). En vert,  $\bar{t}_3$ , pourcentage de tétraèdres internes relatif au total des tétraèdres dans le canal principal (ordonnées : échelle de droite).



**Figure 3.7** : Importance du canal principal du CYP3A4 au sein de l'enveloppe convexe : 2<sup>eme</sup> partie. En rouge,  $V_{pc}$ , volume du canal (ordonnées : échelle de gauche). En bleu,  $t_{pcec}$  : pourcentage de tétraèdres du canal relatif au total des tétraèdres de l'enveloppe convexe (ordonnées : échelle de droite).

En revanche, contrairement au réseau de canaux (figure 3.3), pour cette principale composante il n’y a pas augmentation du taux de tétraèdres internes de l’état ouvert par rapport à l’état fermé. Ceci s’explique par une élimination des composantes connexes isolées à la surface de l’enveloppe, celles-ci étant plus nombreuses dans le cas des formes C et O1 par rapport à O2 (figure 3.5), ce qui compenserait le pourcentage croissant de vide interne dans la principale composante pour la forme O2.

Les structures du 3A4 sont semblables : les valeurs de RMSD (voir section 2.5), calculées par rapport à 1TQN sur les squelettes, varient entre 0.245 et 0.800, générant ainsi un nombre de tétraèdres et des dimensions comparables. Concernant l’importance de la principale composante au sein de l’enveloppe convexe, le volume suit la même tendance que le nombre total de tétraèdres et le nombre des tétraèdres internes. Le volume des composantes suit la même tendance que la proportion de la principale composante par rapport à l’enveloppe convexe, qui elle-même suit les variations des nombres de tétraèdres internes. Ce dernier, considéré comme significatif pour analyser nos réseaux, augmente de la forme fermée à la forme ouverte 1, puis 2 (figure 3.7). Ceci est cohérent du fait que les tétraèdres soient de dimensions comparables.

### 3.1.6 Etablissement des chemins d’accès au site actif

Les MCPs sont des sous-graphes de la principale composante, et un ensemble de MCPs ayant un tronc commun constitue un trajet (voir section 2.3.1). Cette étape consiste à extraire des MCPs de la principale composante et à parcourir cette principale composante afin de trouver les chemins menant des noeuds du site actif (tétraèdres contenant l’atome de Fer) vers les noeuds externes (tétraèdres externes). Les MCPs peuvent appartenir au même trajet ou bien être de trajets différents (voir section 2.3.1). On observe que les MCPs d’un même trajet ont des valeurs de coût minimal proches.

CCCPP permet de définir des parois à l’aide des faces triangulaires infranchissables délimitant les canaux ou les chemins. Cette description détaillée des parois permet des comparaisons précises au niveau des atomes, au niveau des acides aminés et des structures secondaires, et non plus uniquement au niveau des structures secondaires comme c’est le cas dans la littérature ([Cojocar *et al.* 2007]).

Code PDB	$CV_{lim}$	$C_a$	$C_r$	$V_c$	$S_c$	$T_c$	$N_c$
1TQN	6.00	24.8	0.80	613	273	33	36
1W0E	5.75	19.8	0.70	1230	128	29	32
1W0F	6.00	19.8	0.70	589	107	28	31
1W0G	6.50	26.2	0.80	925	265	34	37
3UA1	6.75	20.6	0.70	519	473	28	31
4K9V	6.50	23.8	0.72	520	659	33	36

**Table 3.5 :** Description du canal 2a du CYP3A4 en forme fermée : apparition du premier canal.  $C_a$  : coût total absolu.  $C_r$  : coût total relatif (coût total absolu divisé par le nombre d’arêtes du chemin; ce coût est compris entre 0 et 1).  $V_c$  : volume en  $\text{Å}^3$ .  $S_c$  : surface en  $\text{Å}^2$ .  $T_c$  : nombre de tétraèdres.  $N_c$  : nombre total des atomes bordant le canal, sans redondance.

Code PDB	$CV_{lim}$	$C_a$	$C_r$	$V_c$	$S_c$	$T_c$	$N_c$
1TQN	5.75	19.5	0.70	841	567	28	31
1W0E	5.75	17.8	0.69	855	219	26	29
1W0F	5.75	24.6	0.72	1046	738	34	36
1W0G	5.75	21.7	0.68	714	571	32	33
3UA1	5.50	21.7	0.68	1287	208	32	35
4K9V	6.00	18.8	0.69	1338	434	27	30

**Table 3.6 :** Description du canal 2e du CYP3A4 en forme fermée : apparition du nouveau canal.  $C_a$  : coût total absolu.  $C_r$  : coût total relatif (coût total absolu divisé par le nombre d'arêtes du chemin; ce coût est compris entre 0 et 1).  $V_c$  : volume en  $\text{\AA}^3$ .  $S_c$  : surface en  $\text{\AA}^2$ .  $T_c$  : nombre de tétraèdres.  $N_c$  : nombre total des atomes bordant le canal, sans redondance.

Code PDB	$CV_{lim}$	canal	forme	$C_a$	$C_r$	$V_c$	$S_c$	$T_c$	$N_c$
3NXU	6.25	2a	O1	27.4	0.70	917	342	39	42
4I4G	6.00	2a	O1	20.0	0.67	1284	286	30	32
4I4H	6.25	2a	O1	24.7	0.67	876	656	37	40
2J0D	8.75	2f	O2	21.9	0.76	848	602	29	32
2V0M	7.00	2a	O2	24.3	0.71	774	896	34	37
4K9T	6.75	2f	O2	25.1	0.76	668	265	33	36
4K9U	6.75	S	O2	23.3	0.66	1458	537	35	38

**Table 3.7 :** Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du premier canal.  $C_a$  : coût total absolu.  $C_r$  : coût total relatif (coût total absolu divisé par le nombre d'arêtes du chemin; ce coût est compris entre 0 et 1).  $V_c$  : volume en  $\text{\AA}^3$ .  $S_c$  : surface en  $\text{\AA}^2$ .  $T_c$  : nombre de tétraèdres.  $N_c$  : nombre total des atomes bordant le canal, sans redondance.

Nous avons ainsi pu calculer, pour chaque MCP, le volume et la surface de la paroi, grâce à la liste des tétraèdres le composant (voir tables 3.5 à 3.9). Cette liste permet d'obtenir immédiatement la liste des atomes puis la liste des acides aminés définissant un MCP et un trajet. Cette description précise des canaux et des trajets avec matérialisation des parois (avec calculs de surfaces et de volumes) est originale. Ceci contraste avec les canaux calculés par les méthodes usuelles (CAVER, MOLE), qui sont simplement décrit par des suites de sphères en intersection.

## 3.2 Description des canaux du CYP3A4

Le canal calculé par CCCPP est en forme d'entonnoir. Il comprend les trajets de coûts minimaux calculés avec la seconde version de CCCPP (voir à la fin de la section 2.3.1), et se trouve localisé dans une région commune à toutes les isoformes des P450s, correspondant à la région des canaux 2 dans la nomenclature de [Cojocar *et al.* 2007]. Cette région est comprise entre les 5 structures secondaires suivantes : l'ancre A (position fixe), le feuillet  $\beta_1$ , la boucle BC, la région FG (hélice-boucle-hélice) très flexible en raison de l'ouverture des canaux, et la boucle C-terminale (position un peu dépendante de la

Code PDB	$CV_{lim}$	canal	forme	$C_a$	$C_r$	$V_c$	$S_c$	$T_c$	$N_c$
2J0D	6.75-8.50	bloc 2f*	O2	21.3	0.73	848	602	29	32
2V0M	5.75	2f	O2	32.6	0.71	851	360	46	46
4K9T	5.75	bloc 2f*	O2	33.3	0.69	566	211	48	49
4K9U	6.25	2f	O2	25.3	0.72	780	314	35	38

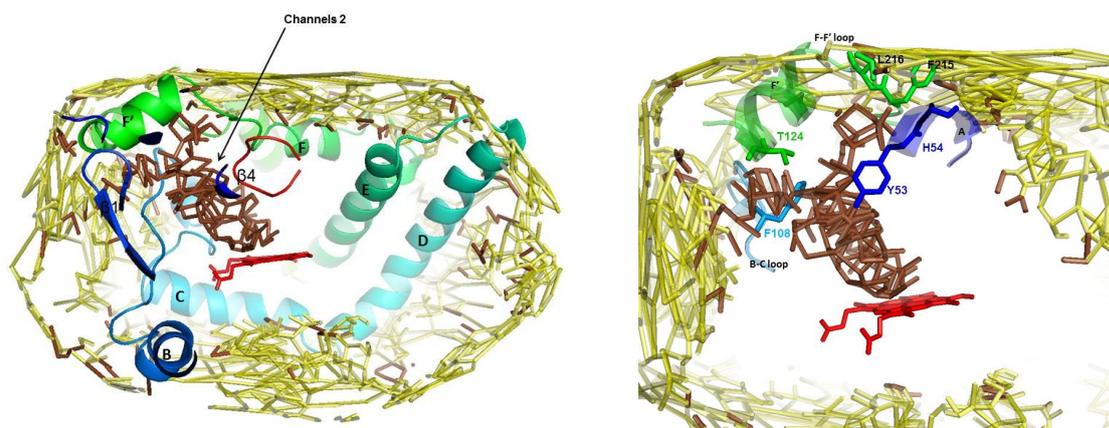
**Table 3.8 :** Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du deuxième canal (2f).  $C_a$  : coût total absolu.  $C_r$  : coût total relatif (coût total absolu divisé par le nombre d'arêtes du chemin; ce coût est compris entre 0 et 1).  $V_c$  : volume en  $\text{\AA}^3$ .  $S_c$  : surface en  $\text{\AA}^2$ .  $T_c$  : nombre de tétraèdres.  $N_c$  : nombre total des atomes bordant le canal, sans redondance. \* : Ensemble des MCPs complémentaires dans la zone du canal concerné.

Code PDB	$CV_{lim}$	canal	forme	$C_a$	$C_r$	$V_c$	$S_c$	$T_c$	$N_c$
3NXU	5.75	2e	O1	23.6	0.66	958	386	36	38
4I4G	5.75	2e	O1	20.4	0.66	1139	312	31	34
4I4H	6.00	2e	O1	22.7	0.69	521	556	33	36
2J0D	5.50	2e	O2	23.3	0.71	1070	228	33	35
2V0M	5.50	2e	O2	19.7	0.64	599	618	31	34
4K9T	5.50	2e	O2	21.0	0.62	673	170	34	36
4K9U	5.75	2e	O2	18.9	0.65	992	173	29	32

**Table 3.9 :** Description des canaux des formes ouvertes au niveau des blocs 1 et 2 : apparition du troisième canal (2e).  $C_a$  : coût total absolu.  $C_r$  : coût total relatif (coût total absolu divisé par le nombre d'arêtes du chemin; ce coût est compris entre 0 et 1).  $V_c$  : volume en  $\text{\AA}^3$ .  $S_c$  : surface en  $\text{\AA}^2$ .  $T_c$  : nombre de tétraèdres.  $N_c$  : nombre total des atomes bordant le canal, sans redondance : il est à noter que dans la structure 4K9U, il y'a apparition du canal 2a à  $CV=5.75\text{\AA}$ .

conformation d'une isoforme donnée) : voir figure 3.8). Cette région 2 varie fortement en séquence et en structure d'un CYP à l'autre, et est connue pour son importance dans la spécificité aux substrats à travers les contacts établis avec les ligands. Tous les canaux de la région 2 (impliquée dans les canaux d'entrée), excepté le 2d et le 2f, sont bordés par la boucle BC. En effet, la boucle BC a été identifiée comme possible région SRS-1, impliquée dans la liaison des substrats [Gotoh 1992]. La flexibilité des deux régions BC et FG est la source principale du changement conformationnel pour une isoforme donnée, car la longueur, la structure secondaire et la séquence de ces deux régions varient considérablement suivant les P450s [Skopalik *et al.* 2008].

Le canal d'accès à la large cavité du site actif apparaît à  $CV_{lim}$  et à  $CV > CV_{lim}$ , et seules les poches de surface sont accessibles (visibles par le calcul CCCPP) : il y a donc ici une évidence d'existence d'un goulot d'étranglement à l'entrée du ou des canaux (figure 3.8 à droite). Un goulot d'étranglement est là où le CV du canal est minimal : c'est ainsi le goulot qui détermine le CV d'apparition d'un canal. Les canaux sont bordés par les chaînes latérales des acides aminés, à quelques exceptions près (lorsqu'un des atomes de la liaison amide du squelette est en bordure); de ce fait la propriété du canal est liée à la propriété de l'acide aminé qui la borde.

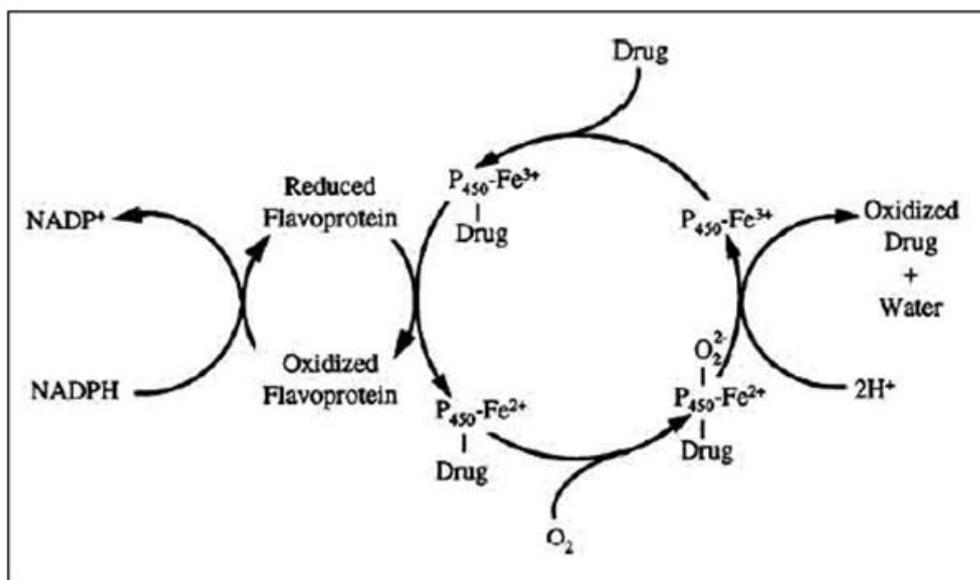
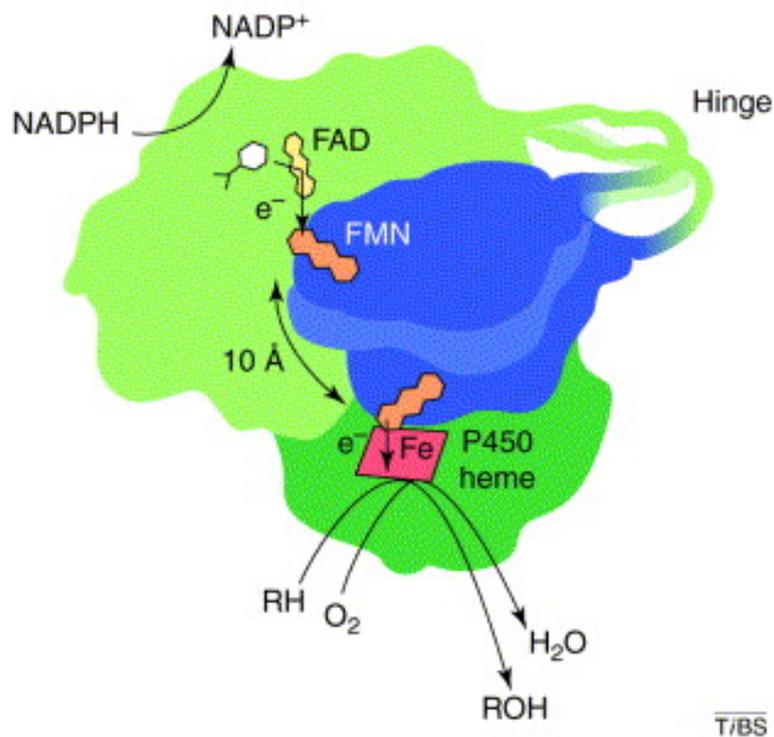


**Figure 3.8** : Réseau de canaux pour le CYP3A4 (code PDB 2V0M). A gauche, le réseau (en forme d'entonnoir), situé dans la région du canal 2 entre les cinq structures secondaires citées dans le texte; en jaune, les poches apparaissant à  $CV > CV_{lim}$ ; en marron, le canal reliant au site actif qui apparaît à  $CV_{lim}$ . A droite (zoom de l'image de gauche) : goulets d'étranglement des canaux 2a et 2f à  $CV_{lim}$ ; les acides aminés formant les goulets sont en bâtons, en vert et bleu. Les acides aminés formant les goulets sont affichés en bâtons et coloriés en vert et bleu.

L'analyse des simulations des cytochromes P450cam, P450-BM3 et P450eryF, a montré que les trajectoires de sortie sont dans la région du canal 2. Les simulations de dynamique moléculaire du P450cam ont montré que les substrats et les produits pourraient sortir du site actif via trois routes, mais que le parcours 2 (c'est à dire celui de la région 2) est énergétiquement favorisée [Lüdemann *et al.* 2000a, Lüdemann *et al.* 2000b].

Dans la littérature, des simulations SMD de l'expulsion hors du 3A4 des produits d'hydroxylation du diazepam, de la testostérone, du temazépam et de la  $6\beta$ -OH testostérone, ont été faites afin d'identifier les routes préférées des substrats ou produits ainsi que les mécanismes d'ouverture. Six routes différentes de sortie des produits ont été localisées, avec différents lieux de sortie. La plupart des canaux contiennent une ou deux phénylalanines empilées contrôlant l'endroit où s'effectue l'ouverture. Pour sortir, le médicament oxydé brise les interactions hydrophobes dues à ces acides aminés, et établit des contacts essentiellement hydrophobes au niveau de l'ouverture. Il est indiqué que les choix de sortie sont régulés par la spécificité protéine-substrat. Il a été trouvé que la forme géométrique du produit et son orientation après oxydation déterminent le choix du canal de sortie, basé sur l'affinité stéréoélectronique de la molécule pour le canal. Ce mécanisme n'est pas conservé au sein des P450s, mais existe pour beaucoup d'isoformes et de canaux. La polyvalence du mécanisme de sortie du 3A4 est significative, si l'on considère le fait que cette isoforme oxyde une grande diversité de substrats. Cette diversité nécessite à son tour un bon nombre de voies de sortie [Fishelovitch *et al.* 2009].

Dans notre étude des canaux d'accès des ligands, nous avons retenu uniquement les canaux distaux susceptibles d'être les voies de circulation des substrats, inhibiteurs, ou produits d'oxydation. Des canaux proximaux existent. Ce sont de petits canaux qui apparaissent dans l'exploration par CCCPP pour les petites valeurs de CV, et, comme décrit section 3.1, c'est leur apparition qui marque le  $CV_{min}$ . Ceux-ci sont supposés être impliqués dans la circulation des petites molécules actives dans le cycle catalytique (molécules d'eau, molécules de dioxygène, protons, électrons : voir figure 3.9).



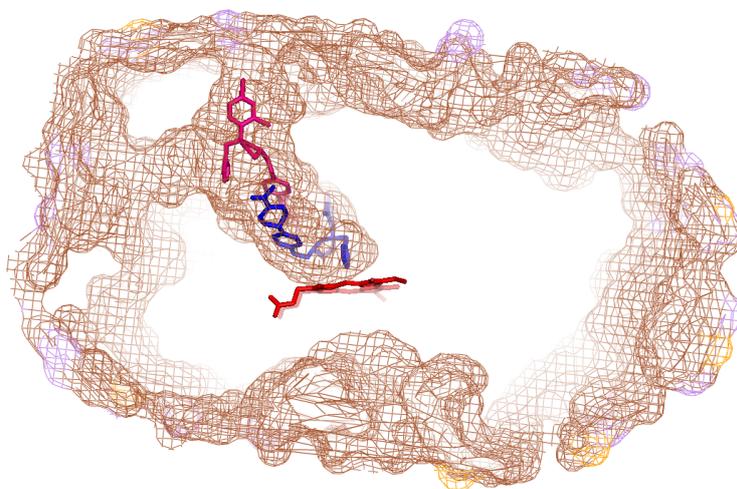
**Figure 3.9** : Cycle catalytique du cytochrome P450. Image du haut extraite de <http://enzymology.lab.ls.manchester.ac.uk/labgroups/leysgroups/> le 24/11/2014. Image du bas : figure 1, [Kalra 2007].

Dans la littérature, des simulations de dynamique moléculaire à solvant explicite ont été faites, avec liaison à un ou deux substrats, ou à aucun, avec ou sans le domaine de liaison flavine-mononucléotide (FMN) de la CPR (Cytochrome P450 Réductase). La CPR aide à réguler le cycle catalytique du 3A4 en utilisant le canal aqueux qui est conservé. Les molécules d'eau ont un rôle majeur dans le cycle catalytique du P450. Il a été observé des entrées et sorties de molécules d'eau via un canal aqueux spécifique et conservé (aqueduc), localisé entre le site actif et le côté proximal de l'hème. Ce mécanisme de contrôle avec un aqueduc est arbitré par R375 (situé sur la boucle K-feuillet  $\beta_1$ ), l'arginine conservée qui offre un pont salin avec l'hème-7-propionate. Lorsque R375 effectue une rotation, elle ouvre l'aqueduc et établit une connexion entre un amas de molécules d'eau du site actif et le reste du solvant. La région de l'aqueduc chevauche le site de liaison de la CPR au 3A4. Lorsque le domaine FMN de la CPR est lié au 3A4, l'aqueduc s'ouvre entièrement, permettant alors un flux de molécules d'eau. L'ouverture de l'aqueduc peut permettre un transfert de protons, transportant ceux-ci au site actif à travers un réseau de molécules d'eau. De plus, l'expulsion de molécules d'eau par l'aqueduc contribue à lier le substrat. Ainsi, la liaison à la CPR a pour fonction de déclencher l'ouverture de l'aqueduc et ainsi occasionner une voie pour le transport d'un proton, nécessaire pour activer le dioxygène. Ce mécanisme pourrait être un paradigme général chez les P450s [Fishelevitch *et al.* 2009].

### 3.2.1 Canal 2a : le canal principal

L'importance des canaux 2a a été soulignée [Schleinkofer *et al.* 2005, Lüdemann *et al.* 2000a, Lüdemann *et al.* 2000b, Winn *et al.* 2002], et les simulations ont montré la capacité des canaux 2a à s'ouvrir pour permettre la sortie des substrats et des produits. Sur la représentation du graphe facial des passages accessibles (figure 3.8), un canal en forme d'entonnoir, toujours ouvert, apparaît dans toutes les formes apo et holo du 3A4. Ce canal principal est le premier apparu pour l'ouverture du bloc 1 (des formes C et O1) lors des calculs avec CCCPP à  $CV_{lim}$  : c'est le plus gros des canaux permettant l'accès au site actif via le bloc 1. Il s'agit du canal 2a, situé dans la zone des canaux 2. Il permet une grande ouverture au voisinage du site actif. Ce canal 2a est bordé par trois régions fortement variables en séquence et en structure parmi les P450s : feuillet  $\beta_1$ , boucle BC, et région FG. Le graphe facial correspondant à ce canal principal est connexe à  $CV_{lim}$ , c'est à dire qu'il n'y a qu'une seule composante connexe comprenant les concavités formant ce canal. Dans cette composante connexe, la majorité des tétraèdres de surface sont des cavités qui représentent les poches de surface (en moyenne 87.5%). Seule une petite partie des tétraèdres internes sont des vides (en moyenne 12.2%), située dans la face distale et connecte l'extérieur à l'hème, et qui représente le canal (figures 3.10, 3.11, 3.12), et c'est celle qui correspond au trajet 2a pour les structures ouvertes dans le bloc 1. On peut considérer que les poches de surfaces, ne représentant de surcroît qu'en moyenne 4.75% du total du canal, sont des concavités n'appartenant pas au domaine protéique.

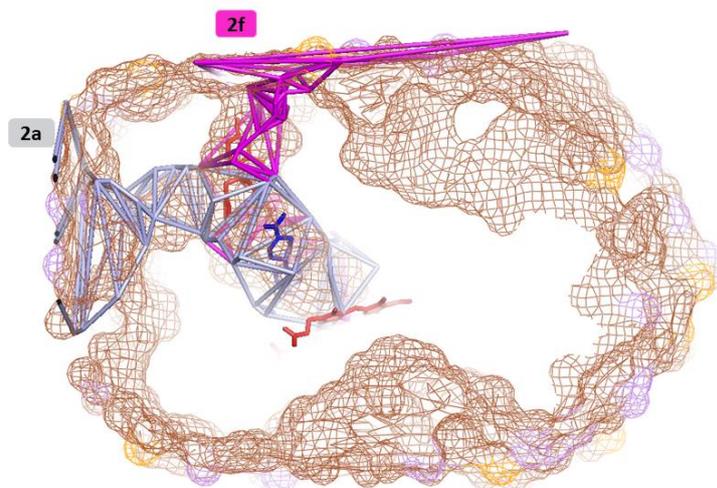
L'hélice G' borde la bouche du canal 2a (figure 3.8). Les acides aminés de l'hélice G' ont un caractère apolaire, car l'hélice G' est ancrée à la membrane plasmique. Le goulot d'étranglement dans le canal 2a est au niveau de F108 et de T224 (figure 3.8 à droite). Dans la conformation ouverte du 3A4 code PDB 4K9T, F108 est orienté vers l'intérieur du canal par rapport à sa conformation dans 2V0M : il y a un resserrement du goulot d'étranglement qui gênerait l'ouverture du canal 2a.



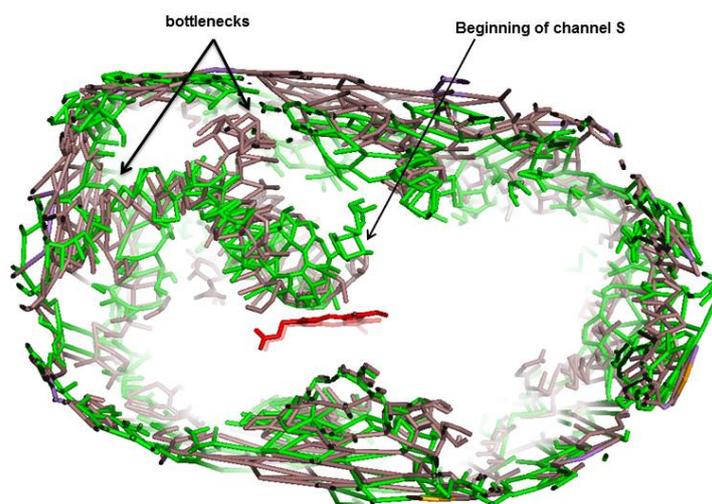
**Figure 3.10** : Graphe facial du réseau de canaux (à  $CV=5.75\text{\AA}$ ) du CYP3A4 code PDB 2V0M, comportant deux molécules de kétoconazole co-cristallisées. Le graphe facial est représenté par un maillage de couleur marron (option de visualisation *mesh* du logiciel PyMOL<sup>TM</sup> (<https://www.pymol.org/>)). Les épaisseurs des deux molécules de kétoconazole (avec les hydrogènes), calculées avec le logiciel RADI (voir section 2.3.1) valent respectivement  $5.19\text{\AA}$  et  $6.52\text{\AA}$  (les hydrogènes sont inclus). Les chemins de coûts minimaux (trajet 2a en bâtons gris, et trajet 2f en bâtons violets), calculés pour  $CV=5.75\text{\AA}$ , sont superposés à l'intérieur des canaux du 3A4. Chacun de ces deux chemins peut être une voie de passage pour chacune des deux molécules de kétoconazole (en bleu et rouge) : voir section 3.7.3.

Le canal 2a apparaît dans les formes fermées à une valeur de  $CV_{lim}$  voisine de  $6\text{\AA}$ . Un élargissement de ce canal est observé lors de l'accueil du ligand dans le cas des 3 formes : ouverte au niveau du bloc 1 (**O1**), ou du bloc 2 (**O2**), ou fermée (**C**) (voir section 3.4). C'est le cas par exemple de la forme O2 de la structure holo 2V0M. L'entonnoir apparaît pour une taille maximale du ligand  $CV_{lim}=7\text{\AA}$  (soit  $1\text{\AA}$  de différence avec la forme apo 1TQN), et permet d'observer 2 voies dans les 2 blocs distincts (voir section 3.4), chacune correspondant à un sous-canal. Dans la forme apo fermée 1TQN, seul le bloc 1 est ouvert à travers le canal 2a, avec une dimension d'ouverture "standard" à  $CV_{lim}=6\text{\AA}$ . Dans la forme O2 holo 2V0M, le canal 2a s'élargit à  $CV_{lim}=7\text{\AA}$ . Dans cette forme O2 2V0M, le bloc 2 à  $CV_{lim}=7\text{\AA}$  est tout aussi ouvert que le bloc 1, sauf que son accès est obstrué par un goulot d'étranglement de  $CV < 6\text{\AA}$  ( $\pm 0.25\text{\AA}$ ). Au-delà de cette limite, le bloc 2 s'ouvre, et est calculé par CCCPP (voir section 3.7.3).

Le canal 2a est calculé sur les deux structures apo 1W0E et 1TQN, avec respectivement  $CV_{lim}=5.75\text{\AA}$  et  $CV_{lim}=6\text{\AA}$ . Ce canal est ouvert pour les formes holo et il s'élargit jusqu'à  $CV=7\text{\AA}$  pour les ligands pouvant arriver au site actif (du point de vue physico-chimique), le canal 2a s'élargit pour accepter une première molécule, puis (ou sinon) d'autres voies sont nécessaires pour accepter d'autres molécules (voir section 3.2.2).



**Figure 3.11** : Chemins de coûts minimaux 2a et 2f, calculés pour  $CV=5.75\text{\AA}$  à l'intérieur des canaux du CYP3A4, code PDB 2V0M. Chacun des deux chemins est supposé être une voie de passage pour chacune des 2 molécules de kétoconazole (en bleu et rouge). Le réseau des canaux est représenté par un maillage en marron (option de visualisation *mesh* du logiciel Maestro de Schrödinger LLC). Les chemins sont représentés par des bâtons en gris (2a) et en violet (2f).



**Figure 3.12** : Superposition des réseaux de canaux de deux structures du CYP3A4, codes PDB 1TQN et 2V0M, respectivement en vert et en marron, et à  $CV_{lim}$   $6\text{\AA}$  et  $7\text{\AA}$ .

Le canal 2a semble être la route la plus probable pour l'arrivée du substrat et la sortie du produit, selon une étude sur des protéines bactériennes [Winn *et al.* 2002]. Pour les P450 bactériens, 2a est ouvert dans la plupart des structures qui ont au moins un canal ouvert. Les simulations de sortie de produit dans CYP101 indiquent que ce canal 2a serait utilisé par le produit aussi bien que par le substrat [Lüdemann *et al.* 2000a].

### 3.2.2 Canal 2f

Comme nous l'avons vu à propos du canal 2a (section 3.2.1), lorsque le bloc 2 est ouvert le graphe facial à  $CV_{lim}$  annonce l'existence des deux voies distinctes connectant l'hème à l'extérieur, correspondant aux voies du bloc 1 et du bloc 2. Par contre, le calcul des MCPs et des trajets avec la seconde version de CCCPP nous montre deux situations.

Dans la première situation, à une CV où apparaissent deux canaux (2a et 2f) dans le graphe facial, CCCPP ne calcule qu'un seul trajet relatif à la voie 2f du bloc 2, et ne calcule pas le trajet relatif au bloc 1. Cela peut s'expliquer comme suit. Lors d'une nouvelle passe de calcul des MCPs (voir section 2.3.1), la suppression des arêtes des précédents MCPs (correspondant au plus petit coût) peut empêcher de trouver un MCP ayant un tronc commun avec les MCPs trouvés précédemment, mais tels que ces MCPs se sépareraient au niveau des sorties. Autrement dit, le plus gros des deux canaux calculés à ce CV partage un tronc commun avec l'autre canal, et donc c'est le MCP du plus gros canal (correspondant au plus petit coût) qui est calculé au détriment de celui du plus petit canal. C'est le cas du canal 2a dans les structures de forme O2 du 3A4 (2J0D), qui n'est pas calculé au détriment du calcul du canal 2f, plus gros (ceci est considéré comme une limite de CCCPP).

Dans la seconde situation, CCCPP calcule le trajet 2f, mais à CV bien inférieure à  $CV_{lim}$ , alors que ce canal 2f est visible à une CV supérieure au CV où est calculé son trajet (cas de 2V0M et de 4K9U : voir table 3.8 et sections 3.6.6 et 3.7.3). Cette seconde situation est due à l'existence d'un engorgement à la sortie de ce deuxième canal 2f, et qui n'est calculé qu'à cette  $CV < CV_{lim}$ , cette CV permettant de surmonter l'obstacle. Ceci démontre la précision de nos résultats obtenus avec CCCPP

Dans toutes les formes holo ouvertes sur le bloc 2, un canal 2f est ouvert (figure 3.11), situé entre la région FG et la boucle C-terminale. Il est identifié comme un trajet distinct et se trouve entre le canal 2a et le canal S. Pour les quatre structures ouvertes sur le bloc 2 (2J0D, 2V0M, 4K9T, 4K9U), le canal 2f est calculé à partir de  $CV=5.75\text{\AA}$  et s'ouvre jusqu'à  $9\text{\AA}$  (chaîne B érythromycine). Dans le cas de deux molécules du ligand, l'ouverture du canal 2f survient pour l'accueil d'au moins une des deux molécules (2V0M, 4K9T et 4K9U). Les canaux 2a et 2f, situés tous deux dans la zone des canaux 2, ont une partie commune, et se séparent à l'entrée au niveau de l'acide aminé T224, situé sur l'hélice F' (figure 3.12).

Pendant la circulation des molécules (entrée du substrat ou sortie de produit), un canal doit être ouvert pour permettre aux ligands d'atteindre ou de partir du site actif [Yu *et al.* 2013]. Le goulot d'étranglement correspond à une disposition des acides aminés bordant le canal à son entrée, qui seraient flexibles en fonction de la forme de la structure, qui elle-même se modifie en fonction du ou des ligands. Il y a un goulot d'étranglement qui entoure l'entrée du canal 2f. Celle-ci est entourée par quatre acides aminés (T53, H54, F215, L216; voir figure 3.8), appartenant aux deux structures secondaires, l'une plutôt fixe, l'hélice A, et l'autre très flexible, la boucle FF' (acides aminés du goulot dans les

trois autres structures ouvertes sur le bloc 2). Trois de ces acides aminés sont aromatiques et se comportent comme des portes. C'est donc la boucle FF' qui orchestre la connexion du canal 2f de l'intérieur vers l'extérieur de la protéine. Pour 2V0M, le goulot du canal 2f est à  $CV < 6\text{\AA}$  ( $\pm 0.25\text{\AA}$ ). Une ouverture à  $CV = 5.75\text{\AA}$  permettrait l'ouverture et l'entrée du second ligand, alors que dans la structure apo, F215 obstrue l'ouverture du canal 2f. Le changement de conformation de F215 et L216 dans la boucle FF' de la structure holo 2V0M permet son ouverture. Les deux acides aminés Y53 et H54 sur l'hélice A ancrée dans la membrane plasmique ne présentent pas de changement significatif par rapport la structure apo, et sont ainsi supposés n'avoir aucun rôle dans l'ouverture du goulot d'étranglement (voir section 3.7.3).

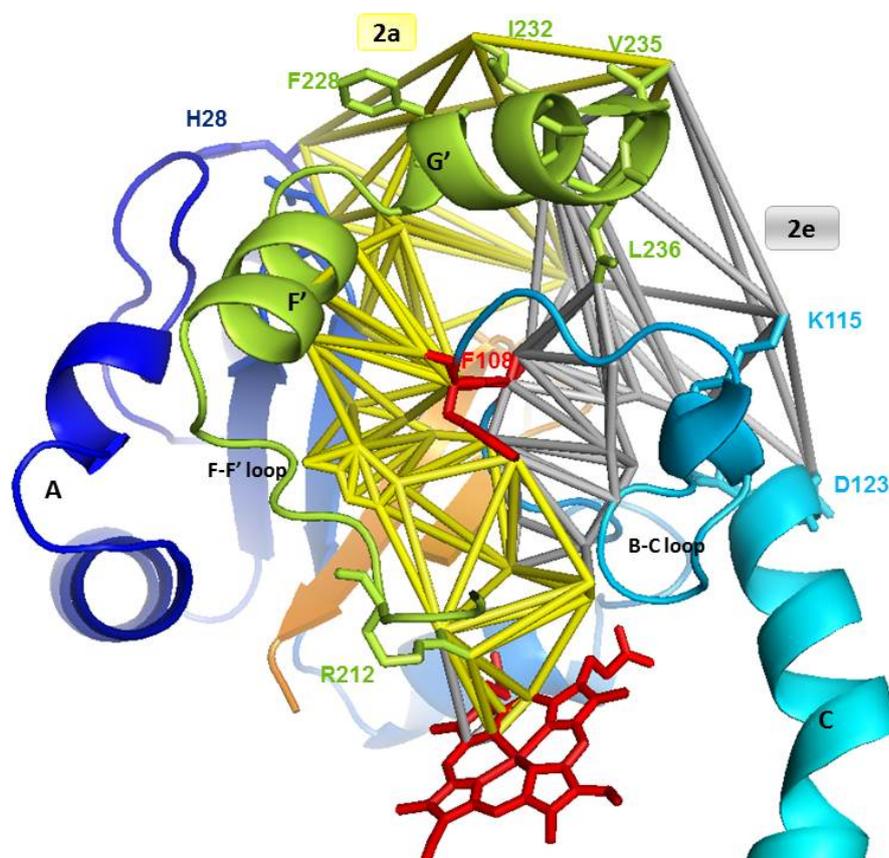
L'hélice F' borde le canal 2f sur tout le long par des acides aminés apolaires (P218, L221, F220, T224). La bouche du canal 2f est bordée par les acides aminés apolaires de la boucle FF' (225). F213 est dans le corps du canal 2f. On estime que l'ouverture du canal 2f est due à la déformation de la boucle FF', ou autrement dit, au changement de conformation des acides aminés de la boucle. Dans le cas de l'accueil d'une molécule hydrophobe, il y a un mouvement de F215 vers l'entrée (ou la bouche) du canal pour accrocher la molécule hydrophobe à la surface. Dans le cas d'une molécule hydrophile, on estime qu'il y a une rotation de F215 (hydrophobe) tout en restant dans le corps du canal 2f, afin d'adopter la forme ouverte pour augmenter d'espace nécessaire à l'ouverture de 2f.

### 3.2.3 Canal S

Le canal S, trouvé comme étant le plus flexible du 3A4, est face au cytosol et contient un acide aminé "porte" chargé menant le produit hors du site actif du 3A4 [Fishelovitch *et al.* 2009]. Le canal S, nommé canal du solvant [Cojocaru *et al.* 2007], se situe entre les hélices E, F et I et la boucle C-terminale. Il est détecté à  $CV = 5\text{\AA}$  dans la structure apo 1TQN (à  $1\text{\AA}$  de  $CV_{lim}$ ), et à  $CV_{lim} = 6,75\text{\AA}$  comme canal principal contenant les ligands dans la structure holo ouverte sur le bloc 2 (4K9U). Il a été signalé comme important dans plusieurs isoformes [Haines *et al.* 2001, Wester *et al.* 2003, Cojocaru *et al.* 2007]. Le canal du solvant est omniprésent chez les CYPs, et on pense qu'il régule l'accèsion de l'eau au site de liaison [Haines *et al.* 2001]. Le canal S a été proposé comme une route contrôlant l'accès et la sortie d'eau du site actif, suite à son observation dans P450-BM3 [Haines *et al.* 2001]. Il peut aussi être utilisé pour la sortie de substrat [Haines *et al.* 2001, Rowland *et al.* 2006]. Dans CYP102A1, il disparaît lorsqu'un substrat de grande affinité se lie [Haines *et al.* 2001]. Avec les mouvements normaux de respiration de la protéine, ceci suggère qu'un produit puisse facilement passer le long du canal S, et en effet celui-ci a été proposé comme passerelle principale d'accès au site actif du 2D6 [Rowland *et al.* 2006]. Des simulations SMD d'expulsion hors du 3A4 ont été faites pour deux produits : le temazepam et la  $6\beta$ -OH testostérone. Le canal S présente la plus grande ouverture pour ces deux produits, l'identifiant ainsi comme canal potentiel pour le substrat. Il a été trouvé comme étant le plus flexible dans le 3A4, et prédit comme étant un canal pour le substrat ou le produit chez d'autres P450s [Rowland *et al.* 2006]. De plus, il a été trouvé que, après oxydation, la forme géométrique du produit et son orientation déterminent son canal d'échappement préférentiel. Pour la majorité des canaux, les portes de contrôle sont des empilements  $\pi$  de phénylalanines. Pour sortir, le produit doit briser cette interaction [Fishelovitch *et al.* 2009].

Il a été suggéré dans le cas du CYP102A1 bactérien (CYPBM3) qu'il existe un canal de solvant contrôlant le passage de l'eau vers le site actif [Haines *et al.* 2001]. Ce canal, de rayon pouvant atteindre 2.2Å (équivalant à une épaisseur d'environ CV=5Å, CV auquel apparaît le canal S dans le 1TQN calculé par CCCPP), a été défini et observé pour beaucoup d'isoformes de P450 [Cojocaru *et al.* 2007], dont le 3A4 [Yano *et al.* 2004, Williams *et al.* 2004]. En considérant les changements conformationnels, on peut supposer que les substrats peuvent passer dans le canal du solvant. On peut aussi supposer que les molécules d'eau peuvent passer dans le canal du substrat. D'ailleurs, pour le CYP2D6 humain, une autre isoforme majeure métabolisant des médicaments, il a été proposé que ce canal du solvant soit aussi le principal canal d'accès du substrat [Rowland *et al.* 2006]. Une molécule d'eau dite "organisée" est une molécule d'eau formant au moins deux liaisons hydrogène avec une autre molécule d'eau, ou bien une liaison hydrogène avec une molécule d'eau et une avec l'enzyme (avec l'hème ou avec un canal). Généralement le site actif est rempli de molécules d'eau, mais peu sont organisées [Rydberg *et al.* 2007]. Par contre, les molécules d'eau présentes sont principalement regroupées à proximité de l'hème et s'étendent au canal du solvant et à l'aqueduc lorsqu'il est ouvert. Les molécules d'eau forment potentiellement un réseau de liaisons H connectant le site actif, les bouches de l'aqueduc et les canaux de solvant. Il y a aussi des molécules d'eau s'étendant dans la direction des autres canaux du substrat, mais elles sont moins organisées. Il a été montré que les voyages des protons dans et hors du site actif des enzymes sont toujours arbitrés par de l'eau [Wikström *et al.* 2005, Zhang 2007 & Bruice 2007]. Des simulations de dynamique moléculaire ont révélé deux organisations différentes des molécules d'eau dans le site actif. L'une s'étend du site actif à l'aqueduc lorsqu'il est ouvert, l'autre s'étend dans la direction générale du canal du solvant impliquant E308 et T309 (qui correspondent à D251 et T252 chez le CYP101A1 bactérien CYPcam). Cette thréonine promeut la seconde protonation de l'oxygène distal des entités peroxy du fer, formant ainsi un premier composé [Nagano & Poulos 2005]. Des études expérimentales du CYP101A1 ont permis la détection de structures aqueuses organisées dans le canal du solvant, en interaction avec T252 et formant une partie du système relayant le proton vers le dioxygène lié au fer [Schlichting *et al.* 2000, Nagano & Poulos 2005]. Il a été observé pour le CYP158A2 bactérien un amas d'eau organisé connectant le site actif à la masse de solvant pouvant participer au transfert de proton, mais que cet amas n'implique pas la thréonine conservée [Zhao *et al.* 2005]. Deux classes de P450s ont été ainsi suggérées, sur la base du mode de transfert de leur proton: l'une utilisant la thréonine conservée du canal du solvant, et l'autre utilisant les groupes hydroxy du substrat dans le site actif afin de stabiliser la voie d'eau pour fournir le proton. Ces résultats confortent l'hypothèse qu'une organisation des molécules d'eau peut coexister avec les substrats, ainsi que cela a été vu avec des simulations dans le cas du diazepam [Fishelovitch *et al.* 2010].

Le canal S a plus d'acides aminés polaires le bordant, de ce fait on lui attribue un comportement globalement polaire (voir plus loin, table 3.10). R212 borde le canal S, par exemple dans 1TQN. Dans les structures fermées, la conformation du F213 gênerait l'ouverture du canal S. Dans la structure 4K9T, le F215 gêne l'ouverture du canal S. Le canal S (figure 3.14) devrait être ouvert dans la structure 2J0D, mais c'est le trajet 2f, plus gros, qui est calculé, et les MCPs correspondant au trajet S ne sont pas calculés. Cela peut s'expliquer comme pour le canal 2f (voir section 3.2.2). Lorsque deux canaux existent à une même CV et partagent des tétraèdres communs, seul le plus gros trajet (correspondant au plus petit coût) est calculé. En effet, après le calcul du premier MCP (de



**Figure 3.13 :** Trajets 2a et 2e calculés par CCCPP respectivement à  $CV=6\text{\AA}$  et à  $CV=5.75\text{\AA}$  sur le CYP3A4 code PDB 1TQN. Le trajet 2a est en jaune, le 2e est en gris. Les deux bouches sont voisines, puis les canaux se séparent vers P107 et F108 dans la boucle BC. Les structures secondaires de 1TQN (feuillet  $\beta_1$ , régions FG et BC) sont en rubans.

coût minimal, voir section 2.3.1), à la nouvelle passe de calcul des MCPs la suppression des arêtes des précédents MCPs contenant des arêtes communes aux deux voies peut empêcher de trouver un MCP correspondant à la seconde voie, alors que ce second canal est visible dans le graphe facial et que l'autre trajet existe.

### 3.2.4 Canal 2e

Ce canal est retrouvé dans toutes les structures du 3A4 et dans les trois formes définies : fermée (C), ouverte au niveau du bloc 1 (O1), ou du bloc 2 (O2) : voir section 3.4. Le canal 2e est plus étroit que les précédents, et il indique généralement la valeur  $CV_{min}$ . Il est détecté vers  $CV=5.5\pm 0.5\text{\AA}$  et ne contient pas de ligand. C'est un canal plutôt commun, observé dans les structures de douze différents types de P450s [Cojocar *et al.* 2007]. Le canal 2e n'appartient ni au bloc 1 ni au bloc 2. Il passe à travers la boucle BC (figure 3.13), et son ouverture dépend vraisemblablement de la longueur et de la flexibilité de cette boucle, parmi d'autres facteurs.

Le canal 2e est détecté comme une route de sortie secondaire pour des substrats ou des produits dans les simulations de dynamique moléculaire de P450s [Cojocar *et al.* 2007], dans les simulations de P450eryF [Winn *et al.* 2002] et CYP2C5 [Schleinkofer

*et al.* 2005]. Il est difficile de conclure sur le rôle du canal 2e comme une issue de sortie tant qu'il n'y a aucune structure RX de complexe connue avec des produits d'oxydation. Des simulations de dynamique moléculaire informent sur les ouvertures et fermetures des tunnels des ligands, mais n'impliquent pas de simulation du processus de sortie du ligand. Pour le temazepam et la 6-hydroxytestostérone, les simulations montrent une préférence thermodynamique pour la sortie par le canal 2e. De plus, la 6-hydroxytestostérone présente un faible travail mécanique lors de la sortie par le canal 3, qui, tout comme le canal 2e, se trouve face au cytosol [Fishelovitch *et al.* 2009]. Lors de simulations de dynamique moléculaire et de simulations SMD, principalement concentrées sur le tunnel de sortie préféré du ligand, le canal 2e a été trouvé comme étant une sortie du temazepam dans 1TQN, associé à l'ouverture des canaux 2b et 3, et comme sortie de la 6 $\beta$ -hydroxytestostérone dans 2V0M [Fishelovitch *et al.* 2009, Krishnamoorthy *et al.* 2010].

Cela suggère l'élargissement de petits canaux ouverts (trajet 2e) pour la sortie du ligand et l'ouverture de canaux initialement fermés (trajets 2b et 3). Une hydratation contrôlée du substrat lié au site actif P450 est extrêmement importante pour la catalyse. Il est probable la fonction du canal 2e soit reliée à l'hydratation de site actif, bien qu'il puisse aussi avoir un rôle dans le transfert de proton. Le canal 2e reliant la masse du solvant à l'hème, il permet à l'eau de circuler et agit comme une pompe à eau. Ce canal serait emprunté par les molécules d'eau éjectées de la cavité catalytique lors de l'occupation du ligand. Le canal 2e est soupçonné d'être impliqué dans la circulation de petites molécules, d'eau et de protons, réactives dans le cycle d'oxydation du CYP. Il a été proposé qu'un canal d'eau fonctionne entre la boucle BC et le C-terminal de l'hélice B [Mustafa *et al.* 2014], ce qui correspond dans le 3A4 au canal 2e.

### 3.2.5 Topologie des canaux : résumé

La table 3.10 résume cette topologie. La majorité des atomes bordant les MCP sont les atomes des chaînes latérales des acides aminés, et non pas les quatre atomes du squelette (N, CA, C, O). Dans le 3A4, le cluster Phe est composé de 8 phénylalanines : 57, 108, 213, 215, 219, 220, 241 et 304.

Les canaux sont classés en fonction du rôle des structures secondaires dans le mécanisme d'ouverture/fermeture, à travers leurs déplacements [Cojocaru *et al.* 2007]. L'augmentation de la flexibilité des boucles BC et FG, qui est définie par leurs longueurs et leurs séquences, favorisent l'ouverture des canaux bordés par ces éléments structuraux. Ces mouvements sont inclus dans le mouvement global de la protéine, qui s'effectue par la liaison de la protéine lors du transfert d'électron au partenaire, comme dans le cas du CYP101 (ouverture de canaux d'accès lors de l'enclenchement du cycle catalytique) [Lüdemann *et al.* 2000a].

## 3.3 Position des canaux du CYP3A4 par rapport à la membrane

Le CYP est ancré dans la membrane plasmique à travers l'hélice N-terminale transmembranaire [Black *et al.* 1994]. Cet ancrage permet le recrutement des ligands hydrophobes. La liaison du domaine globulaire du 3A4 à la membrane reformate la protéine à l'interface membranaire, provoquant des changements conformationnels

Canal	Localisation			Fourchette de CV (Å)	Acides aminés de bordure du canal**			
	Nom*	Dans le CYP*	SRS			Extérieur		
Bloc1	2a	Sort entre les boucles FG et BC et le feuillet β1.	SRS-1	MP	<b>Commun aux 3 formes</b> 5.75 - 7	H28+ <u>F228</u> + <u>P231</u> + <u>V235</u> + <u>D380</u> + <u>K390</u> +Q79+R106+P107+ <sup>108</sup> +I223+T224+ <u>P227</u> +E374+ <u>R105</u> + <sup>304</sup> + <u>A370</u> + <u>R372</u>		
			SRS-2 C <sub>ter</sub> -hélice F Boucle FF'			<b>Forme C</b> 5.75 – 6.75	S29+ <sup>57</sup> + <u>Q78</u> + <u>I232</u> + <u>A305</u> + <u>T309</u>	<u>S119</u> + <u>L211</u> + <sup>213</sup> <u>K115</u> + <u>E122</u> + <u>D123</u> + <sup>213</sup> +G391+ <u>R212</u> + <u>I369</u>
			SRS-3 N <sub>ter</sub> -hélice G				<b>Forme O1</b> 6 – 6.25	<u>K35</u> + <u>L211</u>
						<b>Forme O2</b> 5.75 - 7		# <u>I120</u>
Bloc2	2f	Sort entre l'hélice F'/boucle F-G et la boucle C-terminale.  N'est pas bordé par la boucle BC.	SRS-2 C <sub>ter</sub> -hélice F Boucle FF'	Interface MP/ Cytosol	<b>Forme O2</b> 5.75 – 8.75	<u>F46</u> + <u>L47</u> + <u>I50</u> + <u>L51</u> +Y53+ <u>H54</u> + <u>K55</u> + <sup>57</sup> + <u>R105</u> +R106+ <sup>108</sup> + <u>S119</u> + <u>I120</u> +BoucleFF'+T224+ <u>F226</u> + <sup>241</sup> + <u>I301</u> + <sup>304</sup> + <u>A305</u> +E308+ <u>T309</u> + <u>I369</u> + <u>A370</u> + <u>R372</u> + <u>E374</u> + <u>E470</u> + <u>L479</u> +G480+ <u>G481</u> + <u>L482</u> + <u>E486</u> + <u>K487</u>		
			SRS-6			<b>Boucle FF'</b>	<u>2J0D</u> <sup>#</sup> : <u>R212</u> + <sup>213</sup> + <sup>213</sup> + <u>L216</u> + <u>D217</u> + <u>P218</u> + <sup>220</sup> + <u>L221</u> <u>2V0M</u> : <u>L211</u> + <sup>213</sup> + <sup>213</sup> + <u>L216</u> + <u>P218</u> <u>4K9T</u> : <sup>213</sup> + <u>L216</u> + <u>P218</u> + <sup>220</sup> + <u>L221</u> <u>4K9U</u> <sup>#</sup> : <sup>213</sup> + <sup>213</sup> + <u>L216</u> + <u>L221</u>	
Commun aux 3 formes	2e	Sort à travers la boucle BC.	SRS-4	Cytosol	<b>Forme O2</b> 6.75	<sup>108</sup> + <u>S119</u> + <u>K173</u> + <u>D174</u> + <u>Q200</u> + <u>E205</u> + <u>K208</u> + <u>K209</u> + <u>L211</u> + <u>R212</u> + <sup>213</sup> + <u>F220</u> + <sup>304</sup> + <u>E308</u> + <u>A370</u> + <u>K469</u> + <u>E470</u> + <u>L482</u> + <u>Q484</u> + <u>K487</u> + <u>P488</u>  <u>K168</u> + <u>D214</u> + <u>P218</u> + <u>T309</u> + <u>S312</u> + <u>I369</u> + <u>E486</u>		
			SRS-6			25+H28+ <u>K35</u> + <u>R105</u> + <u>R106</u> + <u>P107</u> + <sup>108</sup> + <u>V111</u> + <u>K115</u> + <u>S119</u> + <u>I120</u> + <u>E122</u> + <u>D123</u> + <u>E124</u> + <u>E125</u> + <u>L211</u> + <u>R212</u> + <sup>213</sup> + <sup>213</sup> +T224+ <u>P227</u> + <u>F228</u> + <u>P231</u> + <u>E234</u> + <u>V235</u> + <u>I301</u> + <sup>304</sup> + <u>A305</u> + <u>A370</u> + <u>D380</u> + <u>K390</u> +G391+ <u>L482</u>		

**Table 3.10** : Localisation des canaux significatifs du CYP3A4 calculés par CCCPP, avec leurs caractéristiques. Les acides aminés du site actif sont en gras. Les acides aminés des bouches sont surlignés. Les acides aminés apolaires sont en vert, les polaires en bleu. Les acides aminés du cluster Phe sont en exposant. Forme O1 : forme ouverte au niveau du bloc 1. Forme O2 : forme ouverte au niveau du bloc 2. Forme C : forme fermée. \* Nomenclature et localisation selon [Cojocaru *et al.* 2007]. \*\* Voir table 1.2 pour les SRS et l'appartenance aux structures secondaires. # Boucle reconstruite.

pertinents pour les canaux d'accès [Baylon *et al.* 2013]. Les CYPs liés à la membrane insèrent leur hélice N-terminale et abaissent leurs hélices hydrophobes F' et G' dans la bicouche lipidique. L'insertion de cette partie de la région FG dans la membrane faciliterait l'ouverture du canal 2a vers la membrane plasmique. Ceci est cohérent avec l'hypothèse que 2a est le principal canal pour les substrats hydrophobes. Les entrées des canaux restent très probablement dans les mêmes régions. Le produit dans le site actif ayant des propriétés différentes de celles du substrat, s'échapperait par d'autres voies, probablement par un canal conduisant vers la région du groupe de tête polaire. En raison de l'association des CYPs mammaliens aux bicouches lipidiques, on s'attend que l'accueil du substrat et la sortie du produit suivent des voies différentes, dépendant de l'affinité de la molécule avec l'environnement extérieur, lipophile (membrane), ou hydrophile (cytosol), ou intermédiaire, à l'interface membrane/cytosol [Berka *et al.* 2013].

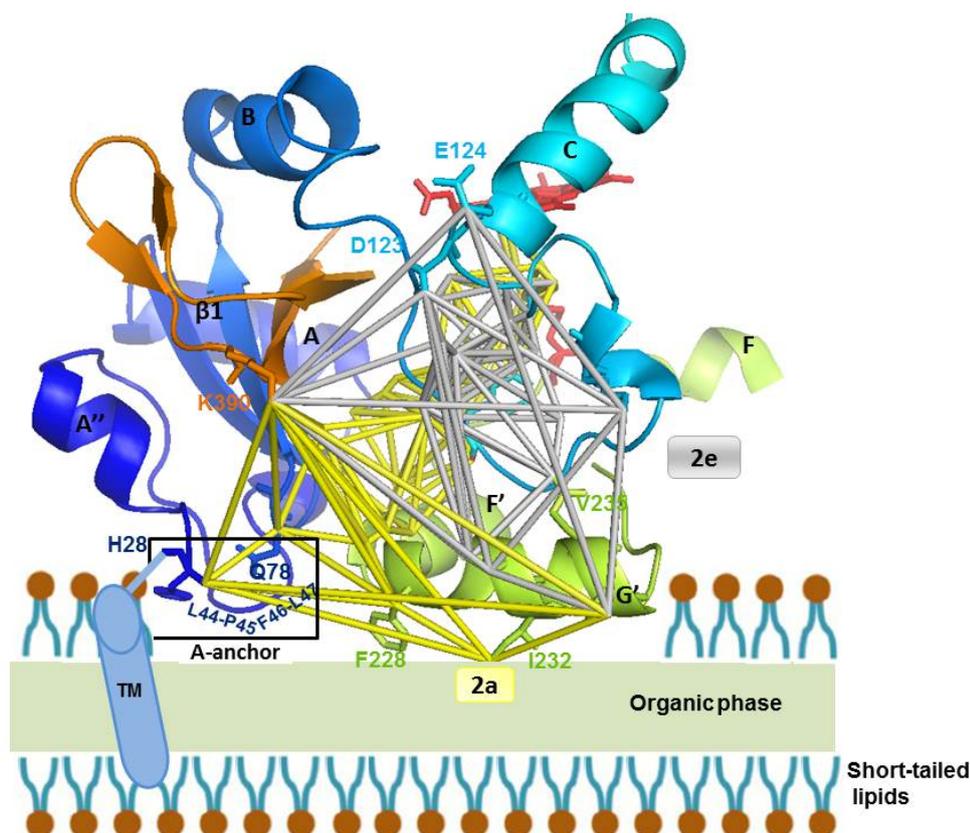
Il existe un modèle structural du 3A4 membrane lié à la membrane compatible avec la plupart des données expérimentales [Denisov *et al.* 2012]. Cette analyse comparative a permis de suggérer que l'insertion F'G' des P450s eucaryotes dans la membrane intervient dans le repositionnement sélectif du domaine  $\beta$ . Ce déplacement du domaine  $\beta$  chez les eucaryotes étend le volume de la cavité catalytique et ouvre un canal situé entre la boucle BB', et les feuillets  $\beta_1$  et  $\beta_3$  correspondant au canal 2a se déversent dans la membrane. Au contraire, dans les P450s solubles, l'accès au site actif par les substrats se fait principalement par un canal formé par la boucle FG et l'hélice B'. Comparé à son équivalent soluble chez les CYPs bactériens, le domaine  $\beta$  des CYPs eucaryotes est déplacé vers le plan de l'hème, et en conséquence, la chaîne latérale propionate de l'hème est orientée vers le côté proximal (voir figure 4 dans [Sevrioukova & Poulos 2013a]).

Les canaux proximaux sont plus petits que les canaux distaux. On retrouve les canaux distaux 2a, 2f, S et 2e (figure 3.14), qui sont décrits dans la section 3.2, et dont le la desserte à la surface de la protéine est indiquée dans la table 3.10.

### 3.4 Conformations ouvertes ou fermées : les trois états du CYP3A4 associés à l'ouverture des canaux

Les 18 structures RX à haute résolution du 3A4 présentes dans la PDB ont permis de comparer les conformations en fonction de leur liaison au substrat. Nous les étudions à travers le calcul des réseaux de canaux du CYP3A4 sous les formes apo et holo. L'objectif est de comprendre la sélectivité à travers les chemins de passage, eux-mêmes liés à la structure des ligands et à l'affinité physico-chimique avec les acides aminés du CYP.

L'interaction CYP-Ligand est régie par leur affinité mutuelle. Cette affinité est physicochimique, mais dépend aussi des contraintes stériques. Deux hypothèses se confrontent concernant le mécanisme d'ouverture des canaux des P450 : "forme induite (induced fit)" vs. "sélection conformationnelle". Elles se différencient selon que la liaison du ligand au 3A4 se forme avant ou après le changement conformationnel dans la protéine [Sevrioukova & Poulos 2012b]. L'interaction (préalable ou postérieure) du ligand avec le 3A4 déformé, a lieu par affinité physico-chimique avec le ligand. De ce fait, le changement conformationnel du 3A4 lié diffère selon le ligand. C'est ce qui détermine la sélectivité, la liaison au substrat et le taux de métabolisation. L'importance de la lipophilie des substrats dans la liaison CYP-Ligand et le taux de métabolisation ont été pointés [Lewis & Ito 2010] (voir section 1.8).



**Figure 3.14** : CYP (ancré dans la membrane plasmique), avec les canaux 2a, 2e, 2f et S calculés par CCCPP.

Contrairement à l'hypothèse initiale ayant motivé le développement du programme CCCPP, le CV du ligand est indépendant du CV du canal emprunté. En effet, la voie de passage principale correspond à celle du plus grand CV. Cette observation a été faite sur graphe facial calculé par CCCPP, où les ligands co-cristallisés apparaissent avec le graphe facial associé au plus grand CV du canal connectant l'hème à l'extérieur. Le calcul par CCCPP des chemins au sein des canaux a confirmé cette hypothèse, en situant les ligands précisément dans les trajets délimités par les acides aminés des structures secondaires bordant ces trajets. Il apparaît que le ligand ne pénètre pas par des canaux "à sa juste mesure", ou, autrement dit, le ligand ne sélectionne pas le canal d'accès correspondant à sa dimension, mais choisit plutôt de passer par le plus gros canal, même s'il existe d'autres canaux plus petits et de dimension plus proche de celle du ligand. Cette observation nous a menés à choisir les plus gros canaux calculés par CCCPP comme canaux d'accès. Et ainsi nous explorons les canaux pertinents par taille décroissante, en repérant les plus gros canaux puis en diminuant le CV, avec un pas de  $0.25\text{\AA}$  pour le 3A4. Les calculs de canaux sont terminés lorsque CCCPP propose des canaux proximaux connus pour ne pas conduire le ligand au site actif, même si leur rôle a bien été identifié dans la littérature. La valeur finale de CV est  $CV_{min} \sim 5\text{\AA}$  pour les différentes structures étudiées du 3A4. Nous estimons ainsi avoir les canaux pertinents pour le transit des molécules dans la fourchette  $CV_{min} - CV_{lim}$ .

L'exploration des canaux des 18 structures (2 apo et 16 holo) du 3A4 par CCCPP, nous permet d'observer deux voies principales de passages pour l'entrée des ligands,

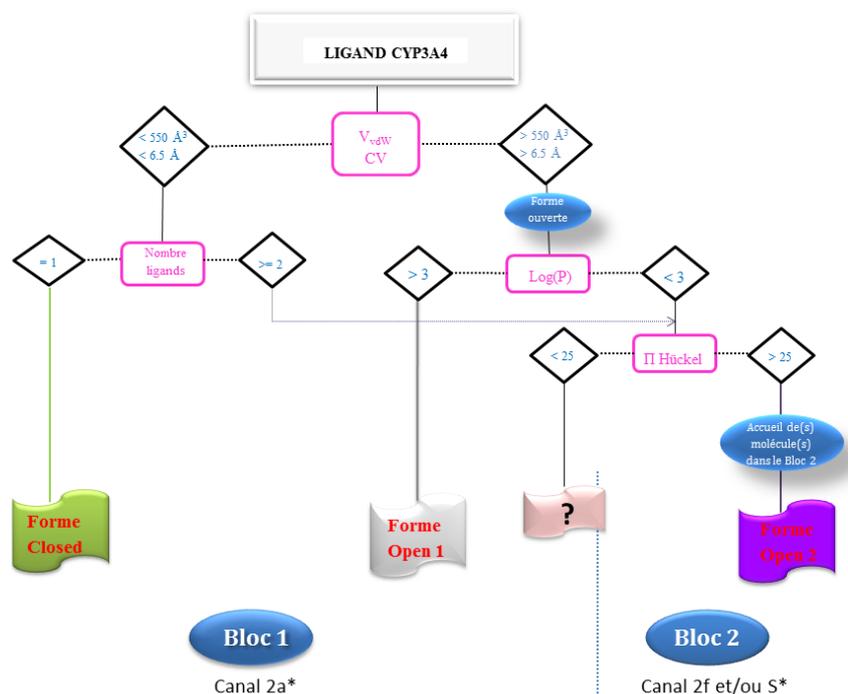
définissant deux blocs. L'ouverture de ces voies est caractérisée par l'ouverture des plus gros canaux situés dans deux régions du CYP précédemment classées pour tous les isoformes, par [Yu *et al.* 2013]. Nous appellerons ces 2 régions **bloc 1** et **bloc 2**, en référence à la classification de [Yu *et al.* 2013]. Le bloc 1 correspond à une crevasse entre les régions BC et FG (avec ouverture des canaux 2a, 2ac et 2c), alors que le bloc 2 correspond à une fissure entre la région FG et la boucle C-terminale (avec ouverture des canaux 2f et S).

Les conformations sont classées dans la littérature comme fermée (C) ou ouverte (O) [Cojocar *et al.* 2007, Yu *et al.* 2013]. Nous avons déterminé avec CCCPP que les structures RX du 3A4 révèlent 1 conformation fermée et 2 ouvertes. Au sein même du bloc 1, nous avons identifié deux types de conformations, relatives à la conformation de la boucle FF'. Elles sont nommées forme "*Closed*" et forme "*Open 1*", et notées respectivement **C** et **O1**. A l'ouverture du bloc 2, une conformation différente de celle du bloc 1 est adoptée, que nous avons nommée forme "*Open 2*", et notée **O2**. Ces changements s'opèrent dans la région de la zone 2, qui attribue à sa conformation l'état du 3A4 : C, O1 ou O2. Les conformations ouvertes sont labellisées O1 ou O2, suivant que ce soit respectivement le bloc 1 ou le bloc 2 qui s'ouvre. L'ouverture du bloc 2 définit l'état O2, que le le bloc 1 soit ouvert ou non. Il y'aurait donc 2 états possibles pour la forme O2 : soit un accueil de molécule(s) dans le bloc 2 uniquement, ou bien un accueil dans le bloc 2 ET dans le bloc 1. La région des canaux d'accès C et O1 se déverse vers la membrane plasmique. La région des canaux O2 s'ouvre vers l'interface cytosol/membrane plasmique via le canal 2f, et vers le cytosol via le canal S. Ainsi, les calculs effectués CCCPP nous ont permis de définir trois familles de conformations adoptées par l'isoforme 3A4 pour accueillir le ligand (figure 3.15). Ces conformations dépendent du ligand et sont détaillées ci-dessous.

### 3.4.1 Diagramme de prédiction des ouvertures du CYP3A4

A partir des diverses conformations du 3A4, nous avons défini trois formes structurales du 3A4 en fonction de la conformation de la boucle FF' (figure 3.15). Ces trois formes sont associées à l'ouverture de canaux déterminés. Les formes apo sont signalées comme le sous-type 1 "*Closed*" [Cojocar *et al.* 2007, Yu *et al.* 2013]. Pour les formes liées (holo), les conformations sont marquées C, O1, ou O2, suivant le ou les ligands en interaction (figure 3.15). Suivant notre définition, nous considérons que pour un petit ligand ( $CV < 6.5\text{\AA}$ ) c'est d'abord le canal 2a qui s'ouvre (état C), et pour un plus grand (ou un second) ligand, une déformation de la boucle FF' s'opère. Dans ce second cas nous passons à la forme ouverte, puis en fonction des comptabilités physico-chimiques, soit la déformation est petite et nous sommes en forme O1, auquel cas le canal 2a s'élargit d'avantage par rapport à la position en forme apo, soit la déformation est grande auquel cas le bloc 2 s'ouvre tandis que le bloc 1 est toujours ouvert mais reste parfois inaccessible car il présente un goulot d'étranglement, c'est la forme O2 (voir figure 3.15).

Nous présentons les structures étudiées sous leurs trois formes énoncées ci-dessus, puis nous explicitons plus en détails l'étude structurale qui distingue les trois formes. D'une manière générale, l'exploration des canaux d'accès ou de sortie dans le 3A4 a révélé l'ouverture d'un canal principal de plus grande dimension. Selon la conformation de la protéine, ce trajet est le canal 2a pour l'ouverture en bloc 1, ou bien le canal 2f (et parfois canal S) qui se situe dans le bloc 2. Ces deux principaux canaux sont supposés être les canaux d'accès car ils logent les ligands co-crisallisés. Ensuite, pour une même valeur de



**Figure 3.15** : Diagramme des formes du CYP3A4 en fonction du paramètre structural et physicochimique du ligand. CV définit l'épaisseur du ligand. \* Canal ouvert et accueillant le substrat.

CV ( $\pm 0.5\text{\AA}$ ), inférieure de  $1.5\text{\AA}$  environ à  $CV_{lim}$  du canal 2a dans la forme C initiale, il y a un canal ouvert pour toutes les conformations de l'isoforme 3A4, c'est le canal 2e. Ce canal est supposé être le canal de sortie, ou bien jouant un rôle dans la circulation des molécules d'eau et/ou des protons acteurs de l'oxydation du CYP (voir section 3.2.4).

Les 18 conformations du 3A4 cristallisées dont nous disposons dans la banque PDB correspondent à différents états : interaction avec un ou deux ligands (table 3.11), ou aucun. Ces changements conformationnels peuvent être corrélés aux différences entre canaux (ouverture, dimension), en fonction du ligand. Il y a parmi les CYPs une forte variation en séquence et en structure autour des régions BC et FG. En général, allant de la boucle à l'hélice puis au feuillet, la flexibilité diminue tandis que la rigidité augmente. Plus la région a une longue boucle, plus elle devrait être flexible. De plus, la longueur des régions BC et FG et leur composition en acides aminés peut aussi influencer sur leur flexibilité. L'influence exacte de la flexibilité dans les changements conformationnels ou dans la conversion entre les états fermés et ouverts, est toujours inconnue [Skopalik *et al.* 2008]. Les éléments de structures secondaires flexibles dans les régions BC et FG subissent généralement les changements conformationnels les plus importants entre des structures ouvertes et fermées, pour permettre l'entrée et l'accueil de composés volumineux.

Le 3A4 entreprend les changements conformationnels importants, dans la région FG et autour du cluster Phe. Les positions de la boucle BC, de l'hélice I et de la boucle C-terminale sont aussi altérées, de façon plus ou moins importante selon la structure. Les variations structurales importantes entre les formes liées du 3A4 sont observées uniquement dans les boucles FF' et C-terminales, bordant la paroi supérieure de la cavité de liaison du substrat (voir figure 5S dans [Sevrioukova & Poulos 2013b]). Donc

Forme	Code PDB	Ligand	Log(P)	H	V	$n_r$	$n_m$
C	1W0F	Progestérone	3.58–4.15	4.10	321		
C	1W0G	Métyrapone	2.03	4.18	215		
C	3UA1	Bromoergocryptine	3.20–3.89	6.02	546	5	
C	4K9V	GS6 <sup>a</sup>					
C	4I3Q	Eau					
C	4NY4	2QH <sup>b</sup>	4.28		389		
C	3TJS	DTMCR <sup>a</sup>	4.27		560		
O1	3NXU	Ritonavir	4.24–5.22	7.39	663		
O1	4I4G	GS2 <sup>a</sup>	5.44	6.57	645		
O1	4I4H	GS3 <sup>a</sup>	6.36	6.69	663		
O1	4K9W	GS7 <sup>a</sup>	5.68		618		
O1	4K9X	GS8 <sup>a</sup>					
O2	4K9T	GS4 <sup>a</sup>	2.22	5.96–6.37	479	14	2
O2	4K9U	GS5 <sup>a</sup>	3.06	5.84–7.89	513	14	2
O2	2V0M	Kétoconazole	4.03–4.19	5.19–6.52	449	7	2
O2	2J0D	Erythromycine	2.37–2.60	7.21	727	7	

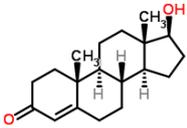
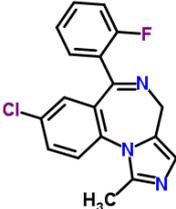
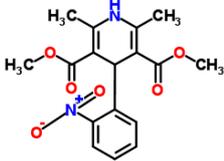
**Table 3.11** : Les 16 structures holo étudiées du CYP3A4. <sup>a</sup> Analogue du ritonavir (voir section 3.6). <sup>b</sup> Dérivé de la pyridine (voir [Bränden *et al.* 2014]). H : épaisseur du ligand en Å. V : volume de van der Waals du ligand en Å<sup>3</sup>.  $n_r$  : nombre de liaisons pivotables.  $n_m$  : nombre de molécules accueillies.

l'hétérogénéité structurale du domaine catalytique du 3A4 semble provenir exclusivement de la flexibilité conformationnelle de ces deux éléments de surface [Sevrioukova & Poulos 2013b]. Le mouvement important de la région FG et du cluster Phe détermine le passage de la forme fermée aux formes ouvertes. Les hélices F et G sont tordues [Yu *et al.* 2013]. La définition des trois formes du 3A4 est liée précisément à la conformation de la boucle FF', qui contient un site de reconnaissance des plus importants dans la sélectivité du 3A4, le SRS-2 (voir table 1.2). On voit donc que la sélection des substrats est liée à la déformation de la boucle FF', qui œuvre comme un clapet, induisant l'ouverture des canaux du 3A4. La dynamique (ouverture/fermeture) des deux blocs est ainsi essentiellement liée au mouvement principal de la boucle flexible FF', plus ou moins coûteux énergétiquement. La position relative de la boucle FF' détermine le degré d'ouverture de la protéine (C, O1, O2), en liaison avec le changement de conformation des acides aminés bordant les canaux. La forme ouverte s'établit avec le déplacement de l'hélice F', et avec l'écartement de l'ensemble de la boucle FG de la boucle BC et du feuillet  $\beta_1$ . La conformation est classée O1 avec l'élargissement du canal existant, ou bien O2 avec l'ouverture d'un ou plusieurs nouveaux canaux. La boucle FF' est orientée vers l'intérieur de la protéine en forme fermée, et vers la surface en formes ouvertes. Lors du mouvement de la boucle FF', un réarrangement des acides aminés s'opère, libérant l'espace de passage d'un bloc (modification spatiale), et délimitation des acides aminés en bordure du bloc d'entrée (modification chimique). Le passage en forme O2 requiert une déformation importante de la boucle FF', qui est déstabilisée; cette région désordonnée n'est en général pas définie en structure RX.

Pour permettre au ligand d'accéder à l'hème, l'acide aminé du site actif R212 adopte différents conformations. R212 est orienté vers l'intérieur de la cavité en forme C, et peut réorienter sa chaîne latérale en adoptant différents rotamères (dans le squelette protéique, le groupe amide de R212 reste inchangé). Cette rotation de la chaîne latérale de R212 n'affecte pas la conformation de la boucle FF'. En revanche, l'ouverture de la structure du 3A4 en formes O1 ou O2 est associée à un changement conformationnel de R212 vers l'extérieur de la cavité, par un mouvement de son groupe amide (dans le squelette), accompagné de la déformation de la boucle FF'. R212, orienté vers l'extérieur de la cavité catalytique, peut se trouver en surface de la protéine ou bien rester dans le cœur de la protéine.

Néanmoins, des mouvements concertés moins importants, et d'autres parties de la protéine accompagnent les mouvements principaux cités ci-dessus dans l'adaptation de la protéine à l'accueil du ligand. Ces mouvements annexes participent à la stabilité du complexe en optimisant l'interaction protéine-ligand (voir section 3.6.2). Le passage de la forme C apo à la forme C holo se fait par le mouvement de la région de la boucle BC qui ouvre le canal 2a et augmente le volume du site actif.

La modélisation structurale du 3A4 par CCCPP nous a permis, d'une part de localiser les voies de passage dans la protéine, et d'autre part d'observer l'influence des acides aminés habillant ces voies à travers leur interaction avec le ligand. Nous avons établi le diagramme de classification des formes du CYP3A4 en fonction des critères stéréo-électroniques, à partir des données sur les 18 structures 3A4 étudiées. Le critère géométrique est basé sur les dimensions des ligands complexés. La limite d'épaisseur du ligand à 6.5Å correspond à la dimension maximale en deça de la dimension maximale d'ouverture du canal 2a en forme C (qui est de 6.75Å pour la bromoergocryptine : voir section 3.7.1). La limite du volume de van der Waals du ligand à 550Å<sup>3</sup> est relative au volume du site actif en forme fermée (520Å<sup>3</sup> pour 1W0E [Williams *et al.* 2004]). L'énergie de Hückel est l'énergie du système  $\pi$  des parties conjuguées de la molécule, acycliques ou cycliques (aromatiques). Les interactions par empilement  $\pi$ - $\pi$  avec les acides aminés aromatiques "portes" ont un rôle important dans le mécanisme de coopérativité dans le 3A4 [Fa *et al.* 2015]. Nous ajoutons que la propriété de conjugaison est privilégiée lors du processus de métabolisation, car elle attribue à la molécule la propriété d'hydrosolubilité, plus facilement éliminable. Lors de l'interaction avec le CYP, ces molécules donneraient plus facilement des substitutions que des additions qui détruiraient la conjugaison. Le nombre de molécules reflète le rapport de concentrations ligand/enzyme, ou la coopérativité par une molécule d'effecteur. Nous avons introduit section 1.5 les travaux sur le 3A4 qui ont mené à l'identification de trois groupes principaux de comportement d'inhibition [Kenworthy *et al.* 1999, Kenworthy *et al.* 2001]. Cette méthode permet de cribler les interactions médicamenteuses potentielles. Elle est basée sur l'évaluation de l'activité d'une nouvelle molécule avec le 3A4 par des réactions sondes représentatives d'une activité spécifique du 3A4. Cette classification, basée sur des données expérimentales, définit trois familles non compétitives de substrats (types testostérone, midazolam, et nifédipine : voir table 3.12 et figure 3.15).

	Testostérone	Midazolam	Nifédipine
Formule brute	C <sub>19</sub> H <sub>28</sub> O <sub>2</sub>	C <sub>18</sub> H <sub>13</sub> ClFN <sub>3</sub>	C <sub>17</sub> H <sub>18</sub> N <sub>2</sub> O <sub>6</sub>
Structure			
Forme supposée	C	C <sup>a</sup> → O2 <sup>b</sup>	O2 <sup>b</sup>
Site de liaison <sup>a,d</sup>	Périphérique	Bloc 1 (canal 2a)	
Site de liaison <sup>b,d</sup>	Bloc 1 (canal 2a)	Bloc 2 (canal 2f)	Bloc 2 (canaux 2f, S)
V <sub>vdW</sub> <sup>c,d</sup>	293.41	266.54	299.95
Log(P) <sup>c,d</sup>	3.37	3.97	1.82
TPSA <sup>c</sup>	37.30	30.18	107.77
Energie π de Hückel <sup>c,d</sup>	9.99	37.19	43.74

**Table 3.12** : Paramètres de 3 substrats prototypes du CYP3A4. V<sub>vdW</sub> : Volume de van der Waals en Å<sup>3</sup>. TPSA : *Topological Polar Surface Area*, mesure la contribution des atomes polaires à la surface moléculaire, en Å<sup>2</sup>. <sup>a</sup> Une seule molécule accueillie. <sup>b</sup> Accueil supposé de 2 molécules ou de la seconde molécule. <sup>c</sup> Paramètres calculés avec Chemaxon (<https://www.chemaxon.com>). <sup>d</sup> Paramètres pris en compte pour la classification des formes selon le diagramme figure 3.15.

Notre étude nous permet d'apporter des précisions quant à cette classification biochimique. En se basant sur la sélectivité par canalisation, nous établissons cette hypothèse de classification, et il serait nécessaire d'utiliser ces données pour valider expérimentalement notre approche in-silico.

L'hypothèse des 3 familles de substrats prototypes est illustrée par le diagramme de la figure 3.15 : l'énergie π de Hückel d'une molécule définirait la possibilité ou non d'accueil de la molécule dans le bloc 2. Le bloc 2 accueillerait des molécules d'énergie π de Hückel dépassant 25kcal/mol. Pour une molécule peu volumineuse (volume de van der Waals inférieur à 550Å<sup>3</sup> ou épaisseur < 5.5Å) avec log(P) > 3, la molécule est accueillie dans le bloc 1 en forme C. Si l'énergie π de Hückel de cette molécule dépasse 25kcal/mol, l'accueil d'une deuxième molécule est possible dans le bloc 2 (exemple, le midazolam). Si l'énergie π de Hückel de cette molécule est inférieure à 25kcal/mol, l'accueil dans le bloc 2 n'est pas possible, et le bloc 1 est trop petit pour l'accueil de 2 molécules, et donc seule une molécule est accueillie par le 3A4 en forme C (exemple, la testostérone). Lorsque la molécule a un log(P) < 3 ET que l'énergie π de Hückel de cette molécule dépasse 25kcal/mole, la molécule est automatiquement accueillie dans le bloc 2, avec possibilité d'accueil de 2 molécules (exemple, la nifédipine), toujours dans le bloc 2. Dans le cas où l'énergie π de Hückel de cette molécule est inférieure à 25kcal/mol, la molécule est accueillie dans le bloc 1 en forme C (car volume de van der Waals < 550Å<sup>3</sup> ou épaisseur < 5.5Å : voir table 3.12).

Le nombre de molécules accueillies dans le 3A4 influence la cinétique adoptée par cette enzyme. A faible concentration du substrat, le taux de métabolisation du midazolam reflète l'activité du 3A4 et suit une simple cinétique Michaelis-Menten. A forte concentration, le taux change, avec une cinétique atypique [Kronbach *et al.* 1989].

Les profils d'interaction observés entre les trois molécules sondes impliquent plusieurs caractéristiques d'inhibition atypiques (partielle, coopérative, concentration-dépendante), représentés par une gamme de modèles cinétiques [Galetin *et al.* 2003]. L'utilisation commune de la réaction d'hydroxylation en 1' du midazolam est une bonne réaction sonde pour évaluer l'activité du 3A4 à de faibles concentrations de substrats [Yuan *et al.* 2002]. D'après notre étude avec CCCPP, et d'après les paramètres stéréo-électroniques du midazolam (voir diagramme figure 3.15), cette réaction commune représente la réaction "standard" du 3A4 en forme C (voir 3.7.1). L'accueil d'une molécule supplémentaire fait adopter à l'enzyme des changements de formes qui s'accompagnent du changement de cinétique. On assimile le cas du midazolam au cas du kétoconazole, étudié ici (voir section 3.7.3). Des hypothèses d'accueil multi-substrats ont été proposées, suite à des études expérimentales effectuées sur les interactions entre les 3 familles de substrats prototypes. Les analyses de cinétique multi-sites supportent l'hypothèse de domaines de liaisons distincts pour chaque substrat prototype, alors que les analyses des interactions inter-substrats (coopérativité hétérotrope) indiquent l'existence d'un domaine de liaison mutuel commun pour chacune des 3 sous-classes [Galetin *et al.* 2003]. Nous avons vu que chaque canal comporte des SRS (table 3.10). La classification que nous proposons ici (figure 3.15), est basée soit sur l'accueil de substrat multi-sites ou sur des sites communs de liaison. L'accueil multi-sites réside en 2 sites dans 2 blocs distincts, ou bien 2 sites distincts au sein du même bloc ou canal (voir 4K9T en section 3.6.6). Cela concerne l'accueil de ligands (effecteur ou substrat) de familles distinctes, ou bien l'accueil de plus d'une molécule de la même famille. L'autre accueil réside en un site mutuel pour des ligands de familles différentes accueillis dans même canal. Des sites de liaison hors des blocs 1 et 2 existent aussi, comme le site périphérique dans le cas de la progestérone (voir 3.7.1).

La complexité de l'interaction enzyme-substrat du 3A4 au niveau moléculaire, rend difficile la prédiction in-vitro des interactions médicamenteuses in-vivo [Ueng *et al.* 1997]. Notre étude in-silico tente d'apporter une connaissance des canaux à l'échelle moléculaire, et sera plus développée dans les sections suivantes.

### 3.4.2 Forme *Closed* : C

Cette forme est caractérisée principalement par la conformation de la boucle FF' dans la forme apo, orientée vers l'intérieur de la protéine. Dans l'état C, la protéine ouvre toujours le canal principal 2a. Aussi bien en forme libre que liée, les CYPs peuvent avoir un état C, que nous définissons comme un état de "repos" de la protéine lorsque la protéine n'est pas complexée. La plupart des structures cristallisées de CYPs mammaliens sont dans un état C [Yu *et al.* 2013]. Les structures du 3A4 concernées sont les formes apo 1TQN et 1W0E, 4I3Q, et les formes holo indiquées dans la table 3.11. La cavité du site actif est petite.

En forme C, seuls des réarrangements mineurs ont lieu pendant la liaison aux molécules volumineuses. Les structures étudiées en forme C sont bien superposées sur toutes les parties de la protéine. La conformation de la boucle FF' est identique, ainsi que la conformation de ses acides aminés, sauf pour l'acide aminé du site actif R212 qui occupe différents rotamères sans changement dans le squelette peptidique associé, et sauf pour F304 dans 4NY4. Ces deux derniers acides aminés changent de conformation pour éviter les collisions avec les ligands.

Les rotamères de R212 sont orientés vers l'intérieur ou l'extérieur de la cavité catalytique (4 rotamères différents observés). Lorsque R212 est orienté du côté du site actif, celui-ci peut être supposé participer à la réactivité du ligand avec l'hème. R212 est orienté vers l'extérieur, du côté opposé au site actif dans 4K9V. F304 dans 4NY4 s'ouvre comme une porte pour libérer la place au ligand volumineux. L211 est orienté vers l'extérieur, alors que F213 et F215 sont orientés vers l'intérieur de la cavité. F215 borde le trajet 2a, au niveau du corps, et R212 les bordent au niveau du site actif. L'orientation de R212 vers le site actif a été révélée comme étant importante lors d'une réaction de déshydrogénation du métabolite 4-hydroxy-tamoxifen (4OHT), en imposant une inhibition stérique et électrostatique. De ce fait, le conformère 1TQN et non 1W0E, présente le mode catalytiquement productif favorisant la réaction [Sevrioukova & Poulos 2013a].

### 3.4.3 Forme *Open 1* : O1

Les structures du 3A4 concernées sont résumées dans la table 3.11. Lors du passage de la forme C à la forme O1, il y a élargissement du canal 2a et élargissement du site actif par décalage de la boucle FF' vers l'extérieur du site actif et mouvement de la boucle BC. Ce sont les mêmes canaux ouverts que dans la forme C, mais avec une conformation plus ouverte de la boucle FF', conséquence de l'interaction avec le ligand. En effet, lors de la pénétration d'un ligand volumineux (volume de van der Waals supérieur à  $550\text{\AA}^3$ ), la boucle FF' s'éloigne de sa position initiale pour éviter les collisions avec les acides aminés de la boucle FF', adoptant ainsi la forme O1 (voir section 3.6).

Nous supposons que le changement conformationnel est induit par la répulsion de la chaîne latérale de R212 vers l'extérieur par le ligand. R212 bascule vers l'extérieur la surface, du côté opposé au site actif, qui induit l'ouverture de la boucle FF' avec libération de l'espace de la cavité catalytique initialement occupée par la chaîne latérale de R212. La réorientation de R212 hors du site actif, supposée intervenir dans la réactivité du substrat dans les formes C, est observée dans le cas de l'interaction du 3A4 avec des inhibiteurs volumineux et très hydrophobes (voir section 3.6). R212 change alors de conformation et entraîne les acides aminés voisins de la boucle. Le mouvement de R212 se concerte avec un basculement de L211 vers l'intérieur, et un espacement des deux acides aminés portes F213 et F215 afin de libérer de l'espace pour le ligand, tout en se maintenant à l'intérieur. R212 ne borde plus le trajet 2a et n'intervient plus sur le site actif. Cette fois, dans la structure O1, c'est F213 qui borde le trajet 2a au niveau du corps (et non plus F215 comme c'est le cas dans la forme C), et L211 le borde au niveau du site actif. La déformation de la boucle FF' est accompagnée par le décalage vers le haut de la cavité de L211, F213, F215 et L216. L'hélice I se décale par l'éloignement de A305 et de l'acide aminé porte F304 pour libérer l'espace du ligand.

Le basculement de R212 est dû à une plus grande dimension des ligands (volume de van der Waals supérieur à  $550\text{\AA}^3$  et/ou épaisseur supérieure à  $6.5\text{\AA}$ ), qui induit des collisions avec les acides aminés de la boucle FF', principalement R212, F213 et F215 (voir section 3.6.3. La proximité trop forte du ligand et des acides aminés de la boucle induisent une répulsion. Cette petite déformation de FF' est moins coûteuse en énergie pour la protéine que celle en forme O2. La forme O1 reste bien définie car la conformation de la boucle FF' est stabilisée par une liaison hydrogène entre le groupe guanidine de R212

et l'oxygène du carbonyle de K208 (situés dans SRS-2) : voir figure 8B dans [Sevrioukova & Poulos 2012a].

### 3.4.4 Forme *Open 2* : O2

Les formes holo O2 présentent une importante plasticité des régions BC et FG, qui permet l'ouverture de deux fissures distinctes. L'une d'elles correspond au canal 2a, alors que l'autre inclut les canaux 2f et S. Ce comportement a été retrouvé dans le CYP2B4, à la différence de l'ouverture des canaux 2c et 2ac dans le bloc 1 à la place du canal 2a dans le 3A4 [Scott *et al.* 2003, Scott *et al.* 2004, Zhao *et al.* 2006]. Les structures concernées sont résumés dans la table 3.11. Pour les 4 structures étudiées en forme O2, il y a ouverture du canal 2f avec ou sans ouverture du bloc 1. Pour la structure 4K9U, il y a ouverture du canal S. Cependant il faut avoir une réserve quant à cette dernière structure car la boucle FF' a été reconstruite sur 6 acides aminés manquants (voir section 2.5 et table 2.2). Dans la forme O2, il y a un élargissement plus important du site actif avec le mouvement de la boucle FF' de l'intérieur de la protéine vers la surface. La boucle FF' s'ancre dans la membrane plasmique pour le recrutement des ligands (par affinité physicochimique avec les acides aminés). La boucle FF' est moins stable dans la forme O2 par rapport à la forme O1 et C.

R212 s'oriente vers la surface et s'éloigne du site actif, qui s'élargit. Dans le cas du 2V0M, les deux blocs sont ouverts, avec ouverture des canaux d'entrée calculés 2a et 2f. Ces canaux sont communs au niveau du corps et du site actif, et ne se séparent en fourche qu'à la sortie, au niveau de T224. Le tronc commun 2a/2f est bordé par L211 et F213. F215 dans les formes C et O1. Il est orienté vers l'intérieur et gêne l'ouverture du canal 2f. Contrairement à la forme O1, pour deux molécules hydrophobes (2V0M), F215 bascule à l'extérieur pour ouvrir le canal 2f, et au niveau de la bouche. Ceci est supposé faciliter le recrutement de la molécule hydrophobe dans le canal 2f. Pour deux molécules hydrophiles (4K9T), F215 se décale par rapport à la position en forme C pour permettre l'ouverture du canal 2f, mais reste à l'intérieur de la protéine. Le mécanisme d'ouverture du canal 2f pour une molécule hydrophile est proposé ainsi : basculement de L211 et de R212 vers l'extérieur, et de F215 (hydrophobe) vers l'intérieur du canal 2f (et non pas vers la bouche).

L'ouverture du bloc 2 expliquerait la coopérativité du 3A4, car elle permet l'accueil d'une deuxième molécule, identique ou différente de la première. Les deux molécules sont logées dans deux canaux différents, ou bien dans un même canal. Le canal du bloc 2 s'élargit et peut accueillir deux molécules. Dans la forme C, le canal 2a s'ouvre dans le cas d'une molécule hydrophobe de petite dimension (volume de van der Waals inférieur à  $550\text{\AA}^3$  et/ou CV inférieure à  $6.5\text{\AA}$ ). S'il y a une seconde molécule, c'est le bloc 2 qui s'ouvre pour des raisons stériques. En revanche, dans le cas de molécules peu hydrophobes ( $\text{Log}(P) < 3$ ), que ce soit une grosse molécule (volume de van der Waals supérieur à  $550\text{\AA}^3$  et/ou CV supérieure à  $6.5\text{\AA}$ ), ou bien deux molécules, c'est le bloc 2 qui prend en charge l'accueil de ces molécules. L'ouverture du bloc 2 induit un changement conformationnel important qui serait coûteux en énergie pour la protéine, la faisant sortir de sa conformation d'équilibre. Une telle déformation est due à une incapacité du ligand de pénétrer par le canal 2a en raison de sa grande dimension et sa faible hydrophobicité. Ceci est dû à des raisons stériques (voir le cas de l'érythromycine ou la deuxième molécule de kétoconazole), ou bien à des raisons physico-chimiques dues à une faible lipophilie de

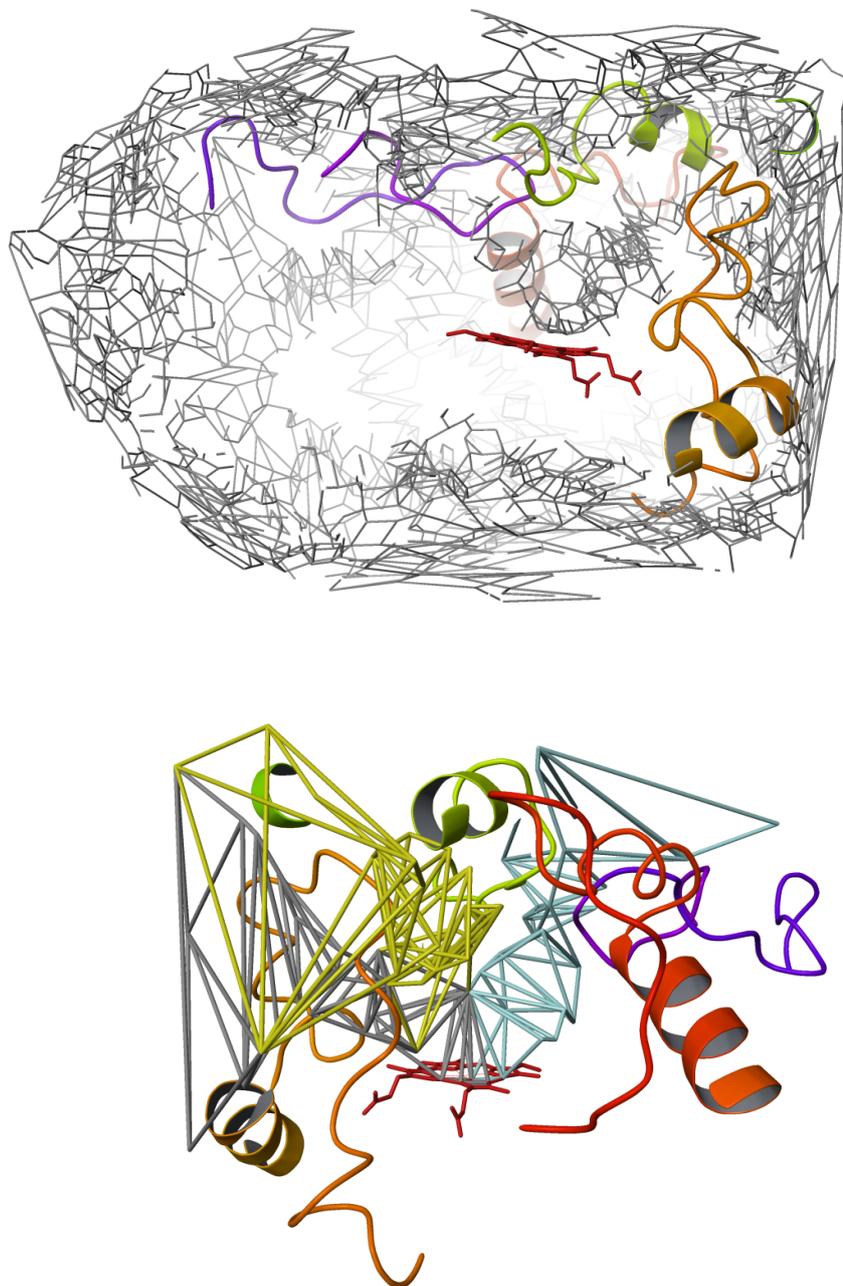
la molécule (voir les molécules de désoxyritonavir dans 4K9T et 4K9U : voir sections 3.6.6 et 3.7.3). On peut supposer qu'une molécule faiblement hydrophobe mais de petite dimension passe par le canal 2a, car cela est plus favorable pour la protéine que d'ouvrir le bloc 2, comme c'est le cas par exemple de la métyrapone (volume de van der Waals égal à  $215\text{\AA}^3$ , épaisseur égale à  $4.2\text{\AA}$ ,  $\text{Log}(P)=2.03$ ).

## 3.5 Conformation de la forme apo du CYP3A4

Deux structures de la forme apo sont publiées, codes 1TQN [Yano *et al.* 2004] et 1W0E [Williams *et al.* 2004]. Elles sont très similaires [Scott & Halpert 2005]. Nous le vérifions avec le logiciel ARMS [Petitjean 1998], en libre accès sur la même page web que CCCPP (<http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>) : sur les 447 carbones  $\alpha$  du squelette commun à 1TQN et à 1W0E, l'algorithme estime que 439 d'entre eux constituent le motif 3D commun maximal, avec une déviation RMS= $0.775\text{\AA}$ . Les caractéristiques les plus remarquables sont les courtes hélices F et G, et un large cœur hydrophobe structuré par des phénylalanines au-dessus du site actif. Le volume accessible au solvant est estimé à  $950\text{\AA}^3$  par [Ekroos & Sjögren 2006]. Le volume du site actif de la forme apo est plus petit que l'on pourrait attendre, comparable à celui du CYP2C8 et du CYP2C9, ce qui a conduit à supposer des changements conformationnels dus à la liaison avec le ligand [Williams *et al.* 2004]. Une troisième structure apo, 4I3Q [Sevrioukova & Poulos 2013c], comprend une molécule d'eau liée à l'hème. Grâce à CCCPP nous avons pu voir que ces changements de volume de cavité sont liés à l'ouverture des canaux (formes O1 et O2). Dans le 3A4 en forme apo, il y a ouverture du bloc 1, en forme C, c'est-à-dire ouverture du canal 2a.

Une grande cavité correspondant à la sortie du canal 2a calculé à  $CV_{lim}=6\text{\AA}$  pour le 1TQN, relativement large, s'étend à l'extérieur le long de la surface du feuillet  $\beta_1$  sous l'hélice F', entouré par ces deux structures secondaires, et par SRS-1 (boucle BC). Un canal plus petit calculé à  $CV_{lim}=5.75\text{\AA}$ , 2e, quitte la cavité du site actif de l'autre côté des acides aminés 106 à 108. Un troisième canal, défini comme étant celui du solvant S [Cojocar *et al.* 2007], apparaissant à  $CV=5\text{\AA}$ , quitte la cavité du site actif sous la région SRS-2 entre SRS-4 (hélice I) et SRS-6 (boucle C-terminale) : voir figure 3.16. Ces trois derniers canaux, repérés et identifiés par CCCPP, ont été présentés comme canaux du solvant [Yano *et al.* 2004], avant l'établissement de la nomenclature de [Cojocar *et al.* 2007]. Les interactions entre les acides aminés stabilisent les ouvertures des canaux. R106 (boucle BC) s'étend à travers une arête du canal 2a et participe à un réseau de liaisons hydrogènes formé par les acides aminés polaires Y53, D61, D76, R106, R372 et E374, stabilisé par des interactions de charge (figure 3.17).

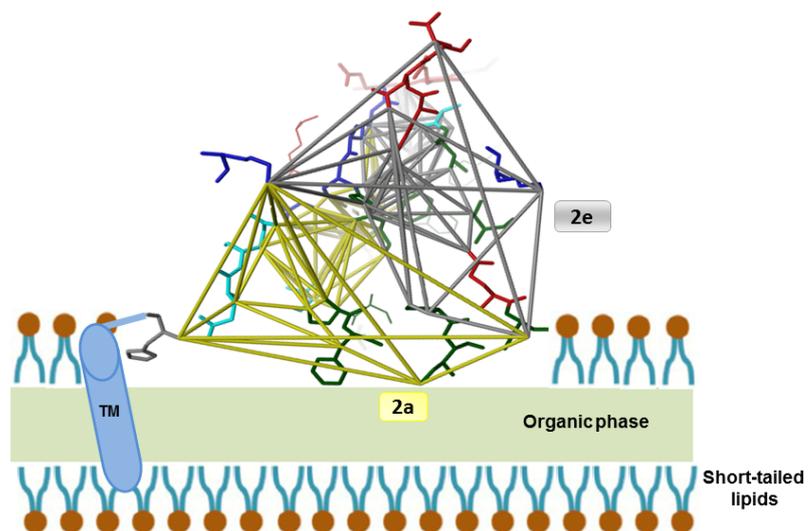
Nous avons pu déterminer avec CCCPP que les acides aminés de la boucle BC sont ceux qui participent à l'élargissement du canal 2a pour adopter la forme O1 (voir section 3.6). Dans le 1W0E, il y a ouverture des canaux 2a et 2e pour la même valeur  $CV_{lim}=5.75\text{\AA}$ . Au repos, la dimension du canal 2a est équivalente à celle du canal 2e. On peut suggérer que c'est l'accueil du substrat qui provoque la déformation des structures secondaires du 3A4, induisant l'élargissement du canal 2a. En revanche le canal S n'est pas calculé pour 1W0E, et son ébauche n'est pas visible sur le graphe facial, contrairement à 1TQN (figure 3.18). On peut supposer que c'est dû à la respiration de la protéine, qui



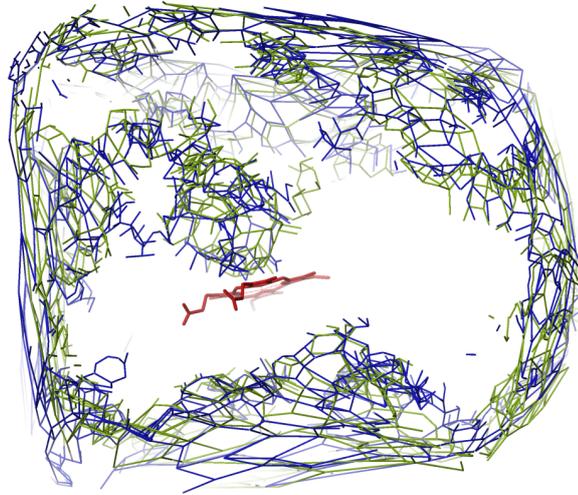
**Figure 3.16** : Graphe facial du CYP3A4 (code 1TQN) à  $CV_{lim}=6\text{\AA}$ . En haut, structures secondaires bordant les canaux 2a, 2e, et l'ébauche du canal S. En bas, trajets 2a (en jaune), 2e (en gris), et S (en bleu), calculés à  $CV=5\text{\AA}$ ; la vue du bas est en rotation de  $90^\circ$  par rapport à celle du haut.

ferme et ouvre des petits canaux comme le canal S au repos. R212 (boucle FF') s'étend à travers une arête du canal S.

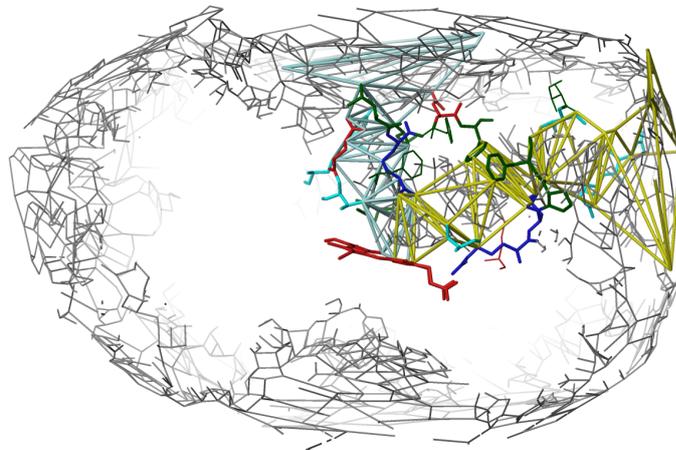
Les mutations L211F et D214E (acide aminés plus encombrés aux mêmes propriétés : influence géométrique uniquement), situées sur le bord extérieur de la région SRS-2 (sur la boucle FF'). Les chaînes latérales de L211 et D214 dans la forme C sont orientées vers



**Figure 3.17** : Trajets 2a, 2e et S du CYP3A4 (code 1TQN) calculés à  $CV=5\text{\AA}$ . Les trajets 2a sont en jaune, les 2e en gris, et les S en bleu. Les acides aminés en bordure sont coloriés suivant leurs propriétés physico-chimiques. En haut, bouches des trajets 2a et 2e, ancrés dans la membrane plasmique. Au milieu, rotation de  $-90^\circ$ . En bas, rotation de  $+90^\circ$ .

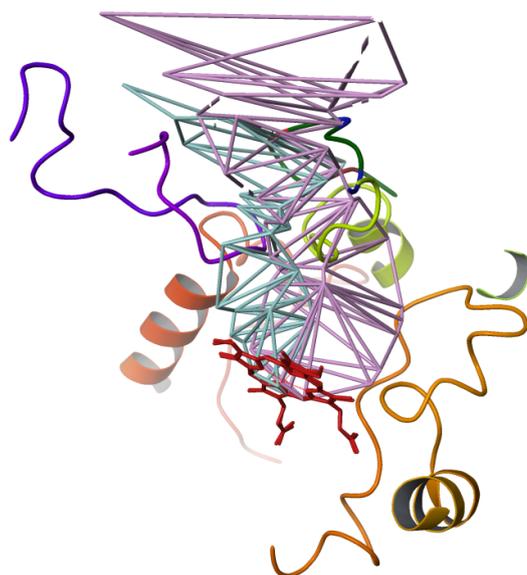


**Figure 3.18** : Graphe facial à  $CV_{lim}$  du 1TQN (en vert) et du 1W0E (en bleu).



**Figure 3.19** : Graphe facial du CYP3A4 (code 1TQN) à  $CV_{lim}=6\text{\AA}$  : superposition des trajets 2a et S. Le trajet 2a est en jaune, le S est en bleu. Les acides aminés bordant les trajets (en tubes épais) et les canaux (en tubes minces) sont coloriés suivant leurs propriétés physico-chimiques.

l'extérieur et ne bordent pas le trajet 2a. En revanche elles bordent l'entrée et la bouche du canal S (bloc 2 : voir figure 3.17). Elles modifient la coopérativité homotropique dans l'hydroxylation de stéroïde catalysé par 3A4 et confère une dépendance hyperbolique à la concentration de substrat [Harlow & Halpert 1998] : la mutation sur le canal S gêne l'ouverture du bloc 2 à  $CV < CV_{lim}$ , ce qui appuie l'hypothèse de la participation des canaux du bloc 2 dans la coopérativité du 3A4. L'hypothèse émise par [Yano *et al.* 2004] de la réduction du volume du site actif est celle de l'augmentation du volume occupé



**Figure 3.20 :** Trajets S pour le CYP3A4 en forme apo (code 1TQN) et holo (code 4K9U). Pour le 1TQN, le trajet S, en bleu, est ouvert à  $CV=5\text{\AA}$  (petit site actif sous la boucle FF'). Pour le 4K9U, le trajet S, en violet, est ouvert à  $CV_{lim}=6.75\text{\AA}$ ; le trajet S et le site actif sont élargis par le mouvement de la boucle FF, et bordés par les structures secondaires de la forme apo; la boucle FF' de la forme holo est en couleurs foncées, suivant les propriétés physico-chimiques des acides aminés.

par les chaînes latérales des acides aminés mutés. Mais les acides aminés en question n'étant pas directement dans le site actif mais plutôt en surface de la protéine (figure 3.17), il est difficile d'établir une relation entre mutation et changement de volume du site actif. Notre étude des canaux avec CCCPP, nous permet de proposer l'hypothèse d'une gêne d'ouverture des canaux du bloc 2 due à la diminution du volume du canal S. Cette hypothèse est d'autant plus justifiée que les chaînes latérales de ces acide aminés sont orientées à l'extérieur du site actif entourant l'entrée du canal S, opérant comme un goulot d'étranglement qui bloquerait l'ouverture du canal S (figure 3.19), et donc empêcherait la coopérativité par le blocage de l'accueil de la molécule. Ceci justifierait mieux le rôle de ces mutations dans le changement de la conformation de protéine. L211 et D214 dans la forme C sont orientés vers l'extérieur et ne bordent pas le trajet 2a.

Il a été aussi démontré que les mutations des acide aminés supplémentaires dans la région SRS-2, comme L210, F213 et F215 de la boucle FF', qui forment le toit du site actif, affectent la coopérativité de l'enzyme [Harlow & Halpert 1997, Domanski *et al.* 2001]. Nous soupçonnons que la structure atypique de la région formant le toit de la cavité du site actif, bordant le tronc commun des canaux 2a/2f, pourrait satisfaire la liaison simultanée de plusieurs molécules de substrat en raison de sa flexibilité, liée à celle de la boucle FF': présence d'une fourche d'ouverture du canal 2f sur région la du toit du site actif au niveau de T224.

Une autre caractéristique inhabituelle du 3A4 est la présence de R212 dans la cavité du site actif. Dans la forme apo, la chaîne latérale est placée vers l'intérieur pour donner

une liaison hydrogène avec le carbonyle du squelette de F304, un acide aminé impliqué aussi dans la coopérativité de l'enzyme [Domanski *et al.* 2001] (acide aminé du site actif sur l'hélice I, bordant les canaux 2a, 2f et 2e), aussi bien que dans un réseau de liaisons l'hydrogène qui inclut un amas de molécules d'eau au-dessus de l'hème. Dans cette position (R212 est orienté vers l'intérieur), il est possible que le squelette des acide aminés L211 et R212 perturbe la voie de transfert d'un éventuel proton entre les acide aminés conservés E308 et T309, sur l'hélice I et sur SRS-4. En effet, E308 est lié par un hydrogène à l'amide du squelette de R212.

L211, R212, E308 et T309 bordent le canal S (figures 3.17 et 3.19). Ces interactions stabilisent l'ouverture du canal S. La structure de l'enzyme suggère que la chaîne latérale de R212 (sur FF', très flexible), peut facilement se réorienter pour résider à l'extérieur de la cavité du site actif. Nous avons conclu de notre étude sur plusieurs structures du 3A4 que cette réorientation de R212 a une cause stérique (passage en forme ouverte : voir figure 3.15). Le repositionnement de la chaîne latérale de R212 vers l'extérieur permettrait le placement d'une molécule d'eau entre E308 et T309 pour le transfert de protons entre ces deux acides aminés, ce phénomène participant au mécanisme catalytique [Yano *et al.* 2004]. La liaison au substrat pourrait influencer l'orientation de la chaîne latérale de R212, laquelle influence la réactivité du 3A4 (figure 3.20).

Les structures 1W0E, 1W0F et 1W0G, en forme C, ont révélé un petit site actif, avec peu de changement conformationnel associé à la liaison de chaque composé. Plusieurs acides aminés du cluster Phe (108, 213, 215, 241 et 304), contribuent à l'enveloppe qui obture le toit du site actif au-dessus de l'hème entre les régions SRS-1, SRS-2, SRS-3 et SRS-4 [Yano *et al.* 2004]. Le cluster Phe se trouve au-dessus du site actif, avec les chaînes latérales aromatiques empilées pour former un important cœur hydrophobe. En se basant sur les valeurs des facteurs B, cette région apparaît très ordonnée, comparée à l'ensemble de la structure [Williams *et al.* 2004]. De plus, en conséquence de cet amas aromatique, le site actif du 3A4 a volume accessible d'environ  $520\text{\AA}^3$  [Williams *et al.* 2004], considérablement plus petit de ce qu'on attendrait au vu de la métabolisation de substrats volumineux. Ceci suggère l'élargissement du site actif pour l'accueil des grandes molécules. Le site actif de la forme C est petit. Le mouvement de FF' dans la forme O agrandit le volume du site actif pour accueillir de plus gros ligands ( $CV > 6.5\text{\AA}$ , et volume de van der Waals  $> 550\text{\AA}^3$  : voir figure 3.15).

La conformation fermée (C) du site actif du 3A4 détermine la position du cluster Phe, qui est semblable à la conformation fermée adoptée par P450 EryF [Cupp-Vickery & Poulos 1995], mais l'arrangement spatial de l'hélice F de la boucle FG n'est pas conservé. La spécificité du 3A4 réside dans sa boucle FF'. Le mouvement conformationnel impliquant le cluster Phe pourrait notamment replacer des phénylalanines, aboutissant peut-être à une extension de l'hélice F et un plus grand site actif. Un tel mécanisme a été observée dans la structure du CYP119 dans lequel, bien qu'il n'y ait aucun analogue du cluster Phe, la liaison du ligand aboutit à l'extension de l'hélice F, à un mouvement de l'hélice G et à l'intervention de la boucle F-G [Park *et al.* 2002].

Le volume global du site actif CYP3A4 est plus petit et de forme différente de celui de CYP2C9 (PDB code 1OG2). La variation de topologie du site actif est une conséquence du cluster Phe (constitué des phénylalanines des boucles FF' et BC très flexibles, l'hélice F dans 2C9 étant moins flexible que la boucle FF' du 3A4) : le cluster Phe est placé sur le sommet du site actif plus près de l'hème, avec, comparé à CYP2C9, le feuillet  $\beta 1$  plus loin de l'hème. L'hème de CYP3A4 a une accessibilité plus grande au site actif

que celui de CYP2C9, qui pourrait permettre à deux molécules de substrat d'accéder à l'oxygène réactif, en accord avec des données indiquant que le 3A4 peut lier et métaboliser simultanément de multiples molécules de substrat. De par la flexibilité de la boucle FF' dans le 3A4, celui-ci adopte la forme la plus compacte dans la forme C, et s'adapte pour accueillir des molécules de caractéristiques différentes.

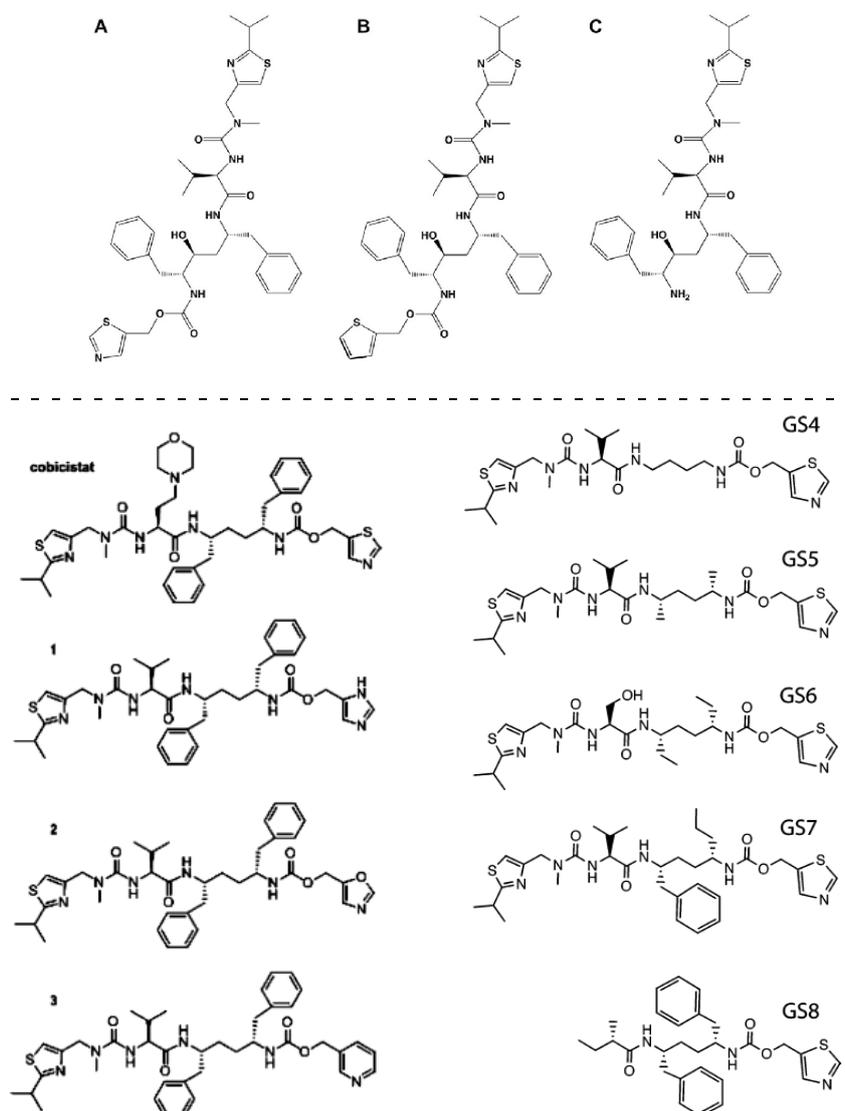
En résumé, la structure de 3A4 en forme C expose une cavité de liaison au substrat relativement grande, avec une déformabilité qui la rendrait compatible avec les tailles des substrats oxydés par l'enzyme. De plus, la taille relativement grande de la cavité près du centre catalytique de l'enzyme, de par la flexibilité de la boucle FF', peut contribuer à la coopérativité hétérotropique en facilitant des modes de liaison alternatifs pour de multiples substrats, en ouvrant le bloc 2. En plus de l'ouverture de deux blocs, le second de ces blocs étant plus grand et accueillent de très gros ligands et une seconde molécule. La grande taille de la cavité reflète des différences significatives des conformations des régions SRS-1, SRS-2 et SRS-5 du 3A4, comparé à la famille 2 des P450s. Les mutations dans la région SRS-2, qui n'est pas hélicoïdale dans le 3A4 (située sur la boucle FF'), altèrent généralement les cinétiques homo et hétérotropiques liant la conformation inhabituelle de l'enzyme dans cette région aux aspects caractéristiques de la fonction du 3A4 : il y a un effet sur l'ouverture des canaux du bloc 1 ou du bloc 2. Ceci prouve que la coopérativité du 3A4 est liée au mouvement de la boucle FF' dans l'ouverture des canaux des bloc 1 et bloc 2.

## 3.6 Inhibition du CYP3A4 par les analogues du ritonavir

### 3.6.1 Le ritonavir et ses analogues

Le ritonavir (figure 3.21) est un peptidomimétique inhibiteur de la protéase HIV qui inactive l'isoforme 3A4 [Sevrioukova & Poulos 2010]. Le mécanisme d'interaction du CYP3A4 avec des analogues du ritonavir (figure 3.21) a été étudié [Sevrioukova & Poulos 2012a], et un modèle de pharmacophore a été établi [Kaur *et al.* 2015].

Les principales caractéristiques structurales nécessaires à une forte inhibition du 3A4 ont été identifiées [Sevrioukova & Poulos 2010]. En revanche, le ritonavir présente des effets indésirables, et il est nécessaire de développer de nouveaux inhibiteurs ayant moins d'effets secondaires, en améliorant leurs propriétés pharmacochimiques et physicochimiques. Les analogues du ritonavir donnent lieu à une liaison N-Fe (comme le ritonavir), mais leur efficacité inhibitrice varie en fonction des interactions non covalentes avec les phényles et avec les groupements aux extrémités : tête thiazolique et isopropylthiazole terminal (noté IPT). D'où l'hypothèse de l'importance de ces groupements dans la liaison avec les analogues ritonavir. Les désoxyritonavirs ont le squelette du ritonavir mais sans le groupement hydroxyle. La suppression de ce groupement confère plus de flexibilité à la molécule, qui est plus apte à pénétrer et à se loger dans le 3A4. Le squelette du désoxyritonavir est flexible, permettant une plus forte liaison Fe-N, et il adopte une conformation minimisant les collisions stériques et optimisant les interactions protéine-ligand, en particulier avec S119. Il a été montré que l'interaction du 3A4 avec ces inhibiteurs est modulée par R212 [Sevrioukova & Poulos 2012a]. Cet acide aminé du site actif est dans la boucle FF' (acides aminés 209 à 219, contenant la région SRS-2), et subit



**Figure 3.21** : Structure du ritonavir et de ses analogues. En haut, de gauche à droite : structures du ritonavir, du déaza-ritonavir (DAR), et du desthiazolylmethyloxycarbonyl ritonavir (DTMCR). En bas, colonne de gauche, de haut en bas : structures du ritonavir, du cobicistat, et des analogues du ritonavir GS1, GS2 et GS3 [Sevrioukova & Poulos 2013c]. En bas, colonne de droite, de haut en bas : structures des déoxyritonavirs GS4 a GS8 [Sevrioukova & Poulos 2013b].

un important changement conformationnel pour lier des molécules volumineuses telles que le ritonavir, le kétoconazole, et l'érythromycine [Ekroos & Sjögren 2006, Sevrioukova & Poulos 2010] (voir sections 3.6.1 et 3.7.3). Dans la forme O1 du 3A4, la boucle FF' est déplacée (acides aminés 212 à 218) par rapport au site actif, mais reste bien définie car sa conformation est stabilisée par une liaison hydrogène entre le groupe guanidine de R212 et l'oxygène du carbonyle de K208 (voir figure 8B dans [Sevrioukova & Poulos 2012a]).

Les données cinétiques, mutagénétiques et structurales sur les interactions du 3A4 avec les dérivés du ritonavir ont aidé à concevoir ses analogues [Sevrioukova & Poulos 2012a], et permis d'en établir les caractéristiques importantes :

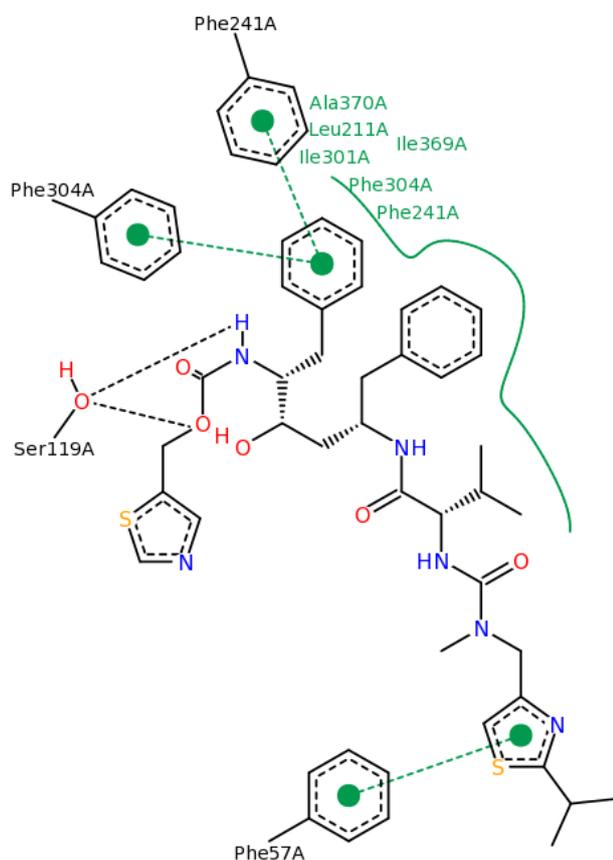
- Un atome d'azote exposé du groupe thiazolyl, qui lie l'inhibiteur au fer héminique.
- Une liaison forte et irréversible à l'hème (côté distal) pour une meilleure inhibition du 3A4.
- Un groupement pyridine donne lieu à une liaison Fe–N plus forte qu'un groupement thiazole [Sevrioukova & Poulos 2013c].
- La coordination de l'hème conduit à associer des ligands de type II (c'est à dire en liaison covalente avec le fer héminique), pendant que les forces hydrophobes définissent le mode de liaison.
- Les interactions non covalentes avec les groupes hydrophobes latéraux et terminaux influencent le mode de liaison au site actif [Sevrioukova & Poulos 2013a].
- Le remplacement d'un phényle du ritonavir par un groupe hydrophobe plus petit pourrait éviter une collision stérique et donc augmenter l'affinité et le pouvoir inhibiteur.
- Le squelette du désoxyritonavir est plus flexible, adoptant une conformation qui optimise les contacts entre protéine et ligand.
- Les acides aminés du site actif contribuent à lier les composés examinés, et peuvent ainsi jouer un rôle de primordial dans la transformation métabolique de xénobiotiques.
- La liaison secondaire périphérique du ritonavir limite la coordination de l'hème et complexifient la cinétique de coordination.
- Capacité à inhiber la catalyse du P450.

### 3.6.2 Relations structure-activité des analogues du ritonavir

La structure des analogues du ritonavir influe comme suit : (i) influence de la diminution de taille et nature hydrophobe des groupements par substitution/élimination, (ii) diminution de taille ou élimination des groupements latéraux hydrophobes (phényle et isopropyle), élimination du groupement final IPT, et élimination du groupement hydroxy du squelette, (iii) interactions non covalentes des groupements latéraux déterminant l'affinité, régie par le processus de liaison et l'orientation de la liaison.

Les relations entre la structure de la protéine et l'activité des analogues du ritonavir apparaissent à plusieurs niveaux :

1. Il y a déformation de trois régions du 3A4 pour l'accueil des molécules. L'espace adjacent aux deux phényles est toujours occupé, et doit donc être un site de liaison très important. Le mode de liaison évite les collisions stériques avec les deux phényles, et favorise les interactions cation– $\pi$  et  $\pi$ – $\pi$ . Les structures secondaires ainsi que les acides aminés entourant ces deux poches sont déplacés pour permettre l'accueil des molécules.
2. Présence d'une poche hydrophobe au niveau du premier phényle (noté Ph-1), stabilisant le ligand par des interactions hydrophobes, et qui augmente le pouvoir inhibiteur. Les ligands essaient de remplir cette poche, même si cela provoque une gêne stérique :
  - Le Ph-1 est implanté dans une poche hydrophobe bordée d'acides aminés ayant une conformation modifiée, stabilisant le ligand par des interactions hydrophobes : F108 (boucle BC), L210, L211 (boucle FF'), F241 (boucle G'G dans SRS-3), I301, F304 (hélice I), L482 (boucle  $\beta$ 4 contenant SRS-6).



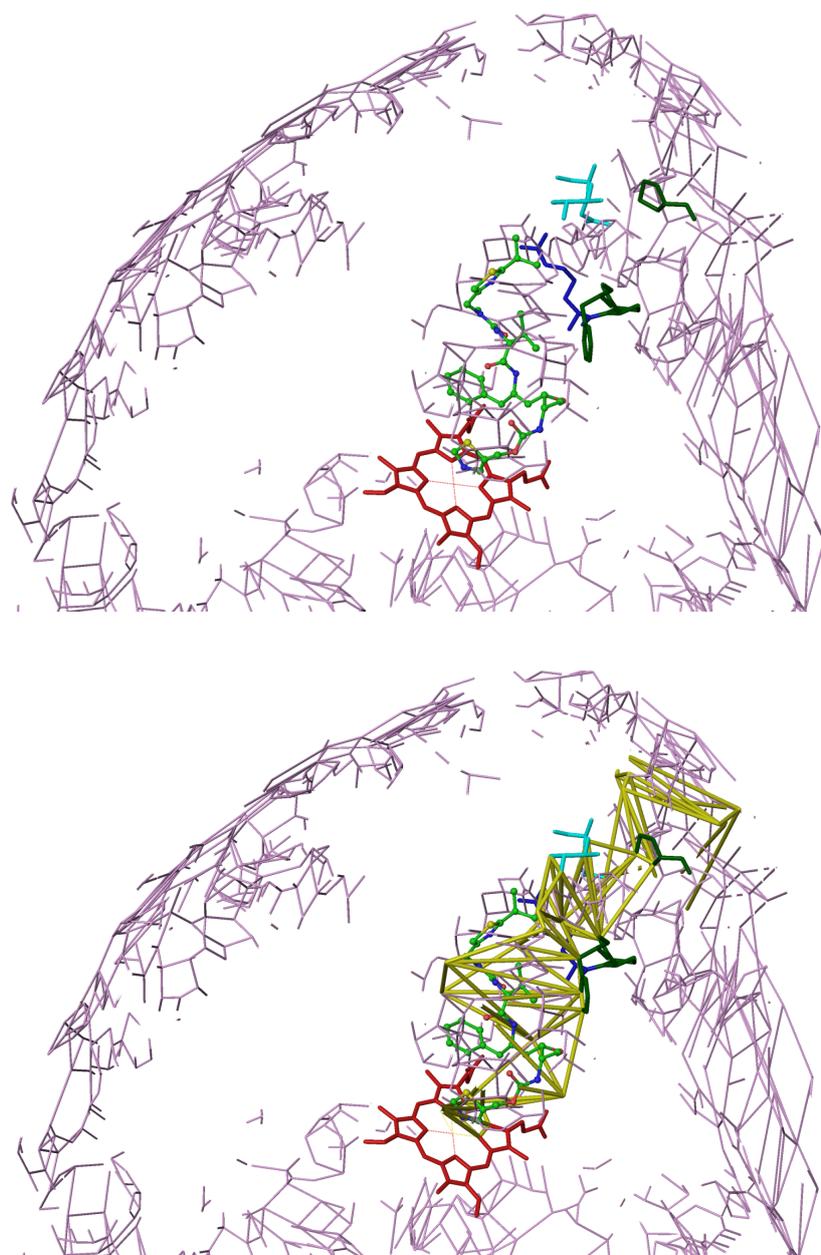
**Figure 3.22 :** Structure du ritonavir en interaction avec les acides aminés voisins du CYP3A4 (code PDB 3NXU). Les lignes pointillées noires indiquent les liaisons H. Les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$ . Cette figure a été générée par PoseView [Stierand *et al.* 2006] à partir des données de la base PDB.

- Une interaction défavorable provenant d'une collision entre le Ph-1 et F304, conduisant au déplacement de l'hélice I par rapport à la forme apo.
  - Ouverture plus ou moins prononcée de la boucle FF' vers la surface. (figure 3.24). R212 (boucle FF', dans le site actif) a un rôle modulateur dans la liaison avec les composés, et ainsi peut jouer un rôle de premier plan dans la transformation métabolique de xénobiotiques.
3. Présence d'une poche hydrophobe au niveau du second phényle (noté Ph-2), stabilisante par des interactions hydrophobes et van der Waals :
- ◆ Existence d'un espace adjacent aux acides aminés 369 à 371 (boucle K- $\beta_1$  et feuillet  $\beta_1$ ). Gêne stérique entre le site Ph-2 du ligand et les acides aminés 369 à 371, et déplacement de l'hème.
  - ◆ Existence d'un espace adjacent à R105 (boucle BC).
  - ◆ Interaction avec le site Ph-2 est importante pour une liaison ferme.

4. Le groupe terminal IPT contribue moins au processus d'association, et les contacts avec l'IPT semblent renforcer la liaison Fe–N :
  - Les liaisons H de S119 (boucle BC, dans le site actif) avec les atomes du squelette du ligand sont importantes, peu importe que la liaison H avec le ligand soit formée ou non [Sevrioukova & Poulos 2013a]. S119 affecte nettement l'affinité et la cinétique de liaison de tous les composés étudiés. C'est un acide aminé clé qui contribue à lier les inhibiteurs de type ritonavir, qui régule l'association, aide à la formation et augmente la stabilité des formes holo. S119 régule la spécificité du 3A4 aux substrats. La mutation S119A diminue la température de fusion pour toutes les formes holo.
  - Existence d'un pont, via une molécule d'eau, avec l'azote de l'IPT et le *parapluie polaire* formé par D61, D76, R106, R372, et E374.
5. Au niveau de l'activité :
  - ◆ Diminution de l'affinité.
  - ◆ Diminution de la force de la liaison à l'hème (Fe–N). Cette force est essentielle à une ferme association avec les analogues du ritonavir. Elle définit l'affinité de la molécule avec le 3A4, et c'est cette liaison forte et irréversible qui est la contribution clé à l'inhibition du 3A4, plutôt que la formation de métabolites réactifs.
  - ◆ Complexification du processus cinétique de liaison à l'hème.
  - ◆ Diminution de l'hydrophobicité : les forces hydrophobes définissent le mode de liaison.
  - ◆ Diminution du pouvoir inhibiteur. Celui-ci dépend des interactions hydrophobes non covalentes dus aux groupes latéraux.
  - ◆ Influence sur le mode de liaison au ligand (ces modes sont multiples).
  - ◆ Diminution de la stabilité thermique de liaison. La dénaturation thermique peut être un outil supplémentaire pour évaluer les modifications du 3A4 induites par le ligand, car la différence de températures de fusion se corrèle avec l'affinité au ligand : plus la liaison est stable, plus la variation du point de fusion est importante [Sevrioukova & Poulos 2013c].

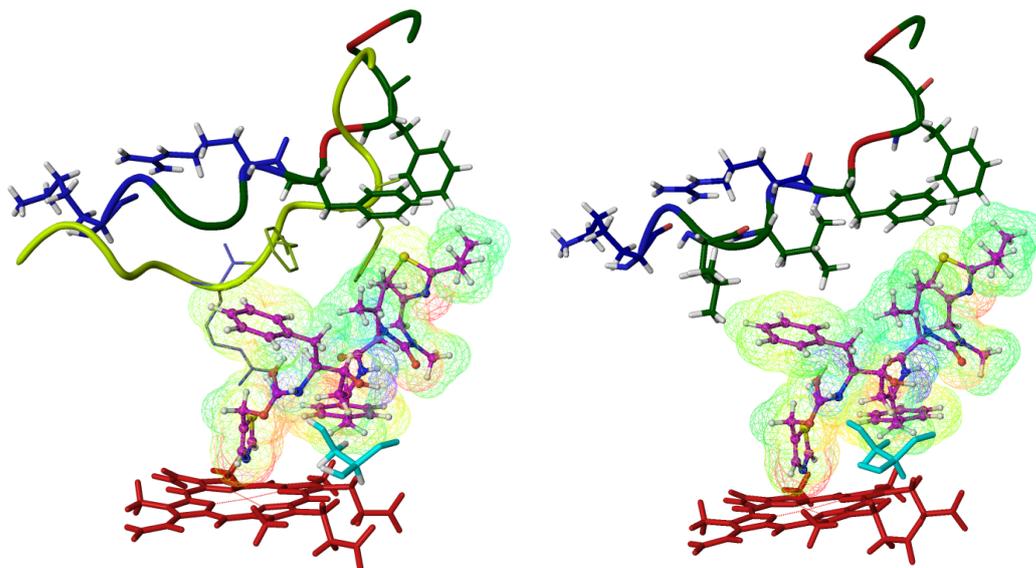
### 3.6.3 Influence des groupes latéraux et terminaux du ritonavir sur l'ouverture du CYP3A4

La molécule rentre bien dans la poche du site actif en raison de contacts hydrophobes additionnels avec ses groupements latéraux (figure 3.22), et fait écran au solvant [Sevrioukova & Poulos 2010]. Les résultats obtenus avec CCCPP le confirment : voir le graphe facial à  $CV_{lim}=6.25\text{\AA}$  du canal contenant le trajet 2a (figure 3.23). Le ritonavir est logé dans une impasse du réseau de canaux qui n'est pas contenue dans le trajet 2a. Deux hypothèses sont envisageables : soit la molécule rentre par un trajet 2a mitoyen à celui calculé par CCCPP, de coût supérieur mais non calculé par CCCPP (voir section 2.3.1), et contenant la partie du canal 2a logeant la queue de la molécule ritonavir, soit la molécule rentre par le trajet 2a de plus bas coût calculé par CCCPP (et donc le plus large), puis la queue de la molécule quitte le trajet 2a et se loge dans le cul de sac du canal 2a pour mieux s'adapter à la poche, stabilisée par interaction avec F213 (empilement  $\pi$ - $\pi$ ). La molécule de ritonavir logée dans le canal 2a est stabilisée par de fortes interactions



**Figure 3.23 :** Graphe facial du canal 2a à  $CV_{lim}=6.25\text{\AA}$ . Le ritonavir est représenté en boules-bâtons. Les acides aminés formant le goulot d'étranglement sont en tubes, coloriés selon leurs propriétés. En bas, le trajet 2a est ajouté en jaune.

hydrophobes et van der Waals avec le Ph-2 et le cluster Phe via les groupes latéraux, et se trouve complètement isolée du solvant (voir figure 7E dans [Sevrioukova & Poulos 2010]). De plus, le ritonavir est ancré dans le site actif directement par, (i) une liaison H entre l'azote de l'IPT et S119, (ii) une molécule d'eau formant un pont avec le *parapluie polaire* (voir fin de la section 3.6.1), et (iii) par un empilement  $\pi$ - $\pi$  entre le thiazole terminal et F213 (figures 3.24 et 3.25). La plupart des acides aminés participant à cette stabilisation



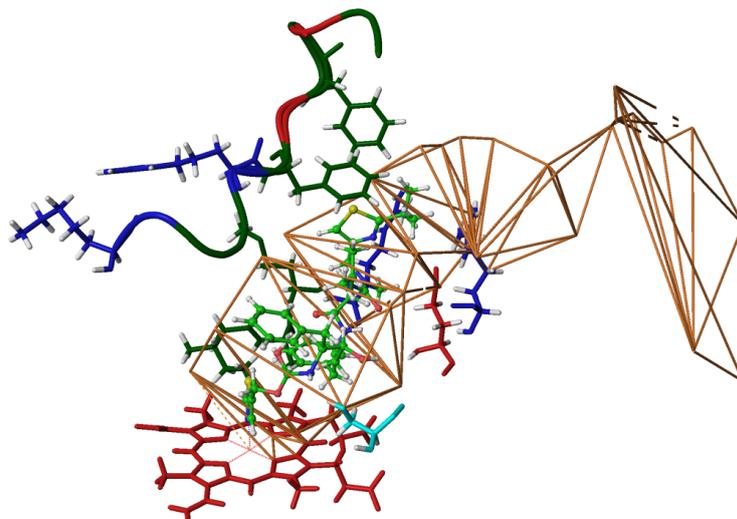
**Figure 3.24** : Maillage de la surface de van der Waals du ritonavir ( $777\text{\AA}^2$ ). Les sphères sont colorées suivant le potentiel électrostatique. Le ritonavir est en boules-bâtons, en magenta. La boucle FF' et les acides aminés du complexe 3NXU sont coloriés suivant leurs propriétés physico-chimiques. A gauche : superposition de la boucle FF' (en vert clair) des acides aminés R212, F213 et F215 (en tubes minces) de 1TQN (forme C). En pointillés noirs, la liaison H entre R212 et K208.

sont directement en contact (en bordure) du trajet 2a (S119, R106, F213, R372, E374) (figure 3.25).

Il a été suggéré que la structure du 3A4 lié au ritonavir soit très semblable à celle liée au kétoconazole et à celle liée à l'érythromycine, du fait d'une liaison similaire dans l'hélice I et dans la conformation de la boucle FF', avec R212 exposé vers l'extérieur [Sevrioukova & Poulos 2010]. En utilisant CCCPP, on voit en effet que les trois structures sont ouvertes, et qui conduit à une analyse plus fine en distinguant les deux modes d'ouverture O1 et O2. Il est effectivement commun aux trois structures que nous venons de citer que R212 soit exposé vers l'extérieur, mais le 3A4 adopte une forme O1 dans sa liaison avec le ritonavir, moins volumineux que les deux kétoconazoles et que l'érythromycine, ne nécessitant pas l'ouverture du bloc 2 en forme O2 comme dans le cas de liaison du 3A4 à ces deux dernières molécules.

La grande taille du ritonavir rend impossible l'accueil de deux molécules dans le site actif. La courbe cinétique atypique (non-hyperbolique, en forme de V) d'association au ritonavir ne peut s'expliquer par une liaison allostérique coopérative, mais plutôt par la capacité du ritonavir à se lier hors de la poche du site actif, sur un site périphérique, avant d'aller dans la cavité du site actif [Sevrioukova & Poulos 2012a] : voir section 3.7.1. Le ritonavir peut s'arrimer sur ce site périphérique, et une fois celui-ci saturé, il y a translocation vers le site actif et interaction avec l'hème.

Les acides aminés du goulot d'étranglement sont Q79, R106, P107, F108, T224 et P227. Ce ne sont pas ceux de la boucle FF' qui interviennent dans l'élargissement du canal 2a pour adopter la structure O1. Le mouvement de la boucle BC est essentiel dans

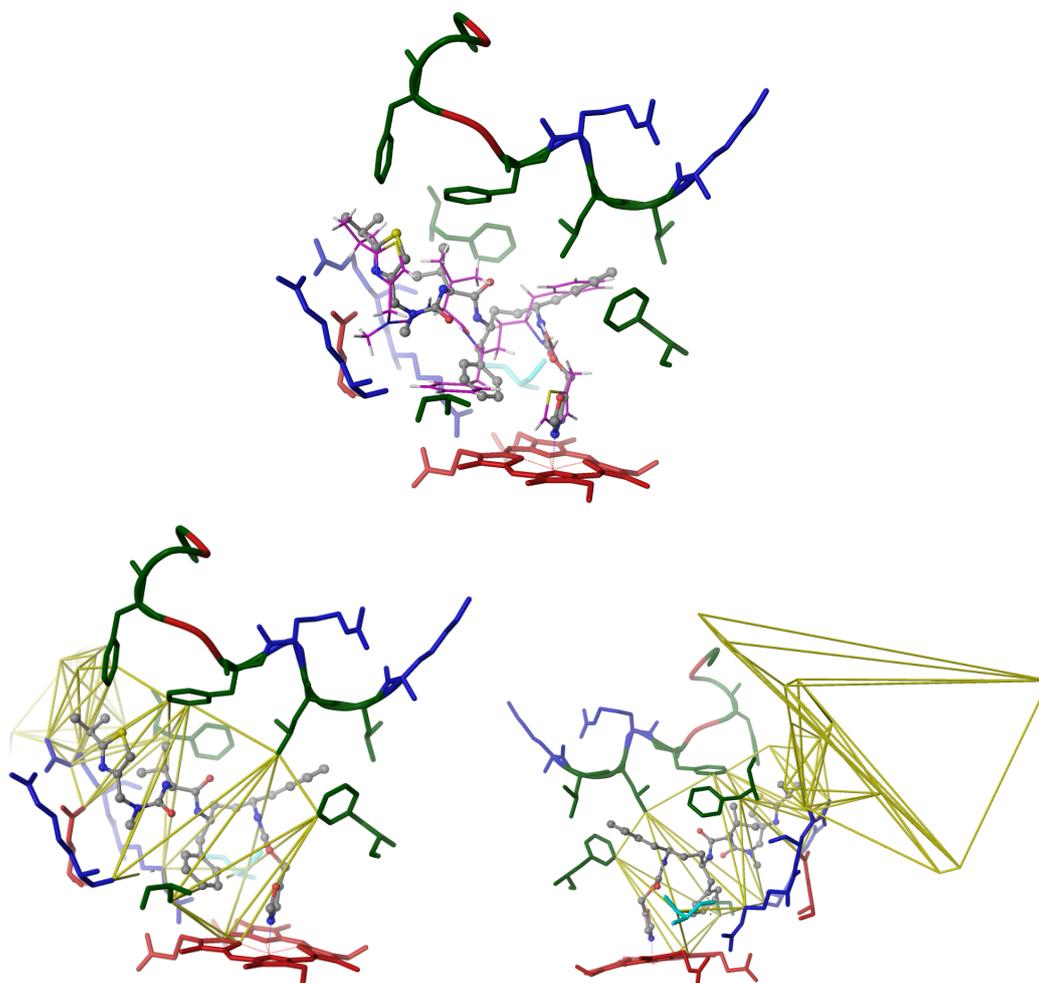


**Figure 3.25 :** Trajet 2a à  $CV_{lim}=6.25\text{\AA}$ . Les acides aminés bordant le trajet 2a sont en tubes colorés suivant leurs propriétés physico-chimiques.

la modulation de l'ouverture de ce canal principal 2a (passage de la forme C à la forme 01). La boucle FF' jouerait un rôle plutôt dans l'ouverture du bloc 2 (figure 3.23).

Alors que le ritonavir rentre bien dans le site actif du 3A4 et fait écran au solvant, sa liaison peut être améliorée. En particulier, il y a une collision stérique notable entre le Ph-2 et A370 (feuillet  $\beta_1$  : voir figure 8D dans [Sevrioukova & Poulos 2012a]), due au déplacement vers le bas du peptide 369–371 et du plan de l'hème, de 2.0 et 0.6Å, respectivement. Cette contrainte conformationnelle et le déplacement de l'hème affectent la liaison N–Fe. La substitution du Ph-2 du ritonavir par un groupe hydrophobe plus petit permettrait l'amélioration de l'accueil stérique du ritonavir dans le site actif du 3A4, avec l'élimination de contacts rapprochés non covalents avec A370, permettant la formation d'une liaison N–Fe plus ferme, permettant par conséquent l'augmentation de l'affinité au ligand et du pouvoir inhibiteur.

La mutation R212A n'a pas d'effet significatif sur l'affinité ni sur la modification spectrale induite par le ritonavir. Cela signifie que R212 n'est pas crucial pour l'association à l'inhibiteur. Par contre la mutation R212A diminue le taux d'association de 20% [Sevrioukova & Poulos 2010]. Cela suggère que R212 peut favoriser le processus de liaison, d'une part en stabilisant le conformère optimal du ritonavir dans la poche réactive (par des liaisons H, hydrophobes et de van der Waals), et d'autre part en stabilisant la conformation O1 du 3A4. En effet, déplacer R212 vers l'extérieur induit une orientation optimale du F213 par rapport au ligand, induisant un double effet stabilisant du complexe (pour le ritonavir et le 3A4). Ce déplacement de R212 d'une part permet d'éviter les collisions stériques avec F213 et le Ph-1 du ritonavir, ainsi qu'entre F215 avec l'isopropyle latéral, et d'autre part permet d'établir des empilements  $\pi$ - $\pi$  entre F213 et le thiazole terminal. La rotation de R212 a également un effet stabilisant sur la protéine, en stabilisant la boucle FF' par liaison hydrogène avec l'oxygène du carbonyle de K208 (figure 3.24).



**Figure 3.26 :** Complexe du CYP3A4 avec GS2 (code PDB 4I4G). GS2 est en boules-bâtons, en gris. Les acides aminés, hormis ceux de la boucle FF', bordent le trajet 2a; ils sont en tubes coloriés suivant leurs propriétés physico-chimiques. Pour plus de clarté, les hydrogènes du complexe 4I4G ne sont pas représentés. En haut, le ritonavir superposé à GS2, en tubes minces, en magenta. En bas, à gauche, visualisation du trajet 2a; à droite, idem, après rotation de  $180^\circ$ .

### 3.6.4 Influence des groupes de tête sur la stabilité du complexe

Le calcul CCCPP sur ces deux structures a révélé une ouverture globale du 3A4 comparable au cas du ritonavir. Un trajet 2a associé à un canal 2a incluant la molécule est calculé dans chacun des deux cas à  $CV_{lim}=6\text{\AA}$  et à  $CV_{lim}=6.25\text{\AA}$ , respectivement. Les deux molécules interagissent avec le 3A4 et se positionnent de façon comparable au ritonavir. Le 3A4 adopte la forme O1 (figure 3.26).

La substitution du thiazole lié à l'hème et la suppression du groupement hydroxy n'affectent pas l'ouverture globale du CYP3A4, qui adopte alors une ouverture de type O1 identique à celle rencontrée dans le cas du ritonavir. Néanmoins, il demeure des changements structuraux pour les deux entités en interaction. Ces changements sont moins déterminants sur l'ouverture du 3A4, s'adaptant aux structures chimiques des désoxyritonavirs, qui permettent d'améliorer l'affinité et le mode de liaison des complexes.

La force de la liaison Fe-N stabilise le ligand et augmente son affinité pour la protéine. GS2 est resté lié au site actif durant la cristallisation. Cependant, seul l'oxazole lié à l'hème est défini dans la structure cristallisée, alors que le reste de GS2 est désordonné car de moindre affinité par rapport au ritonavir. L'oxazole a des propriétés stéréo-électroniques moins favorables que le thiazole et la pyridine, donc GS2 se lie au fer de l'hème de manière plus relâchée. GS2 est alors un inhibiteur moins fort du 3A4, et offre une affinité moins grande que le ritonavir. Bien que GS2 ait un squelette plus flexible pour optimiser les interactions avec la protéine et augmenter son affinité, GS2 est très peu affiné au CYP3A4. Ceci nous suggère que la force de liaison Fe-N de l'inhibiteur est prioritaire sur l'affinité. Nous supposons que la forte liaison à l'hème permet de fixer la molécule, et favorise ensuite la mobilité du squelette de la molécule et les interactions de ses substituants avec les acides aminés de la protéine, ce qui la rendra plus ordonnée et plus affinée. Le cas du GS1 confirme cette suggestion, le groupement imidazole ne se liant pas du tout à l'hème (faibles propriétés oxydoreductrices, affaiblissant le potentiel rédox de l'hème), dissocié durant la cristallisation. GS1 n'est pas du tout retenu dans le site actif par les simples interactions hydrophobes, van der Waals et polaires, non spécifiques des groupements latéraux et terminaux du désoxyritonavir. La liaison covalente Fe-N est nécessaire pour retenir l'inhibiteur [Sevrioukova & Poulos 2013b]. Un comportement identique s'observe pour le déazaritonavir, avec une perte de l'azote du thiazole qui se lie au 3A4 avec une affinité bien moindre [Sevrioukova & Poulos 2012a].

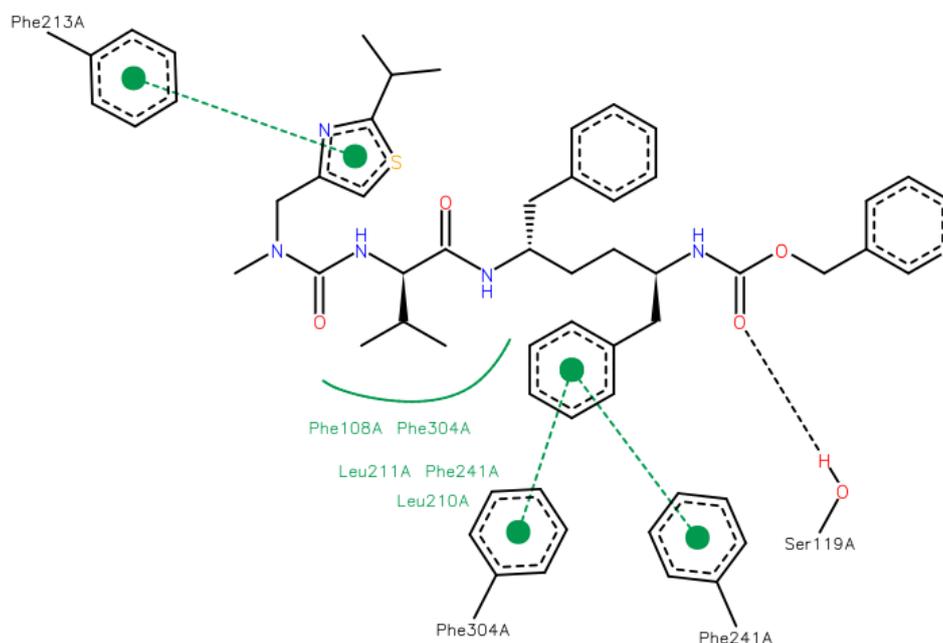
La flexibilité structurale du squelette du désoxyritonavir est plus favorable aux propriétés stéréo-électroniques de l'azote de la pyridine, qui améliorent les interactions avec la protéine et augmentent l'affinité, renforçant ainsi la liaison avec le fer de l'hème. GS3 est un ligand plus fort que le ritonavir en termes d'affinité de liaison et de pouvoir inhibiteur, et tout l'IPT de GS3 est observé dans la structure cristalline [Sevrioukova & Poulos 2013c].

Les acides aminés intervenant dans le mécanisme d'interaction des composés des désoxyritonavirs substitués par les groupes oxazole pyridine, en interactions hydrophobes et van der Waals, et avec des liaisons polaires non spécifiques, sont ceux qui bordent le trajet 2a dans chacune des structures (figures 3.26, 3.27 et 3.28). L'espace de passage de la molécule est alors défini comme étant la plus favorable stéréo-électroniquement.

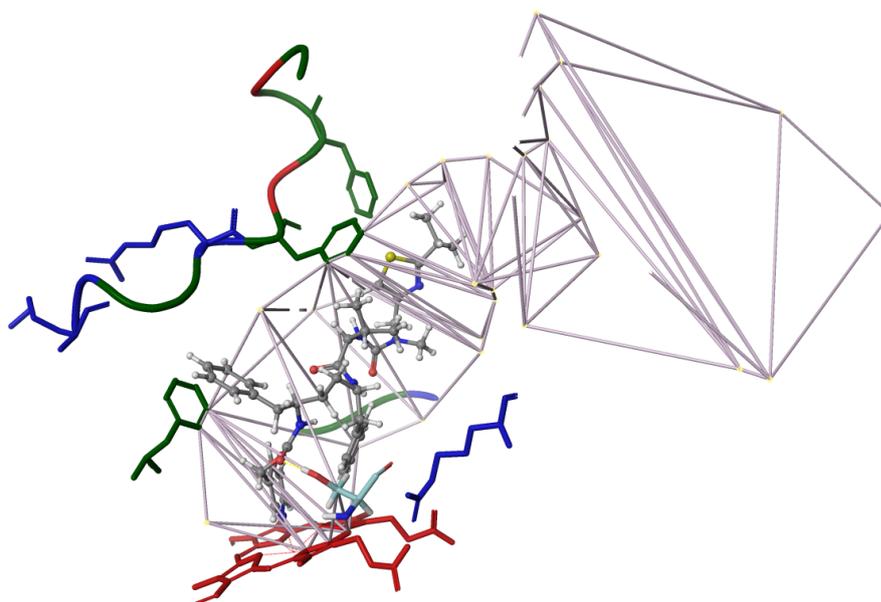
### 3.6.5 Rôle du thiazole de tête et de l'isopropylthiazole terminal

Il y a deux analogues du ritonavir, l'un avec élimination d'un groupement de tête thiazole entier (DTMCR) et l'autre avec un groupement terminal entier IPT (GS8), incluant chacun un fragment du squelette avec un groupement de tête thiazole et le groupe IPT, le reste de la molécule étant identique au ritonavir. Cela a permis d'étudier l'impact de ces groupements manquants sur le rôle des interactions non covalentes stabilisantes.

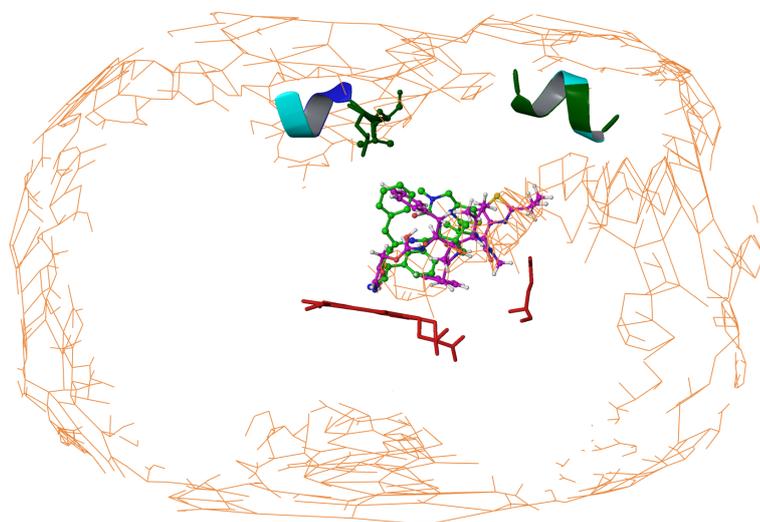
La suppression des deux groupements extrêmes rapetisse les deux molécules, dès lors le groupement final IPT ne peut plus interagir fortement avec la boucle FF'. Cela rend instables les entités en interaction, aussi bien l'inhibiteur que le 3A4. De ce fait, la cristallisation des structures liées au DTMCR et au GS8 (codes respectifs, 3TJS et 4K9X) n'est pas très bien résolue au niveau du toit du site actif, correspondant à l'interaction de la queue de l'analogue ritonavir avec la partie très labile FF' du 3A4. Dès lors, il nous a été difficile dans le cadre de ce travail de faire un calcul des canaux par CCCPP



**Figure 3.27 :** Interactions non covalentes du composé GS3 avec le CYP3A4 (complexe 4I4H). Les lignes pointillées noires indiquent les liaisons H. Les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$ . La figure a été générée par PoseView [Stierand *et al.* 2006].



**Figure 3.28 :** Trajet 2a dans le complexe du CYP3A4 avec GS3 (code PDB 4I4H). Les lignes pointillées noires indiquent la liaison H entre le groupe OH de S119 et l'oxygène de carbonyle de GS3. La boucle FF' adopte la forme O1.



**Figure 3.29** : Complexe du CYP3A4 avec le DTMCr (code PDB 3TJS). Le DTMCr est en boules-bâtons, en vert, superposé au canal 2a de la structure 3UA1 liée à la bromoergocryptine (non représentée), calculé à  $CV_{lim}=6.75\text{\AA}$ . Le canal 2a est en lignes oranges. Par comparaison, le ritonavir est superposé en boules-bâtons, en magenta.

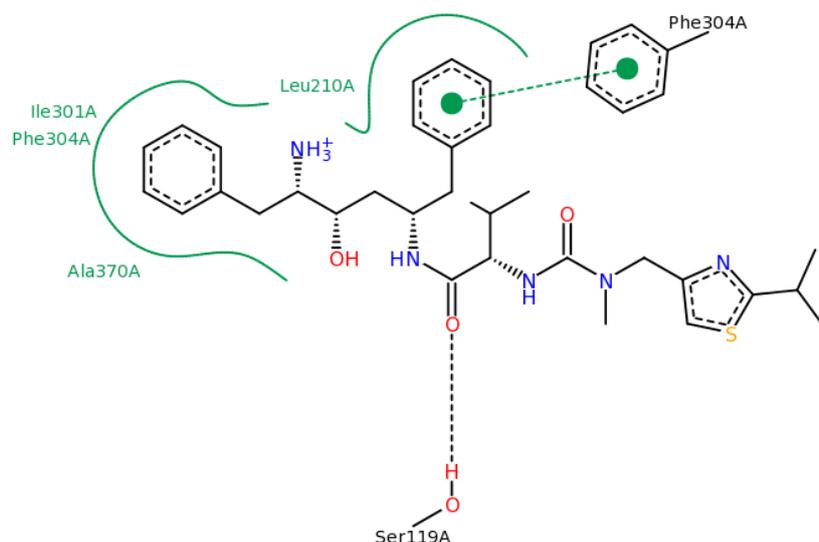
et de déterminer la forme des protéines, dans ce cas de complexation instable, car la reconstruction des parties manquantes de la boucle FF' n'est pas fiable.

Néanmoins, nous avons pu prédire la forme de la protéine et son ouverture, en tenant compte des propriétés stéréochimiques de chacune des molécules et par comparaison à une ouverture de la protéine correspondante. En effet, les données nous ont mené à prédire une forme probable O1 pour le 3A4 en interaction avec le DTMCr et GS8 (figures 3.29 et 3.30).

### Effet des interactions non covalentes de la tête thiazolique

Le DTMCr n'est pas inclus dans le canal 2a du CYP3A4 en forme C. En revanche il est inclus dans le canal 2a de la forme O1 liée au ritonavir. La superposition du ritonavir appuie cette prédiction dans les deux cas, le ritonavir se positionnant de façon globalement comparable au DTMCr (figure 3.29 et 3.31). Le DTMCr est plus petit, donc occupe un volume plus faible dans le site actif et établit moins de contacts hydrophobes et des contacts polaires non spécifiques. Le DMCR adopte un mode de liaison comparable à l'avant et au milieu du ritonavir, alors que le reste de la partie terminale interagit différemment avec la protéine.

La superposition de la boucle FF' du 3A4 lié au ritonavir en forme O1 montre que le DTMCr est dans la poche réactive en interaction stabilisante avec les acides aminés de la boucle FF', déterminante pour l'ouverture du CYP3A4 (figure 3.31 en bas). En revanche, la stabilisation est moindre par rapport à l'interaction du ritonavir. En effet, le raccourcissement du DTMCr dû à l'élimination du groupement de tête en entier ne permet pas l'étalement de la queue de la molécule le long du canal 2a. A ces

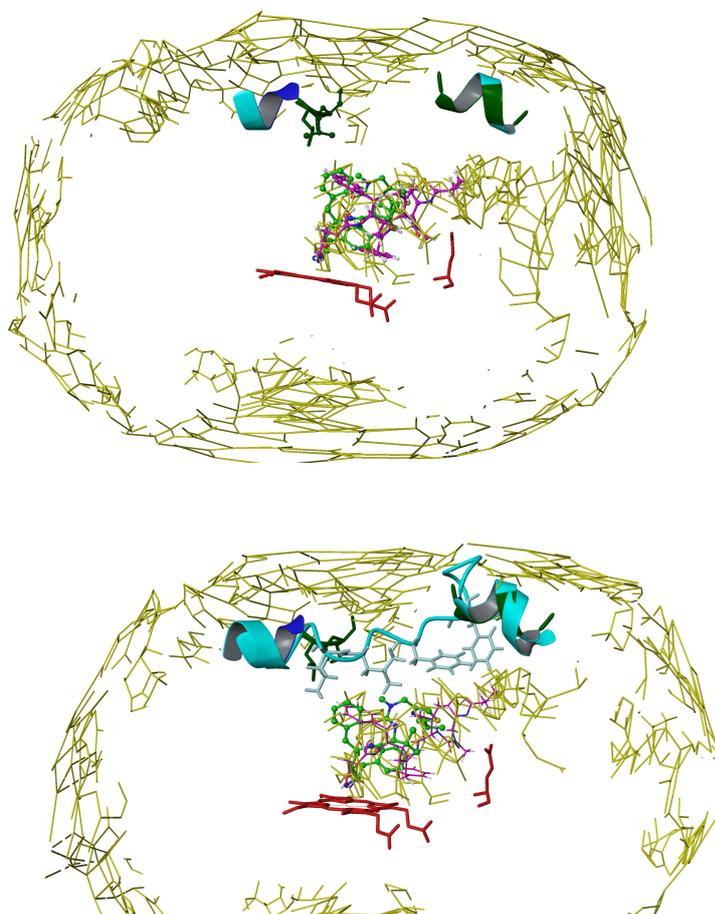


**Figure 3.30** : Interactions non covalentes du DTMCr avec le CYP3A4 (complexe 3TJS). Les lignes pointillées noires indiquent les liaisons H. Les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$ . La figure a été générée par PoseView [Stierand *et al.* 2006].

deux phénomènes s'ajoute l'absence du cycle thiazole, l'ensemble des deux phénomènes étant défavorable aux interactions non covalentes. L'interaction fortement stabilisante des empilements  $\pi$ - $\pi$  entre le cycle thiazolique terminal et F213 est absente. Il se soustrait également l'interaction hydrophobe entre l'IPT terminal et F215.

Pour optimiser l'affinité et donc les interactions non-liantes dans les deux poches, le DTMCr effectue une rotation de  $180^\circ$  par rapport au ritonavir, pour placer les deux groupes phényle dans les poches hydrophobes [Sevrioukova & Poulos 2013b]. La rotation est imposée pour loger le Ph-2 dans la poche 1, car la distance du groupement amine au Ph-1 est trop courte pour lui permettre l'accès à la poche 1. Des interactions hydrophobes extensives au niveau du site Ph-1 déterminent l'orientation de la liaison, optimisent la position du groupe amine et renforcent ainsi la liaison N-Fe.

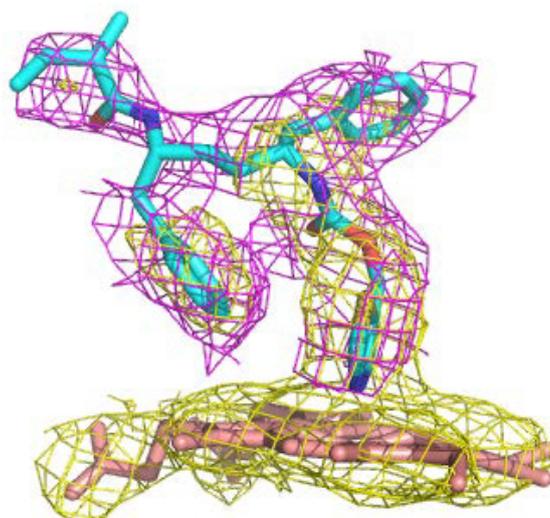
Les principales raisons en faveur d'un site actif plus ouvert et solvato seraient, d'une part l'éloignement entre l'IPT du DTMCr et E374 (feuille  $\beta_1$ ) qui empêcherait la formation du réseau de liaisons H (dû à l'eau) et qui le connecterait à l'amas d'acides aminés chargés protégeant le ritonavir du solvant [Sevrioukova & Poulos 2010], et d'autre part, l'instabilité de la boucle FF' liée à sa faible interaction avec le DTMCr par rapport au ritonavir. Ceci est dû principalement à l'absence d'interactions  $\pi$ - $\pi$  entre le thiazole terminal et F213 dans sa position optimale (pas de contacts hydrophobes entre le DTMCr et F213), qui favorise l'orientation de R212 vers l'extérieur. Ce mécanisme concerté permettrait la stabilité mutuelle de la boucle FF' et de la molécule. La boucle FF' en conformation O1 serait alors stabilisée par interaction hydrogène avec l'oxygène du carbonyle de K208. A cela s'ajouterait la stabilisation mutuelle des deux entités grâce aux contacts hydrophobes entre L211 et le Ph-2 du DTMCr et entre F215 et l'isopropyle du groupe IPT. Il en résulte que les acides aminés 212 à 218 deviennent désordonnés, et que la boucle C-terminale se déplace sur le côté de  $2\text{\AA}$ , laissant la cavité du site actif ouverte au solvant (figure 8 dans [Sevrioukova & Poulos 2012b]).



**Figure 3.31** : Réseau des canaux du CYP3A4 lié au DTMCr à  $CV_{lim}=6.25\text{\AA}$ . Le DTMCr est en boules-bâtons, en vert, superposé au canal 2a du 3A4 (tubes minces jaunes) lié au ritonavir (en boules-bâtons, en magenta), calculé à  $CV_{lim}=6.25\text{\AA}$ . En bas, superposition de la boucle FF' (ruban cyan) avec les acides aminés du 3A4 liés au ritonavir (tubes cyan).

### Rôle du groupement isopropylthiazole dans la stabilisation de la boucle FF'

GS8 contient les deux phényles latéraux mais, comparé au ritonavir, l'IPT manque ainsi qu'un fragment du squelette terminal (figure 3.21). Le positionnement de GS8 dans le complexe formé avec le 3A4 est indiqué figure 3.32. La CYP3A4 complexé à GS8 est instable au niveau des acides aminés 212 et 213 sur la boucle FF', et davantage au niveau de l'hélice F. La structure 4K9X ne contient pas les acides aminés dans ces régions cités (voir tableau 2.2). La reconstruction de l'hélice F n'a pas pu être correctement effectuée, et n'est donc pas fiable. Néanmoins la superposition de la structure 4K9X sur 4I4H et 3NXU nous permet d'observer que la forme O1 est adoptée, notamment au niveau du repliement de la boucle FF' (figure 3.33). On observe un changement de conformation de R212 par rapport à 4I4H. Dans la structure cristalline, seul le thiazole de GS8 et les phényles latéraux sont clairement résolus : GS8 est positionné comme GS3, et forme à peu près les mêmes contacts avec les deux phényles [Sevrioukova & Poulos 2013b]). F304 et l'hélice I subissent un déplacement de même amplitude que dans 4I4H en raison de la gêne



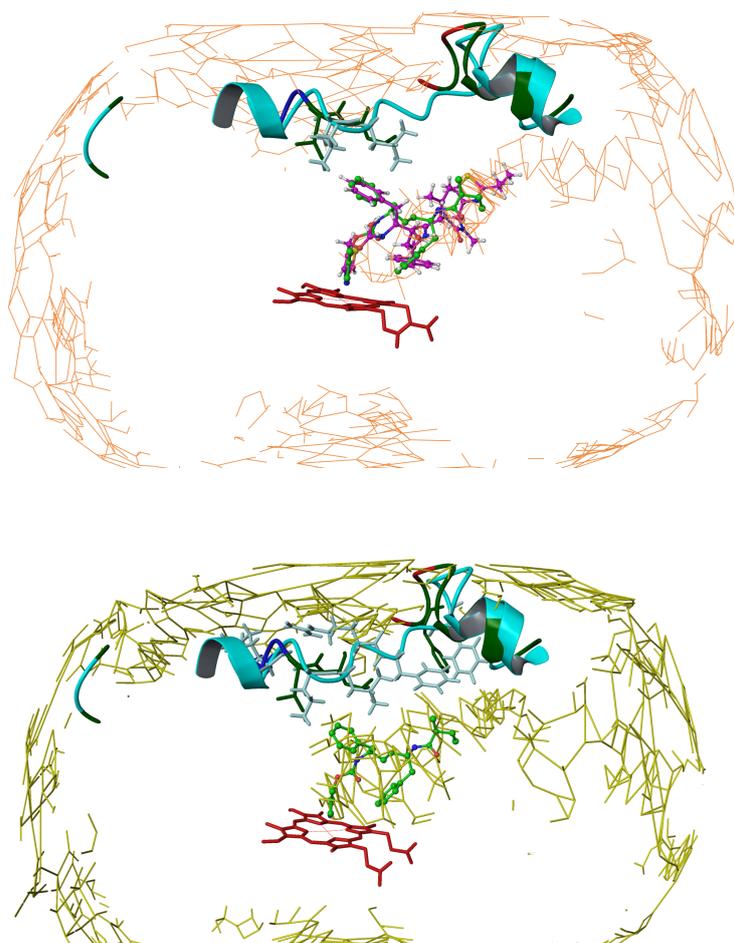
**Figure 3.32 :** Positionnement de l’analogue GS8 du ritonavir dans le complexe du CYP3A4 avec GS8 (code PDB 4K9X). GS8 est partiellement désordonné dans la structure cristalline (image extraite de [Sevrioukova & Poulos 2013b]).

stérique du Ph-1, avec comme différences une distance plus importante entre l’oxygène du carbonyle de GS8 et le groupe hydroxy de S119 (3.5Å dans GS8 contre 2.5Å, dans GS3), ce qui vient probablement du manque de stabilisation de GS8 avec F213 dû à l’absence du thiazole terminal; de plus, F213 est lui-même non résolu car très labile. La boucle FF’ (acides aminés 211 à 217) adopte la même conformation O1 que dans 4I4H et 3NXU, et cette conformation ouverte laisse le site actif accessible au solvant (figure 3.33). A travers les contacts avec la boucle FF’, l’IPT contribue à son adaptabilité et donc à son ouverture. En effet, l’orientation de l’IPT permet la modulation de la conformation de la partie terminale et le renforcement de la liaison Fe–N, à travers les interactions polaires, hydrophobes,  $\pi$ - $\pi$ , et/ou S- $\pi$  [Sevrioukova & Poulos 2013b].

L’affinité de GS8 est affaiblie par rapport au ritonavir. L’absence de l’IPT ne permet pas les interactions dus aux empilements  $\pi$ - $\pi$  avec F213, déstabilisant ainsi le complexe, qui ne permet pas ainsi le maintien de la molécule GS8 pour former la liaison H avec S119 (figure 3.33). La flexibilité du squelette du désoxyritonavir permet au cycle de tête flexible lié au fer hémunique une orientation parallèle au Ph-2 de GS8, et sa stabilisation dans la poche du Ph-2 entre les interactions des empilements  $\pi$ - $\pi$  avec le cycle tête, et les interactions cation- $\pi$  avec R105, contrairement au ritonavir dont le cycle du Ph-2 est perpendiculaire à celui de GS8 (figure 3.33).

### 3.6.6 Rôle des groupes hydrophobes latéraux

Le composé GS3 a un pouvoir inhibiteur plus fort que le ritonavir [Sevrioukova & Poulos 2013c]. Les composés GS4 à GS8 sont des analogues du ritonavir utilisés pour cette analyse car ils sont semblables à GS3 (figures 3.21 et 3.34; voir aussi figure 5 dans [Sevrioukova & Poulos 2013b]). Les résultats fournissent une preuve supplémentaire de la grande flexibilité et de l’adaptabilité du site actif du 3A4, qui prend les formes C, O1 et O2, et de sa capacité à accueillir de multiples ligands. Ces résultats montrent que l’activité



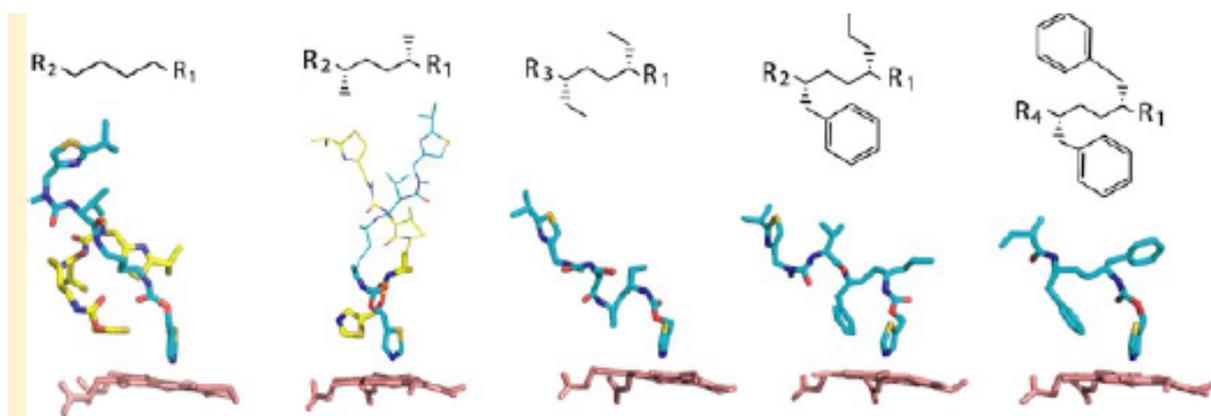
**Figure 3.33** : Superposition du complexe CYP3A4-ritonavir avec le complexe CYP3A4-GS8. Le complexe 3A4-ritonavir est en cyan. GS8 est en boules-bâtons, en vert. Le ritonavir est superposé en boules-bâtons, en magenta. L'interaction de GS8 avec le 3A4 en forme O1 est favorable, contrairement à celle du 3A4 en forme C, défavorable à l'accueil d'une molécule de GS8. En haut : le canal 2a de la structure C (code PDB 3UA1) à  $CV_{lim}=6.75\text{\AA}$  en lignes oranges; le ritonavir est en boules-bâtons, en magenta. En bas : le canal 2a de la structure O1 (code PDB 3NXU) à  $CV_{lim}=6.25\text{\AA}$  en tubes minces jaunes.

de ces analogues dépendent beaucoup de la nature chimique et de la taille des groupes latéraux et qu'elle est arbitrée par des interactions polaires avec S119, l'IPT agissant moins sur l'activité mais plus sur la stabilité du désoxyritonavir dans la cavité.

### Importance du premier phényle dans l'affinité et l'inhibition

L'adoption de la forme O1 par le 3A4 est due à la présence de chaînes latérales dans les analogues ritonavir. Elle a le double avantage, d'une part d'éviter les collisions stériques (les plus importantes étant avec les acides aminés de la boucle FF' dans le site Ph-1), et d'autre part de stabiliser le complexe en optimisant les interactions hydrophobes et de van der Waals.

GS7 possède un groupe propyle à la place du Ph-1 (figure 3.34) mais conserve une forte affinité pour le 3A4. GS7 adopte une conformation semblable à celle de GS3, la chaîne



**Figure 3.34** : Structure des déoxyritonavirs GS4 à G8 et orientation par rapport à l'hème. Codes PDB respectifs : 4K9T, 4K9U, 4K9V, 4K9W, 4K9X. Cette figure est extraite de [Sevrioukova & Poulos 2013b].

propyle occupant le site Ph-1 et déplaçant F304 et l'hélice I d'une même amplitude que le ritonavir par rapport au 3A4 libre.

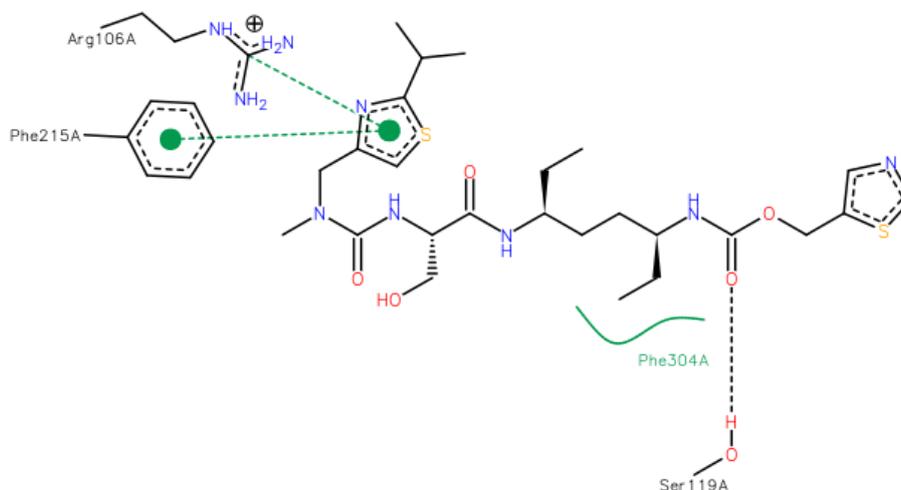
Le complexe 3A4-GS7 (code PDB 4K9W) adopte toujours la forme O1. La substitution du Ph-1 par un groupe hydrophobe plus petit et plus flexible a l'avantage d'empêcher les collisions stériques dans le site Ph-1, mais diminue l'hydrophobicité et donc affaiblit l'interaction de GS7 avec le 3A4 au niveau de ce site Ph-1. Les interactions avec le groupement propyle sont moins fortes. Ceci a pour conséquences, (i) pas de liaison H avec S119 (la distance entre l'oxygène de GS7 et l'hydroxy de S119 est de 3.4Å), et (ii) la perte de cette liaison H induit la rotation du cycle thiazolique de tête (voir figures 5H et 2S dans [Sevrioukova & Poulos 2013b]). Cela nous suggère l'importance du premier phényle sur le pouvoir inhibiteur. Il y a malgré tout une stabilisation de la valine latérale s'opérant grâce à sa réorientation, ce qui induit un réarrangement de la boucle FF' (en forme O1), et des interactions hydrophobes optimales avec L211 et F213.

Comme pour le ritonavir, l'IPT terminal participe à la stabilité de l'ouverture O1. L'azote de IPT de GS7 établit un pont (arbitré par l'eau), et est dirigé vers un amas d'acides aminés chargés (le *parapluie polaire*), en particulier, R106 (boucle BC) et E374 (feuillelet  $\beta_1$ ), et positionné favorablement à des interactions polaires. Il demeure néanmoins des différences de conformations de certains acides aminés, comme le F213. Le soufre de l'IPT est plus proche du cycle de F213, et contrairement au groupe terminal du ritonavir, peut former des interactions S- $\pi$  (voir figures 5F et 2S dans [Sevrioukova & Poulos 2013b]).

GS7 et GS8 ont la plus basse IC50. Les composés GS4 à G8 augmentent la stabilité thermique du 3A4 par rapport à sa forme apo, la plus grande stabilisation s'observant pour GS7 et GS8 [Sevrioukova & Poulos 2013b]. Ceci implique que les interactions hydrophobes dues aux phényles latéraux augmentent le pouvoir inhibiteur et stabilisent la forme liée du 3A4.

### L'élargissement du ligand provoque l'ouverture en forme O1

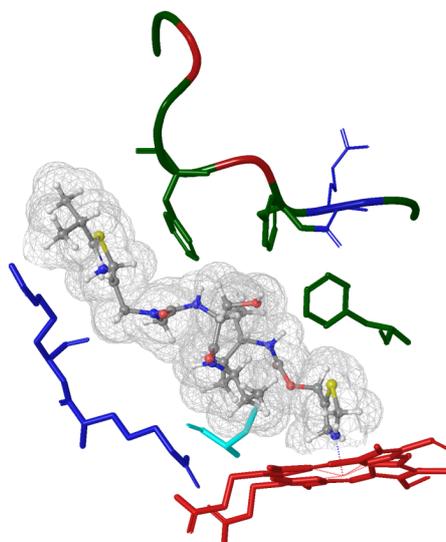
Le 3A4 lié à GS6 (4K9V, figures 3.21 et 3.34) adopte une conformation C distincte. Un canal 2a est calculé avec un site actif élargi par rapport à la forme apo. Trois trajets



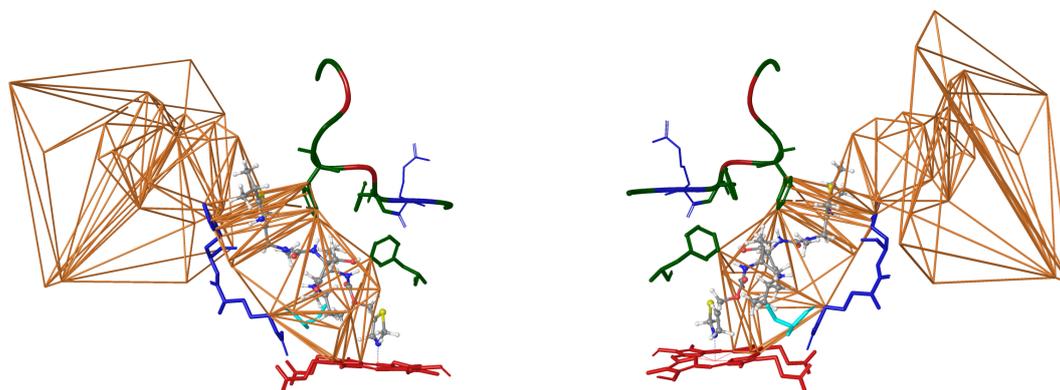
**Figure 3.35** : Interactions du composé GS6 avec les acides aminés voisins du CYP3A4 (code PDB 4K9V). Les lignes pointillées noires indiquent les liaisons H. Les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$  et cation- $\pi$ . La figure a été générée par PoseView [Stierand *et al.* 2006].

2a, à trois différentes valeurs de CV (6.5, 6.25 et 6Å) sont calculées (figure 3.37). Un mouvement dans la région de la boucle BC est observé sans qu'il n'y ait un important réarrangement dans la région FG, l'hélice I et la boucle C terminale (voir section 3.7.1). Cette adaptation de la molécule pour un moindre réarrangement de la protéine, s'effectue grâce à la forme étendue et la grande flexibilité conformationnelle de GS6 qui contient 16 liaisons pivotables. La courbure plus prononcée du squelette de GS6 lui permet de remplir le site Ph-2 avec son groupe éthyle et avec les atomes de la chaîne principale (voir figure 5I-5J-5K dans [Sevrioukova & Poulos 2013b]). Le site Ph-1 reste inoccupé, donc il n'y a pas de déplacement de l'hélice I et de la boucle FF'. Il n'y a pas non plus de déplacement du peptide 369–371 du site Ph-2, car les deux groupements éthyle ne sont pas encombrants et n'engendrent pas de collision stérique pour la position initiale en forme C de la protéine. En raison de la polarité plus élevée de la chaîne latérale de la sérine, une moitié de GS6 se déplace pour former des contacts polaires entre les oxygènes des carbonyles de I369, A370, et R372 (feuillet  $\beta_1$ ), et de G481 (boucle C-terminale). Et donc la boucle FF' adopte une autre conformation par rapport à GS3, afin d'optimiser les interactions avec la portion terminale de GS6. L'IPT s'intercale alors entre le cycle de F215 parallèle au groupe guanidine de R106 (figures 3.35, 3.36, 3.37; voir aussi figure 5J dans [Sevrioukova & Poulos 2013b]). Ceci provoque des interactions, respectivement  $\pi$ - $\pi$  and cation- $\pi$ . Ces nouvelles interactions, comme la liaison H à S119 et la protection complète du solvant, compensent la perte de stabilisation du complexe par les deux phényles. En revanche, le 3A4 ne perd pas d'énergie en changeant de conformation : il reste dans sa position initiale, avec ouverture de son canal 2a initial (figure 3.37).

Dans GS6, la substitution des deux phényles par des groupes hydrophobes plus petits diminue les collisions stériques, tel que la protéine ne se déforme pas et adopte la forme C, mais aussi diminue d'autant plus l'hydrophobicité, défavorisant ainsi l'interaction au niveau des deux sites phényle. La figure 3.36 montre l'étendue de l'intervention de GS6, conférant à la boucle FF' la forme C. Le sens de R212 est inversé par rapport à 1TQN.



**Figure 3.36** : Interactions du composé GS6 avec son environnement dans le complexe 4K9V formé avec le CYP3A4. GS6 est localisé dans le canal 2a (non délimité). Figure réalisée avec Maestro (Schrödinger, LLC, New York), rayon de van der Waals 2.0Å, surface 585Å<sup>2</sup>.

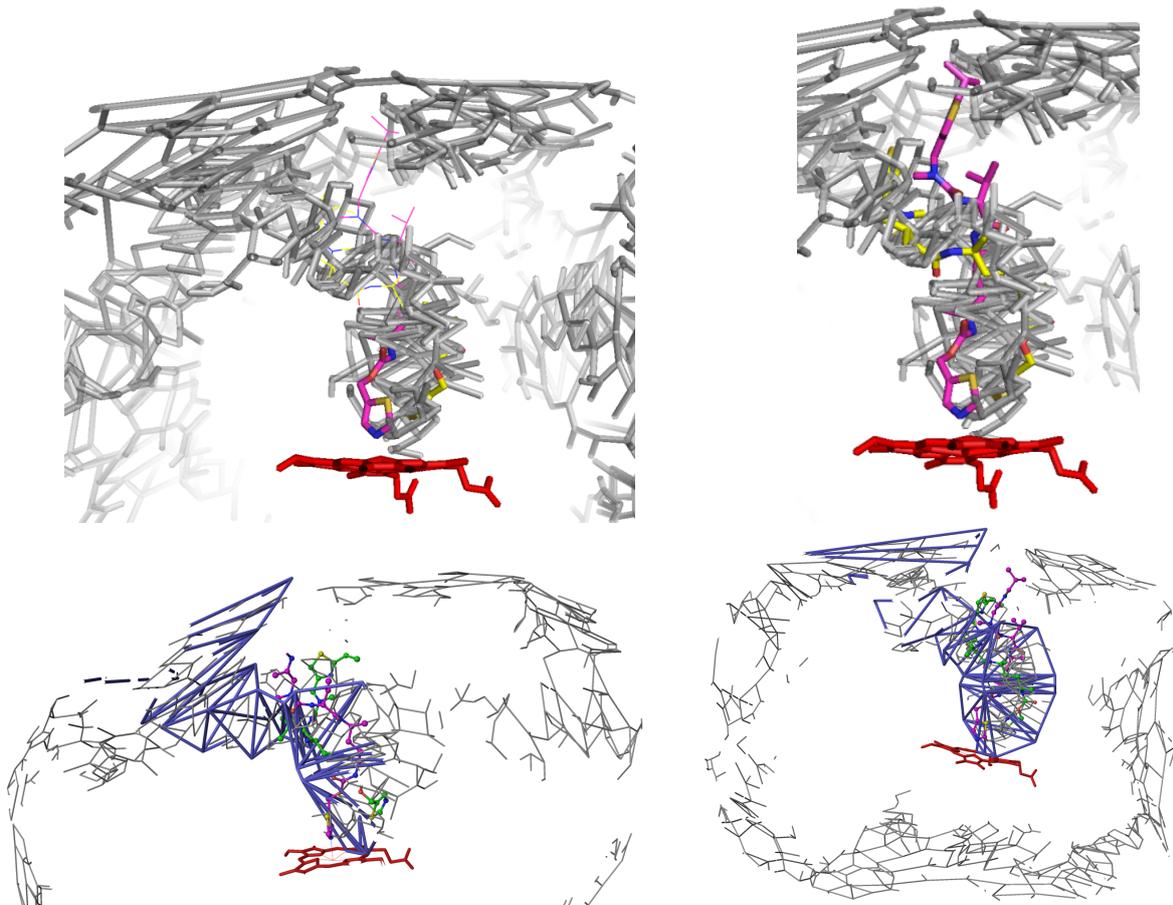


**Figure 3.37** : Visualisation du canal 2a à  $CV_{lim}=6.5\text{\AA}$  dans le complexe du composé GS6 avec le CYP3A4 (code PDB 4K9V). Les acides aminés sont en vert s'ils sont apolaires, en bleu s'ils sont chargés positivement et en rouge s'ils sont chargés négativement. A droite, idem, après rotation de 180°.

### Ouverture en forme O2 pour l'accueil de deux molécules sur deux sites de liaison

GS5 possède deux méthyles à la place des deux phényles latéraux (figure 3.21), et a deux sites d'accueil dans le 3A4, où les deux molécules (GS5-1 and GS5-2) sont liées en conformation allongée (figure 6 dans [Sevrioukova & Poulos 2013b]). Le thiazole de GS5-2 est au niveau du site Ph-2, intercalé entre le thiazole de GS5-1 et parallèle au groupe guanidine de R105, alors que la moitié de GS5-2 occupe le site Ph-1, ce qui se voit au niveau de F304 déformé et de l'hélice I (figure 6 dans [Sevrioukova & Poulos 2013b]).

Les portions de tête résolues aux rayons X des deux molécules de GS5 sont contenues dans le bas du canal d'ouverture, représenté par son graphe facial à  $CV_{lim}=6.75\text{\AA}$  (figure

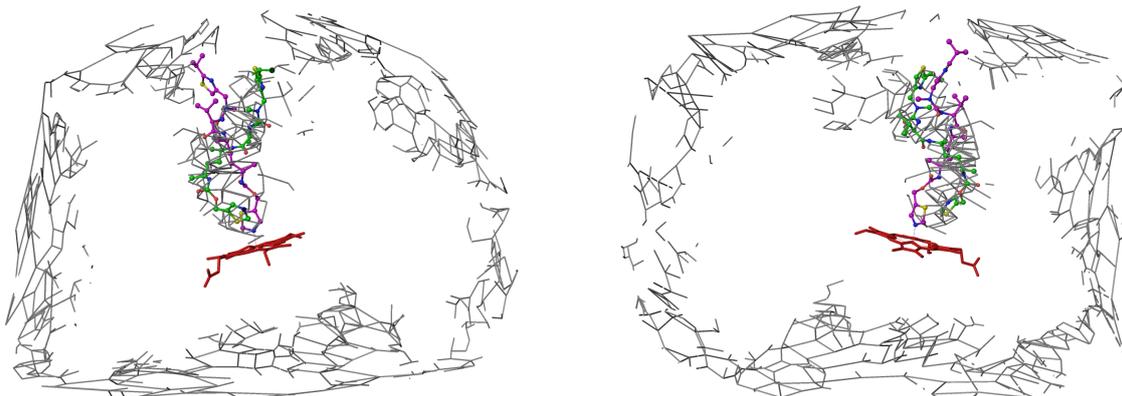


**Figure 3.38** : Graphe facial à  $CV_{lim}=6.75\text{\AA}$  du complexe entre deux molécules du composé GS5 et le CYP3A4 (code PDB 4K9U). En haut, à gauche : GS5-2 est allongé (en bâtons jaunes), légèrement replié autour de GS5-1 (lié et allongé, en bâtons roses); à droite, idem (zoom). Les parties des 2 molécules réelles (résolues) sont contenues dans le canal. La partie modélisée de GS5-1 n'est pas contenue dans le canal mais celle de GS5-2 est contenue dedans (en lignes). En bas, à gauche : superposition du trajet S en bleu; à droite, idem après rotation de  $90^\circ$ ; GS5-1 est en rose, GS5-2 est en vert.

3.38), ainsi que la portion terminale modélisée de GS5-2, sauf l'isopropyle du groupe IPT. GS5-2 est contenu dans le canal S (figure 3.39), mais l'isopropyle de l'IPT de GS5-2 n'y est pas. En effet, l'isopropyle occupe l'espace protéique du côté de la boucle FF', et entre en collision avec celle-ci (figure 3.40). En revanche, la partie modélisée de GS5-2 n'est pas dans le trajet S (figure 3.38 en bas). Ceci nous suggère une mauvaise modélisation de GS5-2, qui emprunterait plutôt le trajet S.

La partie terminale modélisée de GS5-1 n'est pas située dans le canal principal S, et s'avère être sur une position de la protéine (et non dans le vide), au niveau de la boucle FF' (figure 3.40). Il s'agit peut-être d'une erreur de reconstruction de GS5-1 ou de la partie désordonnée manquante de la boucle FF', qui se trouve à cet endroit après reconstruction. La partie terminale modélisée de GS5-1 n'est pas nécessairement contenue dans le canal S (voir figure 3.39).

A  $CV=6.25\text{\AA}$ , il y a apparition du canal 2f. Il contient les deux molécules, sauf le propyle de l'IPT de GS5-1 (idem pour le calcul à  $CV_{lim}$ ). Le trajet 2f ne contient pas les



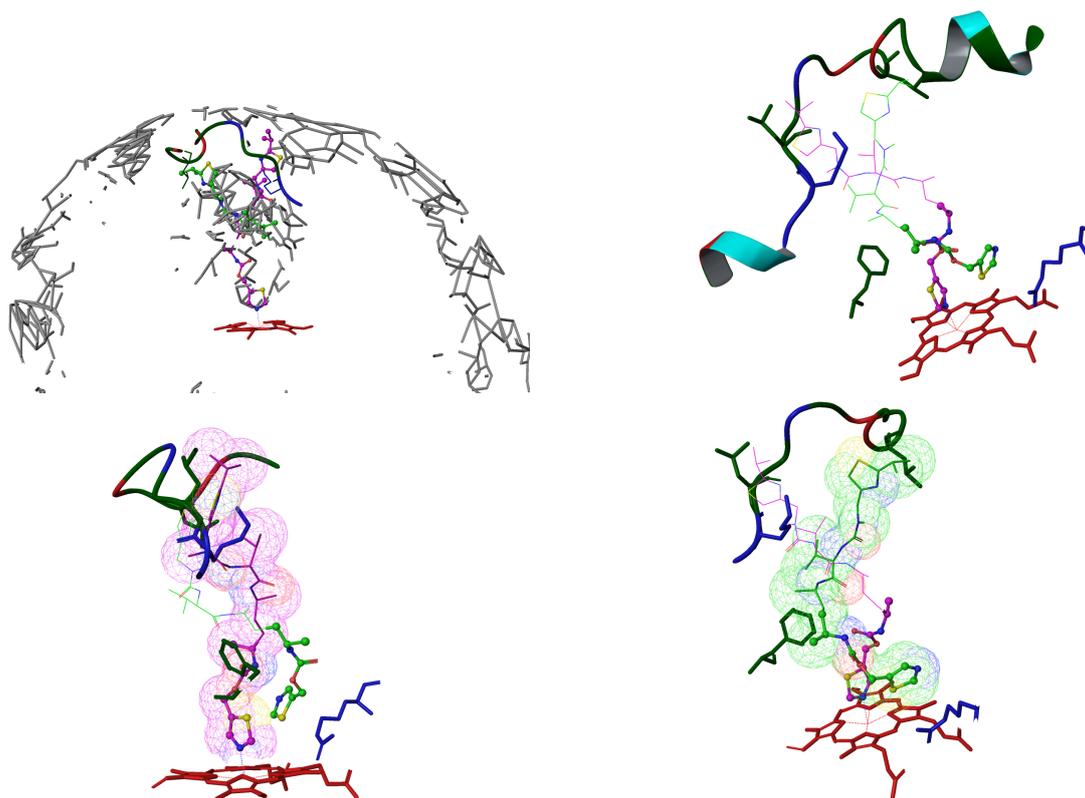
**Figure 3.39** : Visualisation du canal 2f à  $CV=6.25\text{\AA}$  dans le complexe GS5-CYP3A4 (code PDB 4K9U). A droite et à gauche, les deux vues diffèrent de  $180^\circ$ .

parties modélisées des 2 molécules. Les parties définies aux rayons X sont contenues dans le trajet 2f au niveau du site actif, qui est commun au canal S. La différence est dans le sens d'orientation des deux molécules. GS5-1 est allongé en direction du trajet 2f, et GS2-2 l'est en direction du plus gros trajet S.

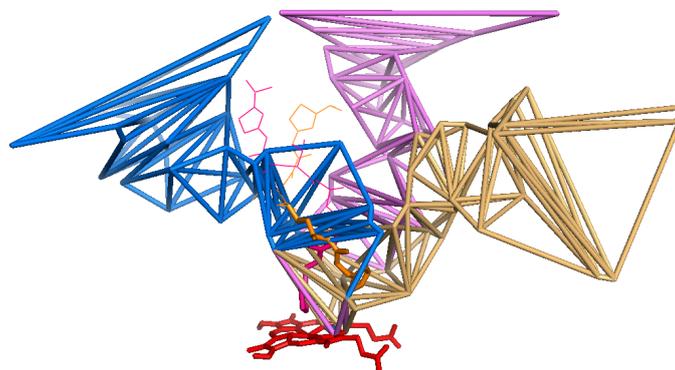
La structure 4K9U ouvre le bloc 2 (canaux S et 2f) et adopte la forme O2. La boucle FF' est désordonnée, non résolue aux rayons RX. Sa grande flexibilité lui permet de se déformer en conformation O2. La boucle FF' reconstruite (L210 à F215) est en surface de la protéine (figure 3.40, en haut). La structure 4K9U est la seule qui ouvre le canal S pour l'accueil d'un ligand. F213 et F215 font mouvement vers l'extérieur de la boucle FF' pour ouvrir le canal S et le canal 2f.

Le canal 2f s'ouvre en premier pour accueillir GS5-1, qui se lie au fer héminique. L'hydrophobicité des molécules GS5 ne permettant pas l'accueil de deux molécules dans le canal 2f, c'est le canal S, plus hydrophobe, qui s'ouvre à  $CV_{lim}=6.75\text{\AA}$ , en supposant qu'il resserre le canal 2f afin d'accueillir GS5-2. Après l'entrée de GS5-1 par le trajet 2f, celui-ci se cale dans une impasse du canal 2f, afin de libérer l'espace pour l'ouverture du canal S. De même, GS5-2 emprunte le trajet S, puis se cale dans une impasse du canal S pour optimiser l'espace d'occupation (figure 3.39). Les impasses des deux molécules sont opposées. La partie terminale de GS5-1 est dans une impasse plus étroite, qui se resserre pour ouvrir l'espace du canal S. La partie terminale de GS5-2 est dans une impasse plus large. GS5-1 suit le passage d'entrée du trajet 2f, se lie au site actif, puis se loge ailleurs. L'espace de passage 2f se resserre afin d'ouvrir l'espace S pour l'entrée de la GS5-2, qui emprunte le plus gros canal S du réseau à travers le trajet S. GS5-2 libère le trajet S puis se cale dans l'impasse : les molécules entrent par les trajets 2f et S, puis libèrent les passages d'entrée en se logeant dans les plus petites impasses, stabilisés entre les acides aminés.

Les parties modélisées des deux molécules ne sont pas logées dans les trajets S et 2f. Les molécules en interaction avec les acides aminés de la cavité sont alors instables. GS5-1 est logé dans le bas du canal 2f au niveau du centre actif, et sa direction laisserait supposer que GS5-1 emprunte le canal 2f et se logerait ensuite dans l'impasse. Sa modélisation serait alors erronée. De même, GS5-2 emprunterait le trajet S, car logé dans le bas du trajet S et allongé suivant la même direction.

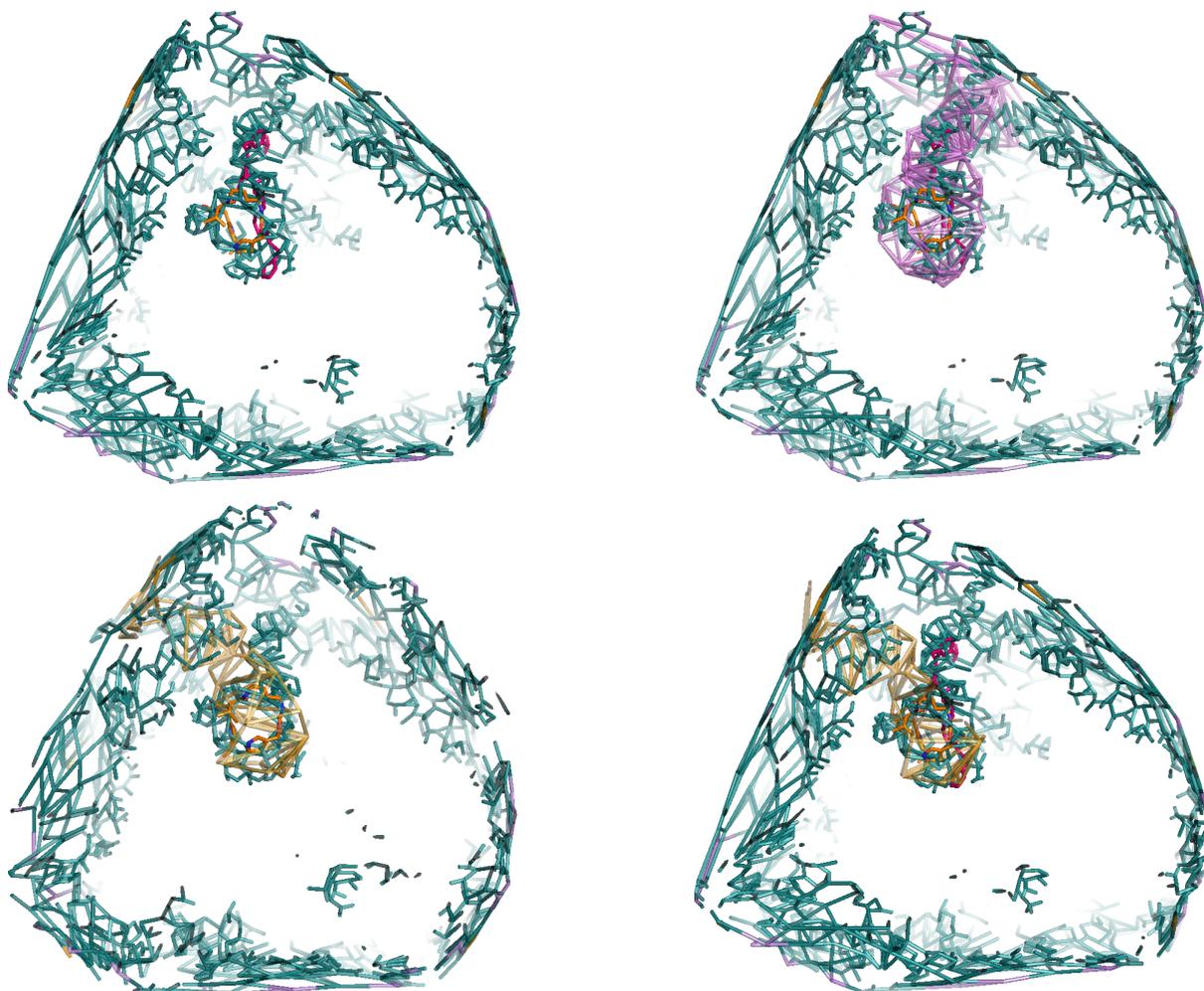


**Figure 3.40 :** Détails du complexe 4K9U formé entre le CYP3A4 et le composé GS5. En haut, à gauche, graphe facial à  $CV_{lim}$  et superposition de la boucle FF' du 3A4; à droite, parties visibles des deux molécules GS5, en boules-bâtons; les parties désordonnées sont modélisées en lignes, les acides aminés sont en tube mince. En bas : visualisation des sphères de van der Waals (deux orientations).



**Figure 3.41 :** Les 3 trajets S, 2f et 2a du complexe 4K9U, calculés à  $CV=6.75\text{\AA}$ ,  $6.25\text{\AA}$ , et  $5.75\text{\AA}$ , respectivement. Les parties réelles des 2 molécules sont contenues dans les trajets calculés, contrairement aux parties modélisées.

Les trajets S, 2f et 2a du complexe 4K9U sont visualisés figure 3.41.



**Figure 3.42 :** Réseau de canaux du complexe 4K9T formé entre le CYP3A4 et le composé GS4; graphe facial à  $CV_{lim}=6.75\text{\AA}$ . En haut : GS4-1 est en rose, GS4-2 est en orange; à droite : canal f ouvert (en violet). En bas : à gauche, représentation du canal 2a fermé avec GS4-2; à droite, représentation du canal 2a fermé avec GS4-1 et GS4-2.

### La flexibilité du désoxyritonavir augmente la stabilité de la forme O2

Le composé GS4 (figure 3.21) n'a pas de groupes latéraux. Comme GS5, il a deux sites d'accueil dans le 3A4, mais contrairement à ce dernier, les molécules GS4-1 et GS4-2 sont bien résolues dans la structure cristalline. GS4-1 est lié à l'hème et offre une conformation allongée, alors que GS4-2 est associé en position perpendiculaire [Sevrioukova & Poulos 2013b] (figure 3.42). GS4-2 est stabilisé dans les sites Ph-1 et Ph-2, respectivement via son groupement terminal et son groupement tête. Cette conformation de GS4-2 est plus stable par rapport à GS5-2, car elle est plus flexible en raison d'un encombrement moindre par ses chaînes latérales.

La poche du site actif est large et accueille les deux molécules. GS4-1 suit le trajet 2f depuis le fer hémistique (auquel il est lié) jusqu'à la sortie. GS4-2 entre par le même canal 2f, large d'ouverture jusqu'à  $CV=5.75$  (figure 3.42), et se replie, peut-être en attente de l'oxydation de GS4-1. GS4-2 est dans la cavité du site actif (figure 3.42), et vers lequel débouche le canal 2f. Lorsqu'on superpose un canal 2a sur le graphe facial hébergeant GS4-2, on observe que GS4-2 est replié dans la cavité du site actif, et qu'il n'y a pas d'issue vers la surface via le canal 2a. En effet, un goulot d'étranglement empêche l'ouverture du site actif via le canal 2a (figure 3.42).

Un second trajet 2f est calculé par CCCPP à  $CV=5.75\text{\AA}$  (figure 3.42). Celui-ci ne couvre qu'une partie de GS4-1, et pourrait être la voie d'entrée de GS4-2. Ces deux trajets 2f sont adjacents et forment un grand bloc d'ouverture 2f (figure 3.42). Le bloc 2f contient GS4-2. F213 borde le canal 2f, mais ceci n'apparaît pas en bordure du trajet 2a car CCCPP ne calcule que le trajet de coût minimal. F215 gêne l'ouverture du canal S, et c'est pour ça que les deux molécules n'y sont pas logées.

Avec un squelette plus flexible et moins ramifié, GS4-2 s'enroule le long d'une moitié de GS4-1, formant des liaisons H avec GS4-1 et avec les atomes de la chaîne principale de F108 et de R372 [Sevrioukova & Poulos 2013b]. Une autre caractéristique est l'orientation perpendiculaire des thiazoles de GS4-1 et GS4-2, qui favorise les interactions S- $\pi$  plutôt que celles des empilements  $\pi$ - $\pi$  observées avec les autres ligands. GS4-2 n'impose pas de gêne stérique au peptide 369–371, mais déplace l'hélice I avec la même amplitude que pour GS3. L'IPT de GS4-1 établit de multiples contacts hydrophobes avec les acides aminés aux alentours, mais il est partiellement exposé au solvant de par l'ouverture en forme O2 du 3A4. En effet, pour accueillir deux molécules et optimiser les interactions avec le groupement terminal, la boucle FF' adopte une conformation O2 [Sevrioukova & Poulos 2013b].

Les interactions polaires arbitrées par S119 sont importantes pour l'association du 3A4 au ligand, indépendamment de la formation de liaisons H. Même en l'absence de liaison H avec S119, le rôle de S119 sur l'affinité et la cinétique de liaison avec GS4 et GS5 est établi ainsi [Sevrioukova & Poulos 2013b] : (i) renforcement de liaison Fe–N des entités liées à GS4 et à GS5, (ii) stabilisation des deux complexes, et (iii) établissement du second site de liaison pour GS4 et GS5 (S119 élimine la liaison dans le site périphérique).

GS5 a une hydrophobicité supérieure (présence de deux méthyles) à celle GS4, ce dernier étant plus flexible et moins encombré que GS5. GS5-1 et GS4-1 s'orientent le long de deux canaux différents, et la boucle FF' adopte un repliement différent dans chacun des deux complexes. Ceci est peut-être dû à l'hydrophobicité. Les repliements de GS5-2 et de GS4-2 diffèrent : le premier est en longueur dans le canal 2f, et le second est en largeur, plus stable, dans la poche du site actif. Ceci peut-être dû à une stabilisation différente du squelette flexible et des chaînes latérales dans la cavité du 3A4. La réduction de la taille du groupement latéral accroît la liberté de mouvement dans le site actif, permettant à la protéine de "modeler" les inhibiteurs de différentes façons, de remplir les vides, et parfois de nécessiter la liaison à plus d'une molécule. Notre étude des canaux éclaire sur la distinction des routes suivies par les deux molécules, montrant que les deux molécules GS5 n'empruntent pas les mêmes canaux. Les modèles GS4-3A4 et GS5-3A4 fournissent un aperçu de la manière dont de multiples ligands peuvent se disposer dans le site actif. Malgré le contraste des modes de liaison au ligand (perpendiculaires dans le cas de GS4, et parallèles dans le cas de GS5), l'hypothèse de l'absence de différence dans le repliement global du 3A4 a été émise [Sevrioukova & Poulos 2013b]. CCCPP le

confirme, en explicitant l'ouverture en conformation O2 de la protéine, en réponse aux caractéristiques des molécules qu'elle accueille.

### 3.6.7 Conclusion sur les analogues du ritonavir

L'étude des analogues du ritonavir à l'aide de CCCPP nous a permis de mieux comprendre le contrôle de l'inhibition par ces analogues (sections 3.6.1 à 3.6.6). L'épaisseur du substrat est le critère de passage des analogues de ritonavir dans le 3A4. Le mode d'ouverture des canaux du 3A4 est indiqué table 3.13. Les interactions chimiques ont pour rôle d'optimiser et de stabiliser l'ouverture. L'ouverture des voies de passage est stabilisée par les interactions pour un meilleur accueil d'une ou plusieurs molécules dans la cavité du site actif. Les analogues du ritonavir adoptent différents modes d'orientation, et les acides aminés bordant les canaux dans différentes conformations, en fonction de leurs propriétés stéréo-électroniques. L'énergie de déformation de la protéine est compensée par les interactions stabilisantes du complexe.

La force de la liaison Fe-N et l'affinité d'interaction avec les groupes fonctionnels du ritonavir optimisent l'ouverture du 3A4. Les groupes de tête analogues au thiazole liés à l'hème déterminent la liaison Fe-N et influencent la stabilité du complexe. Le thiazole de tête et l'IPT terminal stabilisent l'ouverture du 3A4. Le thiazole de tête a un effet sur les interactions non-covalentes et sur le rôle modulateur de R212, dans le mode de liaison structural. L'IPT terminal joue un rôle important dans la stabilisation de la boucle FF' (donc dans l'ouverture du CYP). Les groupements hydrophobes latéraux déterminent le mode de liaison et l'orientation des désoxyritonavirs dans la cavité, ainsi que l'ouverture du 3A4. Le Ph-1 du désoxyritonavir est important pour l'affinité et l'inhibition. L'hydrophobicité et l'élargissement de la taille de la molécule provoquent l'ouverture en forme O1 du 3A4. L'accueil de deux molécules s'effectue sur deux sites de liaison, et provoque l'ouverture en forme O2 du 3A4. La flexibilité du désoxyritonavir augmente la stabilité du complexe en forme O2, accueillant deux molécules sur deux sites de liaison.

## 3.7 Interactions du CYP3A4 avec différents substrats et inhibiteurs

Les différents ligands induisent des changements conformationnels notables du 3A4. Nous en examinons certains.

### 3.7.1 En forme C

Le passage de la forme C apo à la forme C holo se traduit par l'élargissement du site actif et du canal 2a. Il y a déplacement dans la région de la boucle BC sans qu'il y ait un important réarrangement de la région FG, de l'hélice I et la boucle C terminale. CCCPP calcule des trajets 2a à trois valeurs différentes de CV : deux trajets 2a supplémentaires apparaissent en plus de celui à  $CV_{lim}$ ). Ces trois trajets ont des tétraèdres communs et d'autres spécifiques à chacun des trajets. Les trois trajets associés forment un trajet 2a élargi et un site actif élargi par rapport à structure apo.

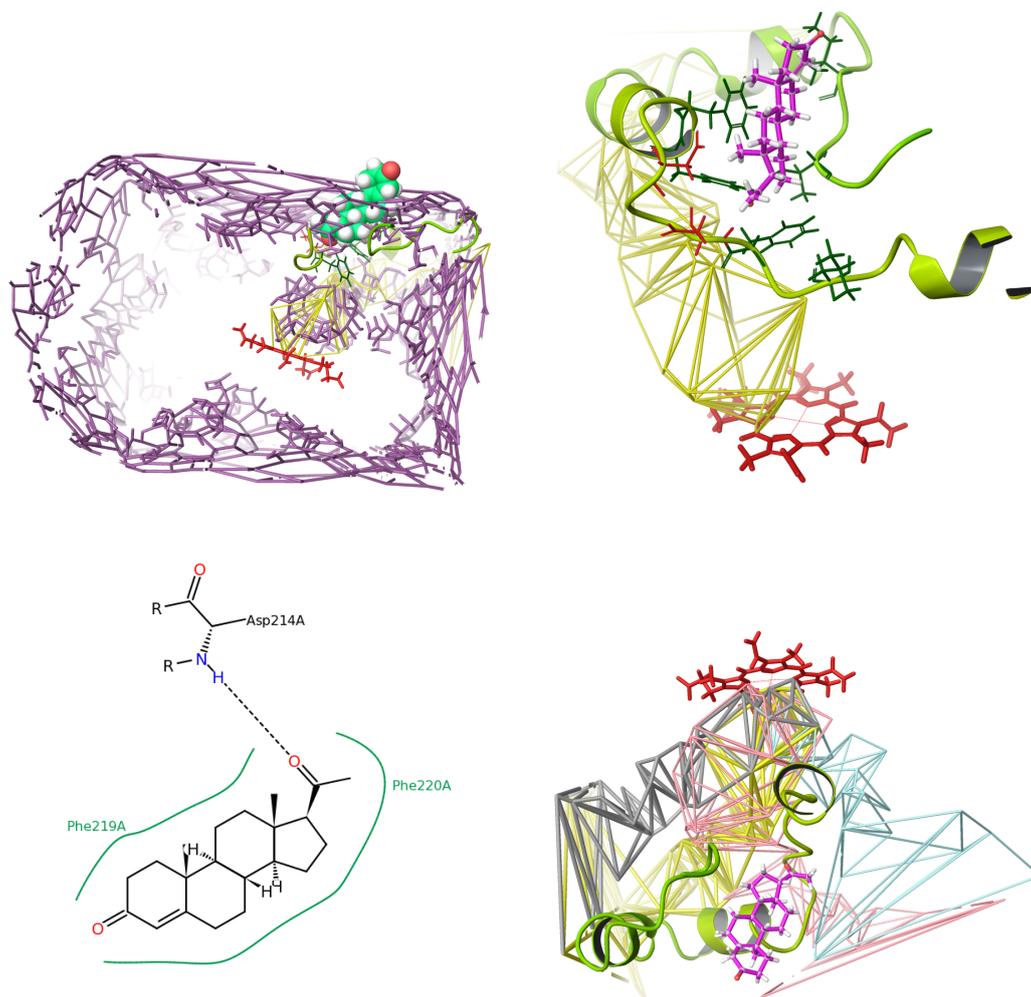
Composé	Code PDB	Mode de liaison (type)	Forme du 3A4 et canaux	Canal de passage
Ritonavir	3NXU	II	O1	2a
DTMCR	3TJS	II	O1	2a
GS8	4K9X	II	O1	2a
DAR	-	I	-	-
GS1	-	I	-	-
GS2	4I4G	II	O1	2a
GS3	4I4H	II	O1	2a
GS4	4K9T	II	O2	2f
GS5	4K9U	II	O2	S et 2f
GS6	4K9V	II	C	2a
GS7	4K9W	II	O1	2a

**Table 3.13** : Mode d'ouverture des canaux dans les complexes du CYP3A4 avec le ritonavir et ses analogues. Les composés GS1 à GS8 sont des désoxyritonavirs (DOR) car le groupement hydroxy du squelette a été supprimé. Les inhibiteurs de type II ont une liaison covalente avec le fer hémique via un azote ou un soufre, alors que ceux de type I n'établissent pas de liaison covalente avec le fer hémique [Schenkman *et al.* 1967, Ahlström & Zamora 2008]. Les composés DTCMR et GS8 permettent d'apprécier le rôle de la présence des groupements extrêmes thiazole et IPT. Le déazaritonavir (DAR) et les composés GS1 à GS3 permettent d'apprécier le rôle du thiazole. Les composés GS4 à GS7 permettent d'apprécier le rôle des groupements latéraux.

### Liaison de la progestérone sur un site périphérique

Dans le complexe cristallisé CYP3A4-progestérone (1W0F), la progestérone est liée dans une fissure à la surface du cluster Phe, dans un site de docking périphérique (figure 3.43, en haut à gauche), et assez accessible au solvant, cette localisation étant une caractéristique unique du 3A4 [Williams *et al.* 2004] (voir figures 4 et S3 dans cet article). La conformation du complexe du 3A4 avec la progestérone est très semblable à celle de la forme apo : la RMSD, calculée entre 1W0F et 1W0E (forme C) sur l'ensemble des C $\alpha$ , vaut 0.55Å. La poche du site périphérique est calculée à une valeur de CV supérieure à  $CV_{lim}$ , c'est-à-dire plus large que l'ouverture des canaux. On note aussi que le site périphérique est calculé également pour les structures apo (1TQN et 1W0E) à des valeurs de CV supérieures à  $CV_{lim}$  comme dans la structure liée à la progestérone. Dans cette zone de docking, cette molécule ainsi que d'autres molécules peuvent se lier avant translocation dans la cavité catalytique [Szklarz & Halpert 1997]. La liaison dans le site périphérique se fait sans perturbation du spectre de l'hème [Sevrioukova & Poulos 2012b], ni modification de l'ouverture de la protéine, comme calculé avec CCCPP. En effet, on voit qu'il n'y a pas d'élargissement du canal 2a par rapport à la forme apo 1TQN ( $CV_{lim}$  ne varie pas).

La progestérone est stabilisée dans la protéine par une liaison H entre l'oxygène de l'acétyle et l'azote de l'amide de D214, compacté entre les chaînes latérales de F219 et de F220 (interactions d'empilements  $\pi$ ) dans le cluster Phe, et apparaît assez accessible au solvant (figure 3.43, en haut à droite). La poche de docking périphérique est proche de L211 et de F213, ces dernières étant avec D214 impliquées dans un processus de coopérativité dépendant de la progestérone [Harlow & Halpert 1998, Domanski *et al.* 2001]. Ces trois



**Figure 3.43 :** Complexe du CYP3A4 avec la progestérone (code PDB 1W0F).

En haut : à gauche, interactions de la progestérone avec le 3A4 (figure générée par PoseView [Stierand *et al.* 2006]); les lignes pointillées noires indiquent la liaison H; les lignes vertes continues indiquent les interactions hydrophobes. A droite, graphe facial calculé à  $CV_{lim}=6\text{\AA}$ ; la progestérone est stabilisée dans la poche de docking périphérique par la région FG; il y a apparition du canal 2a et de l'ébauche du canal 2e; le trajet2a est délimité en jaune. La progestérone est représentée par des sphères de van der Waals avec les carbones en vert.

En bas : à gauche, progestérone liée au CYP3A4 dans le site périphérique au-dessus du cluster Phe, entre F57 (à l'arrière de la figure), F213, F219 et F220; la progestérone (en violet) est en interaction avec la boucle FF' et avec les acides aminés. Il y a une liaison H entre l'oxygène de l'acétyle de la progestérone et l'azote de l'amide de D214. A droite, la progestérone (en violet) dans le site de docking périphérique, par rapport aux trajets 2a (en jaune) et 2e (en gris); superposition des trajets 2f de 2V0M (en rose) et S de 1TQN (en bleu); ces 2 derniers trajets du bloc 2 ne sont pas ouverts dans le complexe 1W0F.

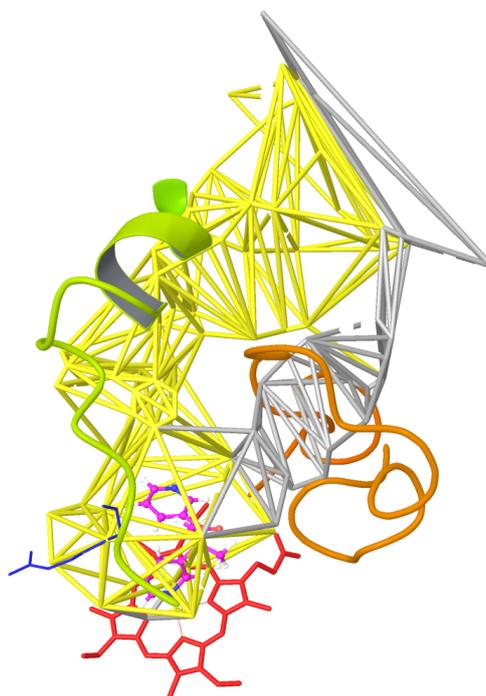
acides aminés impliqués dans la coopérativité (L211, F213 et D214), appartiennent à la boucle FF' et bordent le canal 2a. Le site de la progestérone n'est pas directement en contact des bouches des canaux 2a et 2f (figure 3.43, en bas à gauche). La progestérone se lie à 17Å au dessus du fer de l'hème, et du côté de la surface du 3A4 qui a le plus de chance d'interagir avec la membrane plasmique. La progestérone est dans une poche superficielle en haut du cluster Phe, séparée de l'hème par la région FG (figures 4 et S3 dans [Williams *et al.* 2004]). Un ancrage de cette zone dans la membrane serait peut-être la raison de l'entrecoupeure de la région FG observée dans le 3A4 (hélices F' et G' séparés par des boucles), qui permettrait l'ouverture des voies d'entrée des molécules (2a et 2f) entre ces structures secondaires plus flexibles, et se deversant sur la membrane. Au vu des dimensions de la progestérone (épaisseur de 4.10Å, volume de van der Waals de 321Å<sup>3</sup>) et au vu de son log(P), elle devrait rentrer par le canal 2a, et sa liaison au site effecteur peut être à l'origine du signal d'ouverture du canal d'entrée, et c'est ce qui serait la cause du petit changement conformationnel de la protéine induit par la progestérone et cité par [Williams *et al.* 2004].

Alors que la progestérone peut s'amarrer dans le site actif du 3A4 avec une orientation cohérente avec son site de métabolisation, rien n'appuie cette hypothèse. Certaines études indiquent que les boucles FG et BC, entre lesquelles s'ouvre un possible canal d'accès au substrat (2a, ou 2f suite à un déplacement de FF'), pourraient contribuer à la liaison avec la membrane et orienter le canal d'accès au substrat vers la surface de cette membrane [Graham-Lorence & Peterson 1995, Cosme & Johnson 2000]. La localisation du site périphérique dans la région FG est cohérente avec un rôle de reconnaissance initial du substrat, car elle se situe le long d'un chemin adéquat pour les composés qui vont directement de la membrane au site actif (par le canal 2a ou 2f) : voir figure 3.43 à droite, en haut et en bas.

### Interaction avec une petite molécule : la métyrapone

La métyrapone, methyl-2-di(pyridinyl-3)-1,2-propanone-1, C<sub>14</sub>H<sub>14</sub>N<sub>2</sub>O (volume de van der Waals estimé à 215Å<sup>3</sup>) est liée au fer de l'hème par l'azote de son groupe alkylpyridine, et présente une bonne complémentarité de forme avec le site actif du 3A4 (voir figure S2, [Williams *et al.* 2004]). A l'inverse de ce que l'on peut supposer, la liaison de la métyrapone au 3A4 ne révèle pas de changement conformationnel de la protéine : forme C, très proche des formes apo de 1W0E et 1TQN. La RMSD calculée entre 1W0G et 1W0E (en forme C) sur l'ensemble des C<sub>α</sub>, vaut 0.46Å. L'espace occupé par la métyrapone laisse de la place pour lier une molécule supplémentaire dans le site actif. De faibles déplacements de la boucle BC facilitent l'entrée du composé dans le site actif et augmentent le volume du site actif et du canal 2a.

Deux branches distales à l'hème se séparent dans la cavité catalytique, formant les deux canaux 2a et 2e, séparés par la boucle BC (figure 3.44). Avec CCCPP, on calcule des trajets 2a à 3 valeurs différentes de CV : 6.50Å, 6.00Å, et 5.25Å. Le trajet 2e est calculé à CV=5.5Å. La molécule est logée dans la partie intérieure du trajet 2a au niveau du site actif, partie commune à la partie intérieure du trajet 2e (Figure 3.74). On peut suggérer que la petite molécule accède le long du canal 2a, puis se réarrange dans la grosse cavité catalytique en conformation productive. On note que la chaîne latérale de R212 est orientée vers l'extérieur du site actif, de façon semblable à la forme apo 1W0E, et contraire à la position vers l'intérieur dans le 1TQN.



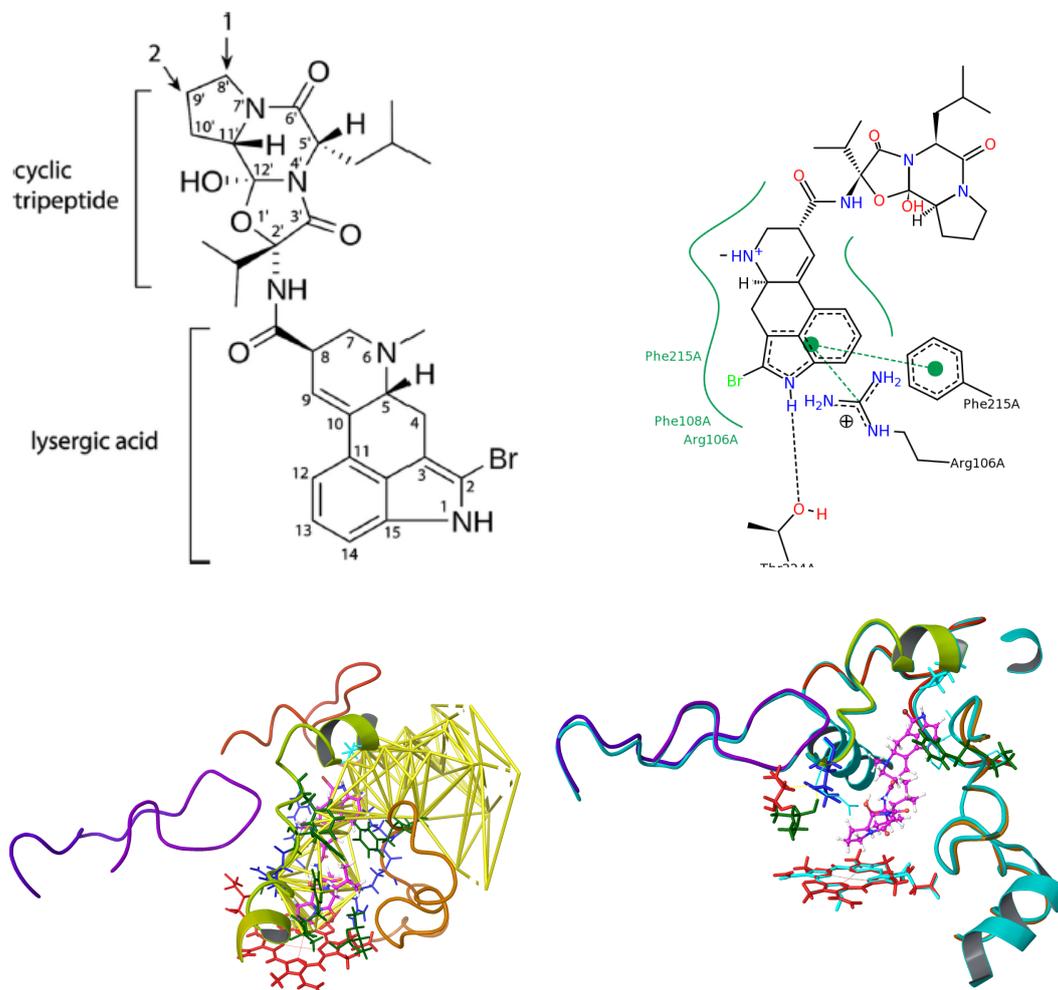
**Figure 3.44 :** Trajets 2a et 2e calculés pour le complexe du CYP3A4 avec la métyrapone (code PDB 1W0G). La métyrapone (en rose) est dans le site actif délimité par l'intérieur des canaux 2a et 2e.

La taille de la métyrapone laisse assez d'espace à des molécules supplémentaires pour se lier dans le site actif. Néanmoins d'après le diagramme (figure 3.15), nous supposons que l'énergie  $\pi$  de Hückel du métyrapone serait en cause du non accueil d'une 2ème molécule. Tout comme la progestérone, la métyrapone est un petit ligand (volume de van der Waals inférieur à  $550\text{\AA}^3$ ) : il faudrait analyser avec un plus gros ligand les changements conformationnels du 3A4.

### Interaction avec une molécule allongée et flexible : la bromoergocryptine

La bromoergocryptine (BEC) est un alcaloïde de l'ergot dérivé de l'acide lysergique comprenant un tripeptide cyclique (figure 3.45). La BEC est un agoniste du récepteur dopamine. C'est un substrat volumineux de grande affinité du 3A4.

Le volume de la cavité, estimé avec LIGSITE [Hendlich *et al.* 1997], est de  $563\text{\AA}^3$  [Sevrioukova & Poulos 2012b]. Aucun changement important de conformation du 3A4 n'est nécessaire pour accueillir une molécule de BEC : la RMSD avec la forme apo est évaluée à  $0.32\text{\AA}$  [Sevrioukova & Poulos 2012b]. Malgré sa taille importante (volume de van der Waals égal à  $547\text{\AA}^3$ ) et sa structure complexe, la BEC tient dans la cavité du site actif sans perturbations structurales significatives, en pénétrant dans le canal 2a ouvert, élargi par rapport à la forme apo ( $CV_{lim}$  varie de  $0.75\text{\AA}$ ). La BEC se lie au 3A4 dans une conformation étendue le long du trajet 2a (figure 3.45 en bas à gauche). Comme dans le cas de la métyrapone (section 3.7.1) et de GS6 (section 3.6.6), CCCPP détermine le trajet 2a à 3 différentes valeurs de CV :  $6.75\text{\AA}$ ,  $6.25\text{\AA}$ , et  $5.75\text{\AA}$ . Pour accueillir l'encombrante



**Figure 3.45** : Structure chimique de la bromoergocryptine (BEC) et trajet 2a calculé dans le complexe 3A4-BEC (code PDB 3UA1).

En haut : à gauche, les deux sites primaires d'hydroxylation de la BEC sont indiqués par des flèches (figure extraite de [Sevrioukova & Poulos 2012b]). A droite, interactions de la BEC avec le 3A4 (figure générée par PoseView [Stierand *et al.* 2006]); les lignes pointillées noires indiquent les liaisons H; les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$  et cation- $\pi$ .

En bas : à gauche, trajet 2a calculé à  $CV_{lim}$  pour le 3A4 lié au BEC (en violet); la BEC est allongée dans le trajet 2a, bordé par les structures secondaires et les acides aminés (en tubes colorés selon leurs propriétés physico-chimiques). A droite, comparaison des structures secondaires autour du canal 2a dans le 3A4 lié au BEC avec la structure apo (code PDB 1TQN, en cyan); la BEC (en rose) est allongée dans le canal 2a (non délimité); il y a déplacement du peptide P107-G109 sur la boucle BC.

molécule de BEC, le peptide P107–G109 (boucle BC) se déplace de 2.4Å, sans qu'il y ait un important réarrangement dans la région FG ni dans l'hélice I et dans la boucle C-terminale (figure 3.45 en bas à droite), comme c'est le cas pour l'érythromycine (voir section 3.7.3). Cette différence peut provenir de la conformation allongée de la BEC et de sa plus grande flexibilité conformationnelle (5 liaisons pivotantes).

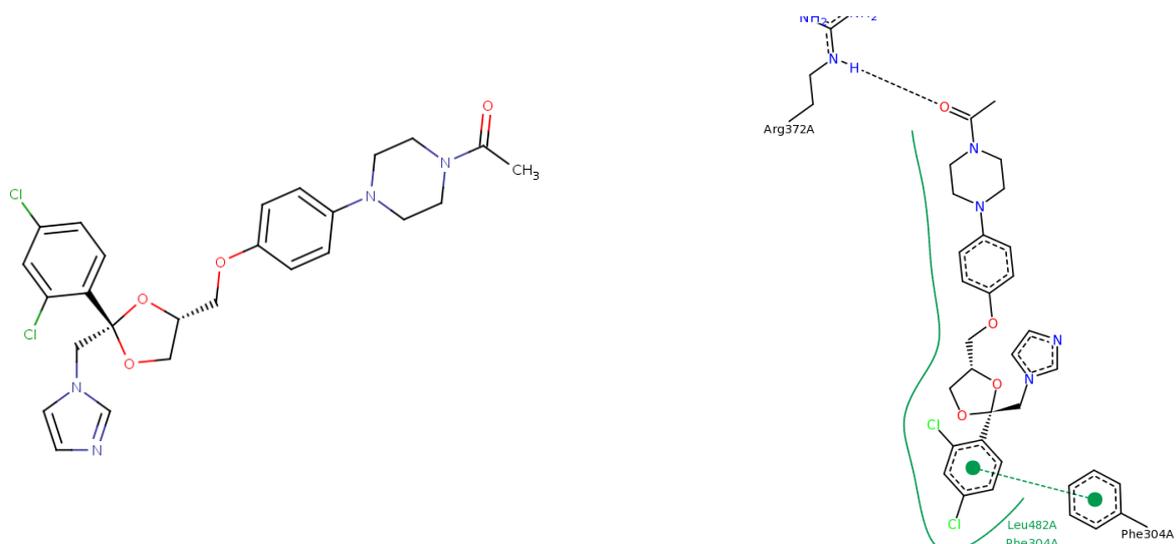
Le groupement acide lysergique de la BEC est coincée entre les chaînes latérales parallèles de R106 et de F215, et offre une liaison H avec T224 (hélice G') via l'azote N1 (figures 3C et 3D dans [Sevrioukova & Poulos 2012b]). Le groupe tripeptide, situé dans la cavité catalytique, n'établit ni interactions polaires spécifiques ni interactions électrostatiques, mais offre des contacts van der Waals extensifs avec R105, R212, I301, F304, A305, A370 et R372. Ces acides aminés entourent la molécule sur toute la surface du site actif, ce qui est révélé par leur position bordure du trajet 2a dans sa partie intérieure dans la cavité, hormis I301 qui demeure dans la cavité bordant quant à lui le canal 2a (figure 3.45 en bas à gauche).

R212 and T224 définissent l'affinité et participent à la liaison en orientation productive de la BEC dans le 3A4 [Sevrioukova & Poulos 2012b]. Ces deux acides aminés cruciaux dans le mécanisme de liaison de la BEC, bordent le trajet 2a calculé par CCCPP (figure 3.45, en bas). Il a été proposé un mécanisme de liaison de la BEC avec le 3A4 en 3 étapes. La BEC possède différentes conformations et peut d'abord s'associer avec un site périphérique distant de l'hème, sans perturber le spectre de l'hème [Isin & Guengerich 2006]. Dès la translocation dans la cavité du site actif, via le canal 2a, la BEC établit une seule liaison H avec T224, via le groupement acide lysergique à l'entrée de la cavité (figure 3.45 en haut à droite), qui redirige le groupe tripeptide vers le fer de l'hème, contrôlé par R212. Une fois dans la cavité catalytique, la conformation du tripeptide de la BEC dépend de celle de R212. La conformation de ce tripeptide nécessiterait ou non un réajustement conformationnel de la chaîne latérale de R212 (une rotation d'environ 180° autour de la liaison amide), avant de s'insérer dans le site actif et de se lier à l'hème. Un nouveau rotamère de R212 peut s'en suivre pour optimiser la position de la BEC, qui influencerait sur l'équilibre de spin (voir figure 8 dans [Sevrioukova & Poulos 2012b]). Ceci fait valoir le rôle de l'acide aminé flexible R212 entourant le canal d'accès dans la réactivité des substrats (figure 3.45 en bas à droite). La nouvelle conformation de R212 est stabilisée par deux liaisons H, entre d'une part le groupe guanidine et d'autre part le carboxy de E308 et l'oxygène du carboxy de I369.

Dans le cas de l'ouverture du canal 2f, celui-ci s'accompagne du réarrangement du squelette et de la chaîne latérale de R212. R212 compose alors le tronc commun des canaux 2a et 2f, alors que T224 représente la fourche qui sépare les 2 canaux à l'entrée. Dans le cas d'une seule molécule de BEC, seul le canal 2a est impliqué, mais dans le cas d'une seconde molécule, l'implication de ces deux derniers acides aminés (R212 et T224) dans l'ouverture du canal 2f en suivant un mécanisme de coopérativité, est une hypothèse cohérente (voir section 3.7.3).

### 3.7.2 En forme O1

L'agrandissement du site actif via l'élargissement du canal 2a s'effectue en raison du mouvement de la région de la boucle BC, du cluster Phe, et de la boucle FF' (voir section 3.6).



**Figure 3.46 :** Formule structurale du ketoconazole. A droite, les interactions avec le 3A4 (figure générée par PoseView [Stierand *et al.* 2006]); les lignes pointillées noires indiquent les liaisons H; les lignes vertes continues indiquent les interactions hydrophobes, et les lignes pointillées vertes indiquent les interactions  $\pi$ - $\pi$ .

### 3.7.3 En forme O2

Il y a ouverture des canaux 2f et S, avec ou sans ouverture du bloc 1. L'agrandissement du site actif se fait par l'ouverture du bloc 2 en raison du mouvement du cluster Phe et du déplacement important de la boucle FF'.

#### Interaction du CYP3A4 avec deux molécules de kétoconazole

Le kétoconazole est un antimycotique inhibiteur du 3A4. Il possède un groupe imidazole (fortement polaire) et il est très flexible car il a 9 liaisons pivotables (figure 3.46). L'étude de l'échantillonnage conformationnel du kétoconazole a montré qu'il existait de nombreuses familles de conformères [Meslamani *et al.* 2009, Benkaidali *et al.* 2012]. Le kétoconazole n'est pas très sélectif et inhibe plusieurs autres isoformes des familles 1B, 2B et 2C [von Moltke *et al.* 1998]. C'est une molécule assez volumineuse, avec un volume de van der Waals de  $450\text{\AA}^3$ . Dans le complexe (code 2V0M) du 3A4 cristallisé avec deux kétoconazoles, les épaisseurs de chacune des deux molécules sont respectivement de  $5.19\text{\AA}$  et  $6.52\text{\AA}$  (calculs effectués avec RADI, cité section 2.3.1). Les squelettes de 2V0M et 1TQN ne diffèrent pas significativement :  $\text{RMSD}=0.749\text{\AA}$  sur 1397 atomes du squelette. Les principales différences sont dans la boucle FF', et il y a un petit changement conformationnel de la boucle BC et de la boucle C-terminale.

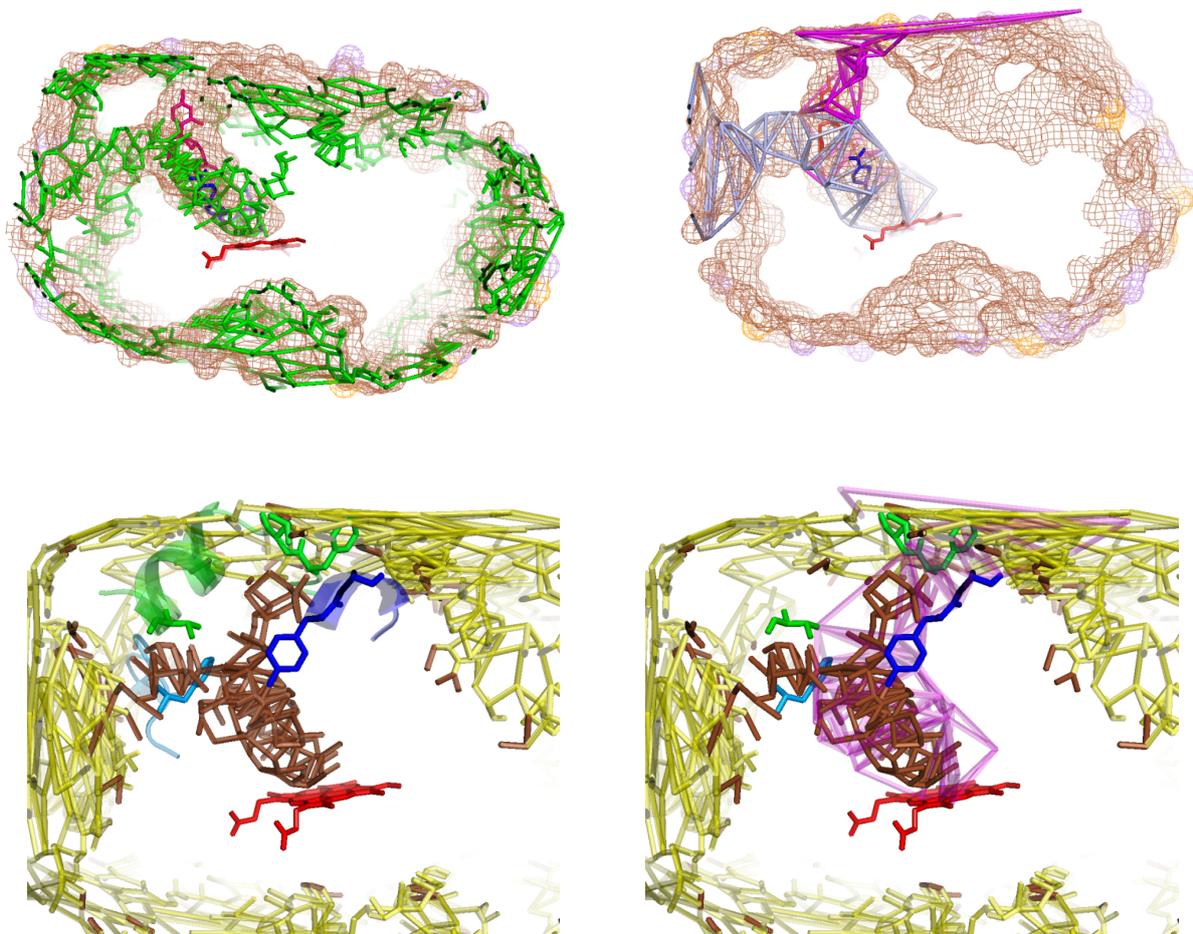
Le kétoconazole est une molécule hydrophobe avec un  $\log(P)$  d'environ 4.25. D'après sa taille, il serait accueilli par le canal 2a suivant le diagramme des formes présenté sur la figure 3.15. Mais une intéressante caractéristique de la structure du complexe qu'il forme avec le 3A4 est la présence de deux molécules liées dans la cavité du site actif dans une configuration en tandem antiparallèle (figure 3.47). Le bloc 2 s'ouvre alors. L'un des kétoconazoles a son groupe imidazole proche de l'hème : l'azote N3 du cycle imidazole se

lie au fer de l'hème [Davis *et al.* 2009]. Le cycle aromatique bi-chloré est dans le site actif et la chaîne latérale est dans le tronc commun des canaux 2a/2f. Le second kétoconazole a une orientation opposée, avec le groupe acétyle près du site actif et le cycle aromatique bi-chloré près de l'entrée du canal 2f (figure 3.48). Un élargissement du canal 2a est nécessaire à l'accueil d'un ligand volumineux ou d'une deuxième molécule (figure 3.48). Pour accueillir les deux kétoconazole, le 3A4 entreprend un changement conformationnel significatif, particulièrement dans la boucle FF' et autour du cluster Phe. L'hélice I, la boucle BC et la boucle C-terminale opèrent aussi un changement conformationnel (figure 3.49). Le volume accessible au solvant dans la cavité du site actif augmente de  $1650\text{\AA}^3$  par rapport à la forme apo [Ekroos & Sjögren 2006].

Les poches de surface sont visibles pour  $CV > CV_{lim}$  et le canal calculé pour 2V0M apparaît à  $CV_{lim} = 7\text{\AA}$ , et donc il est évident qu'il y a un goulot d'étranglement à l'entrée des trajets 2a et 2f. Vers le canal 2a, ce goulot est au niveau de F108 (sur la boucle BC) et de T224 (dans l'hélice F'). Vers le canal 2f, il est au niveau de Y53 et H54 (tous deux dans l'hélice A), et de F215 et F216 (tous deux dans la boucle FF'). Les déplacements concernent ces acides aminés situés dans des régions flexibles de la protéine, et donc induisent l'ouverture ou la fermeture des goulots. Ces 6 acides aminés bordant les goulots d'étranglement sont tous le long du canal 2f : quatre sont localisés à l'entrée (dans l'hélice A et dans la boucle FF'), F108 est dans la partie commune à 2a et 2f, et T224 représente la fourche de séparation des canaux 2a and 2f (voir figure 3.47 en bas, et figure 3.49).

Le canal 2a est le premier retrouvé par CCCPP à  $CV_{lim} = 7\text{\AA}$ , et donc c'est le plus large. Un second canal est trouvé, 2f, dans le bloc 2, situé entre la région FG et la boucle C-terminale, et contenant le second kétoconazole. Dans le complexe 2V0M, les canaux 2a et 2f ont une importante partie commune puis se séparent à la sortie de l'hélice F' au niveau du groupe hydroxy de T224. Le canal 2f est calculé avec CCCPP à  $CV = 5.75\text{\AA}$ , et est orienté comme le second ketoconazole. A  $CV_{lim}$ , le canal 2f apparaît dans une impasse, non connectée à l'extérieur (figures 3.47 en bas, et 3.48). Un goulot d'étranglement à  $CV > 5.75\text{\AA}$  ( $\pm 0.25\text{\AA}$ ) obstrue la connexion à l'extérieur de la protéine. Une ouverture à  $CV = 5.75\text{\AA}$  devrait permettre l'entrée du second ligand. Pendant l'entrée du substrat ou la sortie du produit, un tunnel du ligand doit s'ouvrir pour permettre aux ligands d'accéder ou de s'échapper du site actif [Yu *et al.* 2013]. Le trajet 2a contient dans sa partie incluse dans la cavité du site actif, le premier kétoconazole, lié à l'hème (figure 3.47 en haut, et figure 3.48). On peut supposer que ce canal constitue le premier chemin trouvé par le ligand. Mais le canal 2a, dans toute sa partie de la cavité, est commun avec le canal 2f, et ne se sépare de ce dernier qu'à l'entrée. On peut suggérer que l'élargissement du canal 2a est dû à l'ouverture du bloc 2 pour s'adapter à l'accueil de deux molécules. De plus, l'orientation du premier kétoconazole ne conforte pas l'hypothèse de son entrée par le canal 2a, du fait que ce premier kétoconazole est parallèle au second, et comme ce dernier, il est dirigé vers l'ouverture du trajet 2f. Le premier kétoconazole serait situé non pas dans le canal 2a, mais plutôt dans le tronc commun au niveau de la cavité des trajets 2a/2f (figure 3.48). Le positionnement du second kétoconazole à l'entrée du trajet 2f nous permet alors de proposer sans ambiguïté son accès par le canal 2f. Le second kétoconazole est à l'intérieur du canal, et pointe depuis l'hème suivant la direction du canal 2f. Le canal 2e est calculé dans sa gamme d'apparition à  $CV = 5.5\text{\AA}$ , et comme pour toutes les structures 3A4, il ne contient aucune molécule (figure 3.50).

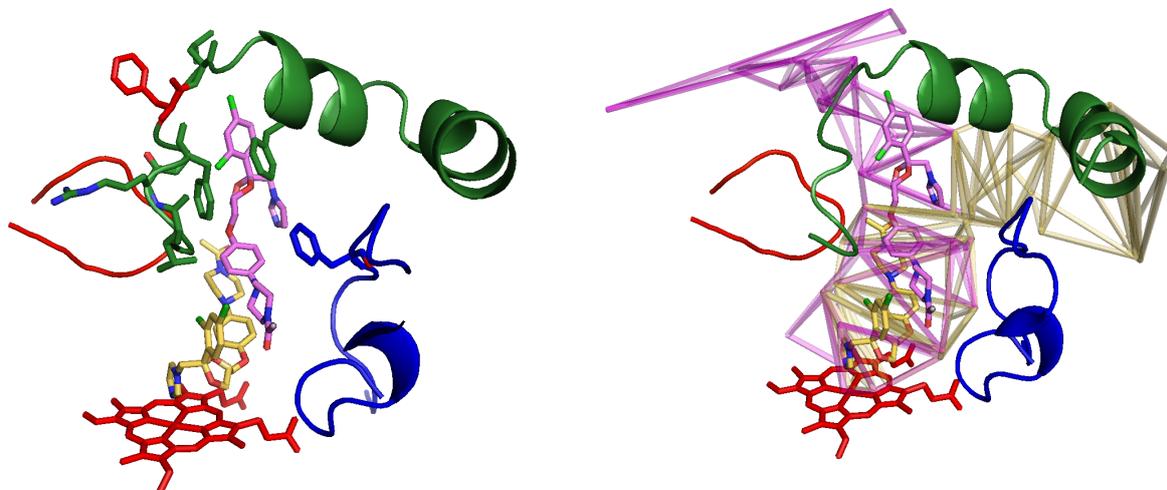
Pour lier le kétoconazole, la protéine subit un grand changement conformationnel, tel que les structures secondaires sont déplacées par rapport à la forme apo (figure 3.49).



**Figure 3.47 :** Graphe facial du canal du 3A4 holo (code PDB 2V0M).

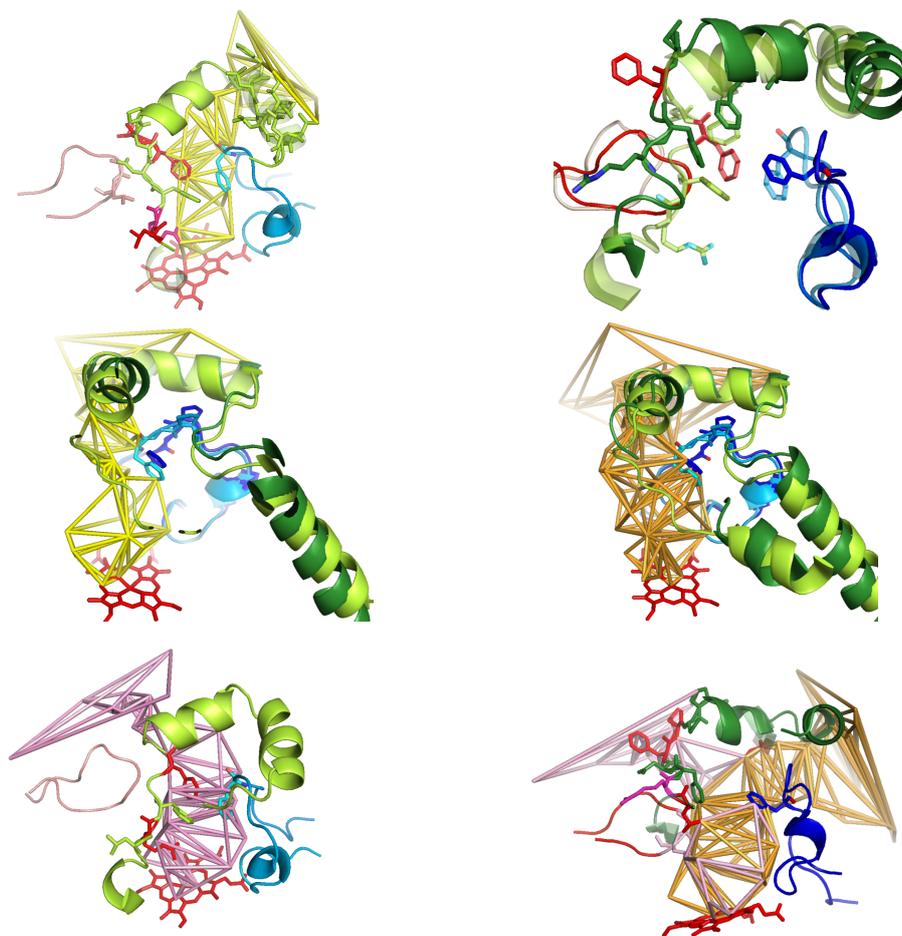
En haut : à gauche, graphe facial à  $CV=5.75\text{Å}$  (maillage en marron), avec superposition du canal du 3A4 apo (code PDB 1TQN) à  $CV_{lim}=6\text{Å}$  (en bâtons verts); les deux molécules de kétoconazole (en bleu et rouge) sont contenues dans le canal du 2V0M; à droite, idem, pour le canal du 2V0M avec superposition des trajets 2a (en jaune) et 2f (en violet; un des atomes au niveau de la bouche du trajet 2f est au bord d'un tétraèdre très aplati).

En bas : à gauche, graphe facial à  $CV=7\text{Å}$ ; les poches sont représentées en bâtons jaunes, le canal principal menant au site actif est en bâtons marrons; l'ébauche du canal 2f n'est pas connectée à l'extérieur; les acides aminés des goulets d'étranglement à l'entrée des canaux 2a et 2f sont représentés en bâtons; à droite, idem avec superposition du trajet 2f calculé à  $CV=5.75\text{Å}$ . L'atome éloigné du triangle aplati du trajet 2f a été supprimé.



**Figure 3.48** : Trajets 2a et 2f bordés par les structures secondaires du complexe 2V0M. Les acides aminés et les deux kétoconazoles sont représentés en tubes. A gauche, les deux kétoconazoles sont dans les canaux 2a et 2f, entourés des structures secondaires délimitant ces canaux (ces derniers ne sont pas délimités). A droite, idem, avec délimitation des trajets 2a and 2f (les acides aminés ne sont pas affichés).

La superposition du complexe avec le kétoconazole avec la forme apo donne une RMSD de 1.6Å pour l'ensemble des  $C\alpha$ . Les changements majeurs de conformation se situent dans les hélices F et G et dans les boucles intermédiaires. Le cluster Phe hydrophobe de la forme apo est décomposé, et certaines chaînes latérales hydrophobes sont exposées au milieu environnant [Ekroos & Sjögren 2006] (figure 3.50 à droite). L'ouverture du canal 2f est dû au basculement de la boucle FF' vers l'extérieur de la protéine. Une conséquence de la nouvelle conformation de la boucle FF' est que R212, qui se trouve dans le site actif de la structure en forme apo, se localise à la surface du complexe formé avec le kétoconazole. R212 change de côté pour libérer de la place au tronc commun des canaux 2a et 2f, et permettre ainsi aux deux kétoconazoles d'accéder à l'hème. R212 bascule vers l'extérieur et n'est alors plus en mesure d'interagir avec le ligand dans la cavité. A l'inverse, L211, qui se trouve à la surface de la protéine en forme apo, se localise dans le site actif du complexe formé avec le kétoconazole (figure 3.49). Cette orientation se corrèle avec l'observation que la substitution de L211 par une phénylalanine affecte la coopérativité homotropique dans la liaison avec la testostérone [Harlow & Halpert 1998]. Dans la structure holo, F215 est déplacé vers la bouche du canal 2f, et capture le ligand hydrophobe à l'interface avec le solvant aqueux. Ce basculement de l'acide aminé hydrophobe vers la surface a un double intérêt : (i) lever la gêne stérique afin de permettre l'élargissement du tronc 2a/2f au niveau de la cavité (figure 3.49, en bas à gauche), (ii) favoriser l'interaction avec la molécule hydrophobe en surface (figure 3.49, en bas à droite). L'exposition induite par le ligand d'une partie hydrophobe de la protéine à la surface fournit une possible explication à la nécessité d'une zone tampon d'assez faible polarité. D'autres changements notables sont un déplacement de 4Å de la boucle C-terminale, et une distorsion de la crevasse de l'hélice I autour de Y307 (le  $C\alpha$  est déplacé au plus de 2Å) [Ekroos & Sjögren 2006]. Une

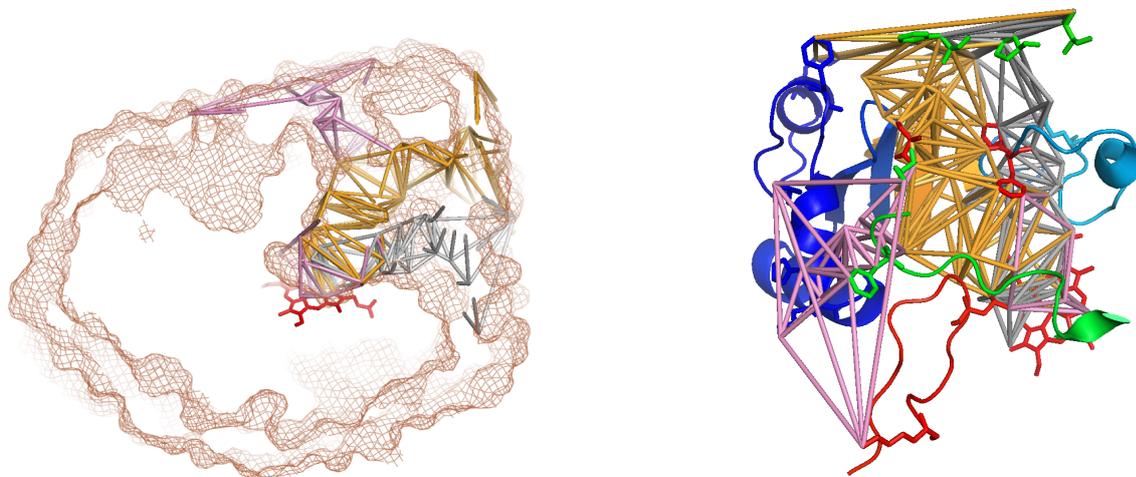


**Figure 3.49 :** Comparaisons des formes apo (1TQN) et holo (2V0M). Les formes apo sont en couleurs claires, les formes holos en couleurs sombres. Les structures secondaires de la forme holo (foncées) sont décalées par rapport à celles de la forme apo (claires).

En haut : à gauche, trajet 2a dans 1TQN avec quelques un de ses acides aminés en bordure dans les boucles FF' et BC; la bouche du canal est délimitée par les acides aminés hydrophobes de l'hélice G; les acides aminés portes F213 et F215 (boucle FF') et F108 (boucle BC) délimitent le trajet 2a en forme C; à droite, 1TQN superposé à 2V0M; les structures secondaires sont en rubans et les acides aminés sont en bâtons; les acides aminés portes sont L216 et R212; 1TQN est transparent, indiquant la conformation du 3A4 avant l'entrée des deux kétoconazoles; ceci suggère ce que pourraient être les déplacements des régions BC et FG et de la boucle C-terminale, impliqués dans l'ouverture des canaux d'accès 2a and 2f.

Au milieu : superposition de 1TQN et de 2V0M au niveau des blocs FG et BC, qui délimitent le canal 2a; pour plus de clarté, un morceau de la boucle FF' est supprimé; à gauche, superposition du trajet 2a de 2V0M.

En bas : à gauche, trajet 2f de 2V0M superposé sur 1TQN, montrant l'obstruction stérique des acides aminés en bordure du trajet, dans les boucles FF' et BC (R212, et les acides aminés portes F213, F215 et F108), ainsi que dans l'hélice F', pour ouvrir le canal 2f; à droite, trajets 2a et 2f dans 2V0M avec quelques acides aminés en bordure; la partie commune aux trajets 2a et 2f est délimitée par les boucles FF' et BC et par la boucle C-terminale; la bouche de 2a est délimitée par l'hélice G' et celle de 2f par le boucle FF'; 2a et 2f se séparent à la sortie par T224 dans l'hélice F'; la partie commune à 2a et 2f est en orange, et les parties séparées en rose clair; les acides aminés portes, F213 et F215 dans la boucle FF' et F108 dans la boucle BC, bordent la conformation en état ouvert de 2a et de 2f.



**Figure 3.50** : Trajets 2a, 2f, et 2e calculés avec 2V0M pour les valeurs de CV respectivement égales à 7Å, 5.75Å, et 5.5Å. A gauche : superposition du canal (en maillage marron) calculé à CV=5.5Å. A droite : trajets 2a (en orange), 2f (en rose), et 2e (en gris), avec les structures secondaires en bordure (feuillelet  $\beta_1$ , régions FG et BC, et boucle C-terminale); les bouches de 2a et 2e sont voisines, puis les trajets se séparent vers P107 et F108 dans la boucle BC; les trajets 2a et 2f sont séparés par T224 dans l'hélice F'; ceci montre la localisation des bouches des trajets 2a, 2f, et 2e, par rapport à la membrane plasmique; 2a conduit à la membrane et 2e au cytosol; 2f est transmembranaire.

distorsion similaire est aussi observée dans la structure du complexe formé par P450eryF avec le kétoconazole [Cupp-Vickery *et al.* 2001].

Les ouvertures de la protéine et les deux molécules de kétoconazole sont stabilisées par des interactions hydrophobes. Le kétoconazole lié forme une liaison H avec R372 et des empilements  $\pi$  avec F304. Le kétoconazole non lié forme une liaison H avec S119. La fonction cétone du kétoconazole est dans une poche polaire délimitée par les chaînes latérales de R106, R372, et E374. Puis d'autres interactions hydrophobes sont stabilisantes, comme les empilements  $\pi$  entre le cycle chlorobenzyle du kétoconazole lié et le cycle de F304 (sur l'hélice I) : voir figure 3.46 et figure 6B dans [Sevrioukova & Poulos 2013a]. Il existe des liaisons H entre les groupes cétoniques et les chaînes latérales de S119 (boucle BC bordant le tronc commun 2a/2f), et de R372 (feuillelet  $\beta_1$ , bordant le canal 2a). Le kétoconazole non lié s'adapte bien à l'intérieur de la poche du site actif, établissant de vastes interactions hydrophobes et van der Waals avec la boucle K- $\beta_1$  et le cluster Phe via les groupes latéraux, et est entièrement entouré de solvant. Des études expérimentales [Pearson *et al.* 2006], et théoriques ont prouvé une coopération homotrope positive dans la liaison du kétoconazole au 3A4 : il y aurait au moins trois étapes consécutives, commençant probablement par une association à un site périphérique [Isin & Guengerich 2007]. D'autres études ont démontré que l'association du premier kétoconazole augmente l'affinité pour la liaison au deuxième de 5kJ/mol, ce qui explique la coopérativité [Bren & Oostenbrink 2012]. Chaque structure du 3A4 simulée (sans ligand, ou liée avec un, ou deux kétoconazoles), était capable de lier l'inhibiteur

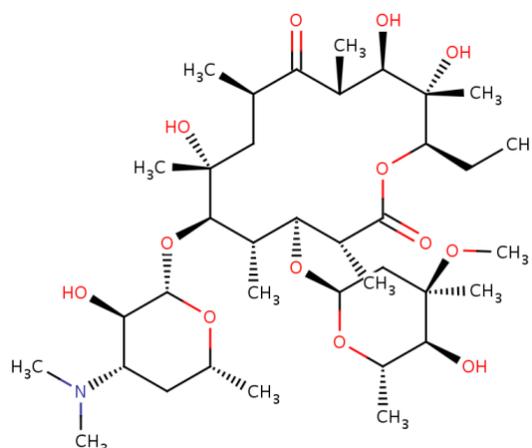
dans seulement quelques conformations sélectionnées, ce qui appuie ainsi le mécanisme de sélection conformationnel. Ceci démontre l'importance de la déformation de la structure du 3A4 pour l'accueil des ligands. C'est en totale concordance avec nos résultats obtenus avec CCCPP : nous confirmons cette hypothèse par, respectivement, l'ouverture initiale en forme C, l'ouverture en forme C avec élargissement du canal 2a, et l'ouverture du bloc 2.

Ces études suggèrent que la structure du 3A4 liée au kétoconazole ne soit pas un artéfact de cristallisation, mais reflète bien un complexe enzyme-inhibiteur biologiquement pertinent [Sevrioukova & Poulos 2013a]. La structure 2V0M conforte l'évidence d'une liaison simultanée au ligand pour un P450 mammalien, et indique comment surviennent de telles interactions entre molécules médicamenteuses, même relativement volumineuses. Il a été montré que la conformation du 3A4 lié au kétoconazole s'accordait bien avec les prédictions des sites de métabolisation faites pour beaucoup de substrats du 3A4 [Afzellus *et al.* 2007]. Ceci prouve le rôle de la déformation structurale du 3A4 dans son processus de sélection des ligands, et appuie le mécanisme de sélection conformationnel.

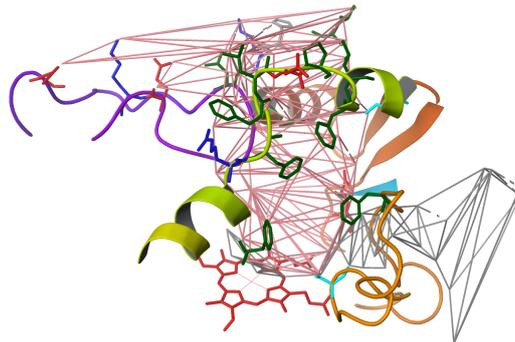
### Liaison à une molécule volumineuse : l'érythromycine

Afin d'étudier la large ouverture en forme O2, nous examinons le changement conformationnel du 3A4 induit par un gros ligand, l'érythromycine ( $CV=7.2\text{\AA}$ , volume de van der Waals égal à  $727\text{\AA}^3$ ). L'érythromycine est un macrolide antibiotique, et un ligand du 3A4 de type I, faible. L'érythromycine a été choisie en raison de ses propriétés en terme de coordination à l'hème, de ses dimensions, et de sa distribution de charges très différente de celle du kétoconazole, ce qui peut augmenter les chances d'identifier d'autres interactions avec le 3A4 [Ekroos & Sjögren 2006].

Dans le complexe 2J0D, la liaison au ligand fait subir un important changement conformationnel à la protéine. Contrairement à la BEC (section 3.7.1), le complexe 3A4-érythromycine présente un réarrangement conformationnel dans la région FG (c'est le plus visible), dans l'hélice I, et la boucle C-terminale se place pour effectuer le passage à la forme



**Figure 3.51** : Formule structurale de l'érythromycine.

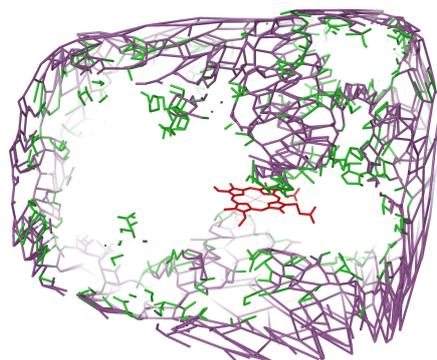


**Figure 3.52 :** Trajets 2f et 2e avec leurs structures secondaires en bordure. Le trajet 2f (en rose) représente une association de MCPs apparus à  $CV=8.75\text{\AA}$ , et à  $6.75\text{\AA}$ . Le trajet 2e (en gris) apparaît à  $CV=5.50\text{\AA}$ . Certains acides aminés en bordure des trajets sont également représentés.

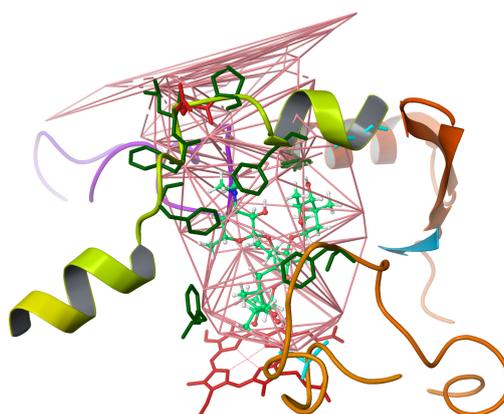
O2, avec une augmentation du volume du site actif de plus de 80% (figure 3.52). La boucle FF' est mal définie, signe de sa grande instabilité et de son manque d'interaction avec l'érythromycine [Ekroos & Sjögren 2006]. Nous l'avons donc reconstruite (voir table 2.2). Le déplacement de l'hélice F a provoqué l'ouverture d'une large fissure, la délimitation du site actif devient arbitraire, mais il est estimé que son volume augmente d'environ  $2000\text{\AA}^3$  [Ekroos & Sjögren 2006].

Dans 2J0D, un grand canal 2f s'ouvre à  $CV=8.75\text{\AA}$  (figure 3.52). CCCPP ne calcule pas de canal 2a, mais celui-ci est visible sur le graphe facial et connecte l'hème à l'extérieur à  $CV=6.5\text{\AA}$  (figure 3.53). Le canal 2a est donc bien ouvert dans la structure 2J0D, et son ouverture est située entre des structures secondaires très flexibles : le feuillet  $\beta_1$ , la boucle BC et l'hélice F'. C'est le mouvement de ces dernières qui modulerait l'ouverture du canal 2a. Le canal 2e est calculé à  $CV=5.5\text{\AA}$ . F215 bascule vers l'extérieur pour libérer l'ouverture du canal 2f, mais reste dans le corps du canal (et non en surface) : F215 n'intervient pas chimiquement dans le recrutement de la molécule polaire à la surface de la protéine. F213, orienté vers la cavité du site actif, constitue le corps du canal 2f. Il a été suggéré que les quatre acides aminés du cluster Phe (57, 213, 220, 304) stabilisent la molécule [Ekroos & Sjögren 2006]. Ceci est confirmé par le calcul fait avec CCCPP : en effet, les chaînes latérales de ces quatre acides aminés, en plus de F108 et F215, tapissent le large canal 2f (figure 3.54).

Le calcul effectué avec CCCPP nous révèle que l'érythromycine interagit dans la même zone que le kétoconazole. La liaison à deux molécules de kétoconazole ou à une molécule volumineuse d'érythromycine, provoque l'ouverture du bloc 2, tout en maintenant l'ouverture du bloc 1. La différence est que l'érythromycine opère une très grande ouverture du bloc 2, telle que le canal 2f est plus large dans le cas de l'érythromycine (CV varie de  $3\text{\AA}$ ). Cette différence s'observe aussi par l'importance de l'ouverture du bloc 1, mais à l'inverse, l'ouverture du canal 2a, qui s'effectue avec une variation de CV de  $0.5\text{\AA}$ , est plus petite pour le complexe 2J0D que pour le 2V0M. On peut suggérer que l'élargissement du canal 2f par le mouvement des structures secondaires en bordure, empièterait sur la zone d'ouverture du canal 2a, et deviendrait plus faible. Les structures 2V0M et 2J0D représentent deux conformations O2 distinctes du 3A4, de largeurs différentes. La liaison des deux molécules de kétoconazole au site actif du 3A4 et l'indication nette de multiples modes de liaison pour l'érythromycine ont des conséquences



**Figure 3.53** : Graphe facial des canaux calculés à  $CV=6.5\text{\AA}$  pour le complexe CYP3A4-érythromycine (code PDB 2J0D). Le graphe facial est en violet. Il y a ouverture du canal 2f (qui apparaît initialement à  $CV=8.75\text{\AA}$ ) et du canal 2a; superposition du graphe facial calculé à  $CV=5.5\text{\AA}$  (en vert) où il y a en plus ouverture du canal 2e.



**Figure 3.54** : Trajet 2f calculé pour le complexe CYP3A4-érythromycine (code PDB 2J0D). L'érythromycine, liée au site actif, est en boules-bâtons, en vert clair. Le large trajet 2f (en rose) représente un ensemble de quatre MCPs apparaissant de  $CV=8.75\text{\AA}$  à  $CV=6.75\text{\AA}$ . Les structures secondaires en bordure sont représentées.

sur l'interprétation des données cinétiques atypiques souvent observées pour le 3A4. Cette cinétique atypique est expliquée par la liaison de la molécule logée dans l'ouverture dans le bloc 2 calculée par CCCPP.

La boucle FF' bascule vers la surface de la protéine pour border l'ouverture du canal 2f. Ceci, comme pour le complexe 2V0M, a un intérêt stérique en libérant l'espace d'ouverture, puis un intérêt physico-chimique en favorisant le recrutement de la molécule à l'interface cytosol-membrane plasmique. En effet, les acides aminés apolaires de la boucle FF' et de l'hélice F' (L216, P218 et L221), constituent alors la bouche du trajet 2f vers la membrane plasmique, afin d'équilibrer le caractère polaire au niveau du cytosol (K55, E470, E486, K487) : voir figure 3.52.

L'érythromycine, dont le macrocycle est proche de l'hème, est liée au 3A4 dans le canal 2f suivant une orientation non productive. Elle est métabolisée par N-déméthylation du groupe D-désosamine, qui n'est pas exposé à l'hème mais en est à  $17\text{\AA}$  (voir figure 5B

dans [Sevrioukova & Poulos 2013a]). L'érythromycine n'est pas dans sa conformation la plus stable, n'interagissant pas optimalement par des interactions non covalentes avec les acides aminés bordant le canal 2f. De ce fait, la molécule ne se caractérise pas par de fortes interactions avec la protéine, contrairement au cas du complexe 3A4-kétoconazole. L'érythromycine offre une faible affinité, et la perte de liaison avec le 3A4 expliquerait l'absence de la molécule dans le site actif de la seconde chaîne de l'unité asymétrique de 2J0D et dans la structure 1TQN [Yano *et al.* 2004, Ekroos & Sjögren 2006]. Il n'y a pas d'interactions polaires spécifiques entre l'érythromycine et la protéine, mais une forte proximité avec les chaînes latérales des phénylalanines du cluster Phe suggère que le complexe est partiellement stabilisé par des interactions hydrophobes. Il a été suggéré d'autres modes de liaison avec l'érythromycine [Ekroos & Sjögren 2006]. Ces données confirment l'observation que le substrat peut se localiser dans le site actif sans orientation préférentielle pour la liaison [Cameron *et al.* 2005]. L'érythromycine avait été arrimée dans la cavité catalytique de la structure 1TQN en forme C dans une orientation productive [Yano *et al.* 2004]. La mauvaise conformation de la molécule dans la structure cristallisée est peut-être due à la cristallisation sans la membrane plasmique. En effet, le 3A4 est lié à la membrane, et le canal 2f est ouvert sur la région transmembranaire. Le macrocycle a plus d'affinité pour la membrane plasmique que le groupement osidique azoté, qui en premier pénètre le canal 2f s'ouvrant sur la région transmembranaire, puis ensuite le macrocycle, plus affiné à la membrane, pénètre en second. Les conditions de cristallisation en milieu aqueux ont pour effet de retenir le groupement azoté plus hydrophile par rapport au macrocycle. Les différences structurales entre les chaînes A et B sont surtout dans la région F'G' (hélice/boucle/hélice : acides aminés 219 à 237), au niveau de l'ancrage dans la membrane plasmique, où les différences moyennes entre les positions des C $\alpha$  sont d'environ 1Å.

## 3.8 L'hypothèse multi-canaux dans l'accueil multi-substrats du 3A4

### 3.8.1 Mouvement concerté des canaux du CYP3A4 et canalisation du substrat

La boucle FF', qui détermine les 3 formes du 3A4, borde les deux principaux canaux d'accès du bloc 1 (canal 2a) et du bloc 2 (canal 2f). Les mouvements de ses acides aminés modulent l'ouverture/fermeture de ces canaux. Le rôle des canaux 2a et 2f est de donner accès au site actif, tandis que le rôle des canaux 2e et S demeure peu clair. Une synchronisation entre l'accès du substrat et la sortie du produit via des routes différentes est possible, mais encore non prouvée. Dans les formes apo et holo, le canal 2e est ouvert simultanément avec les canaux 2a et 2f, indiquant que les canaux d'accès peuvent s'ouvrir alternativement par éloignement de la région FG de la boucle BC, sans affecter le canal 2e. Ceci est en accord avec une étude sur la structure de CYP2C5 (1NR6) [Lee *et al.* 2003]. De nouvelles études sont nécessaires pour confirmer ou non cette hypothèse. Les Canaux 2e auraient deux rôles potentiels dans le 3A4. Le premier est pour la sortie des produits, comme pour les autres canaux [Cojocar *et al.* 2007]. Dans les différentes conformations du 3A4, globalement les canaux 2a, 2f, S et 2e s'ouvrent indépendamment, mais avec des largeurs différentes. L'éloignement de la boucle BC de l'hélice G' n'a pas induit la

fermeture de certains canaux pour laisser d'autres s'ouvrir. Cependant, on peut suggérer que l'ouverture d'un canal peut influencer le rétrécissement d'un autre canal. Ceci serait en accord avec la conclusion proposée pour le CYP51 [Podust *et al.* 2004], indiquant qu'un canal 2a fermé et un canal 2e ouvert suggèrent que 2e doit se fermer pour laisser 2a s'ouvrir. Cette situation est due à l'hélice B' qui s'éloigne de l'hélice G, ce qui comprime la boucle BC et donc ferme le canal 2e. Le second rôle peut être lié au solvant aqueux. Le canal 2e pourrait servir à relarguer l'eau hors des canaux afin de libérer de l'espace pour les ligands. Dans la forme apo, les canaux 2a, 2e et S sont simultanément ouverts pour laisser passer l'eau. Dans la forme holo, les canaux 2a ou 2f sont occupés par le substrat alors que 2e et S contiennent une chaîne de molécules d'eau entre l'hème et la surface de la protéine. Le canal 2e, et peut-être le canal S, peuvent permettre à l'eau de sortir du site actif après être apparue comme produit de la peroxygénation, en accord avec ce qui a été observé pour les structures du CYP152A1, du CYP2C5 et du CYP2B4 [Lee *et al.* 2003].

La multiplicité des canaux suggère la canalisation du ligand depuis et vers les P450s enchâssés dans (ou sur) la membrane. La topologie du P450 favorise la formation d'un canal du côté distal de l'hème. Le côté proximal est le site de liaison probable de la réductase, et correspond aux plus petits canaux trouvés avec CCCPP. Ainsi, la liaison proximale à la réductase n'interfererait pas directement, par gêne stérique, avec toute voie d'accès distal définie par la topologie du P450. Par contre, il est évident que la liaison à la réductase altère la dynamique de la protéine du côté distal [Rui *et al.* 2006]. Elle pourrait ainsi moduler l'ouverture des canaux et aider à synchroniser le transfert d'électrons avec les étapes réactionnelles de liaison au substrat et de relargage des produits. Nos calculs confirment que, pour le 3A4, les canaux 2a et 2f sont vraisemblablement les canaux principaux d'entrée du substrat et de sortie des produits, et qu'en plus il y aurait deux canaux secondaires, 2e et S. La polyvalence du mécanisme de sortie du 3A4 se comprend parfaitement, lorsqu'on considère que cette isoforme est ubiquitaire ("*promiscuous*") et oxyde une large variété de substrats. Cette variété requiert, à son tour, un bon nombre de voies de sortie.

### 3.8.2 Flexibilité du 3A4 dans l'accueil multi-substrats

Le 3A4 existe sous multiples conformations cinétiquement distinctes, en présence ou absence de molécule(s) de substrat, effecteur, ou inhibiteur [Koley *et al.* 1995, Koley *et al.* 1997, Davydov *et al.* 2003]. Les trois conformations du 3A4 identifiées dans cette étude contribueraient à l'allostérie. Il a été proposé l'existence de trois sous-poches dans le site actif du 3A4, liants des groupes de 10 substrats-sondes classés en fonction de leur taille ou leur classe chimique [Kenworthy *et al.* 1999]. Le 3A4 se déformerait pour rendre accessible ces sous-poches au(x) molécule(s). Ceci est en total accord avec notre étude avec CCCPP, ouvrant de nouvelles cavités en adoptant l'une des trois formes.

Le calcul des canaux par CCCPP nous a permis d'identifier les acides aminés en bordure de ces canaux, et de repérer les acides aminés clés identifiés comme étant impliqués dans la réactivité et la coopérativité du 3A4. Les acides aminés clés de bordure contribuent à orienter les substrats dans le site actif [Sevrioukova & Poulos 2015]. En plus de la couverture physico-chimique des canaux et l'orientation des acides aminés, la matérialisation des canaux et leur topologie (leur localisation au sein de l'enzyme) ont permis d'éclairer l'implication des acides aminés clés et les phénomènes observés dans les

mutagenèses (modification cinétique, structurales, etc). Ceci nous donne une vision de la liaison coopérative du 3A4 au niveau moléculaire. En effet, les ligands se lient dans les sites effecteurs (loin, ou dans le site actif). Ces sites sont bordés par les canaux calculés par CCCPP.

Nous supposons qu'après le recrutement du ligand à l'entrée des canaux, et que, en raison de contraintes stériques, l'orientation du ligand ne change pas jusqu'à ce qu'il atteigne le site actif. On suppose qu'à l'arrivée au site actif, le ligand peut changer d'orientation en fonction de ses dimensions. Nous avons observé lors de notre étude que les canaux se modulent (élargissement/rétrécissement, suivant les valeurs de CV) et modulent avec eux le site actif. Aussi, l'existence de deux blocs d'accès ou de sortie, et du canal 2e, seraient impliquées dans l'allostérie. L'adaptabilité du site actif et des canaux capables de lier plusieurs composés simultanément explique la grande diversité chimique des substrats 3A4, et la large spécificité.

Un mécanisme "porte" des canaux a été défini pour le 3A4. Nous pointons l'importance du rôle de ce mécanisme dans la sélectivité et la coopérativité du 3A4. Les acides aminés portes ont un rôle fonctionnel dans le contrôle d'accès et le maintien de la flexibilité des canaux. En effet, les acides aminés portes jouent un rôle dans la régulation du trafic entre le site actif et la surface, par le contrôle des changements conformationnels, permettant l'accès vers et depuis le site actif [Pravda *et al.* 2014]. Les acides aminés portes, hydrophobes et volumineux, sont principalement des phénylalanines bordants les canaux. Néanmoins, ces acides aminés portes ne possèdent pas tous un groupement aromatique, et peuvent être une lysine ou une valine, ou une arginine chargée positivement. Malgré la conservation du repliement du CYP, il y a une grande diversité dans le nombre et dans les dispositions des portes, incluses dans les régions variables des SRS situées sur les bords des canaux (voir table 3.10). Le mécanisme "porte" n'est ainsi pas conservé dans les canaux pour les différentes isoformes, et joue un rôle dans la sélectivité. Par exemple, la spécificité de la sortie du ligand par le canal 2a est contrôlée par les acides aminés portes, qui diffèrent pour chaque isoforme [Winn *et al.* 2002]. Les acides aminés portes sont empilés parallèlement (interactions  $\pi$ ) ou perpendiculairement, et le produit brise ces interactions lors de la sortie de l'enzyme. Pour passer la porte, la molécule oxydée brise ces interactions hydrophobes pour former à son tour principalement des contacts hydrophobes avec les acides aminés portes. La forme du produit oxydé et son orientation sélectionnent ainsi le canal par lequel il sortirait préférentiellement [Fishelovitch *et al.* 2009], ou inversement, nous pouvons supposer que c'est le canal qui sélectionne son produit. Le contrôle d'accès a lieu au niveau des goulots d'étranglement, par les importants mouvements effectués par ces acides aminés volumineux, aliphatiques ou aromatiques [Pravda *et al.* 2014]. La chaîne latérale de la phénylalanine peut pivoter, effectuant des mouvements d'ouverture des portes dans le passage du ligand par le canal. Chaque canal du 3A4 est constitué d'une ou deux phénylalanines empilées par des interactions  $\pi$ - $\pi$ , qui contrôlent l'accès par l'ouverture et la fermeture du canal. L'analyse des structures cristallines par CCCPP et le rôle du mouvement des acides aminés portes dans le changement de formes, suggèrent que le réseau des portes régule coopérativement l'ouverture et la fermeture des tunnels. Les dimensions des canaux et l'orientation des acides aminés portes définissent les voies physiquement possibles pour l'entrée du substrat ou la sortie du produit. Il y a trois phénylalanines (108, 213, 215) et une arginine (212), qui constituent les portes du tunnel 2a : ils couvrent le ligand dans l'état C lié, tandis qu'ils sont ouverts dans un état O,

permettant aux ligands de circuler. Ainsi, les acides aminés de bordure agiraient comme un filtre sélectif pour des substrats ou des produits.

Trois types de mécanismes portes des canaux ont été identifiés par une étude théorique : (i) rupture d'empilements  $\pi$  entre phénylalanines, (ii) rupture de contacts hydrophobes aliphatiques (comme dans le canal S entre R212 et L482, ou dans le canal 2e entre F108 et I120), et (iii) rupture de liaison hydrogène (comme dans le canal 2e entre R105 et S119). Le mécanisme prédominant est l'empilement  $\pi$  [Fishelovitch *et al.* 2009]. Une étude récente a montré l'importance des interactions  $\pi$ - $\pi$  entre les acides aminés portes et les ligands : ces interactions favorisent la liaison coopérative au second ligand et favorisent l'orientation du premier ligand dans le site actif, en vue de métabolisations ultérieures [Fa *et al.* 2015]. Ceci expliquerait l'influence du paramètre d'énergie  $\pi$  de Hückel dans notre diagramme proposé figure 3.17. D'après notre étude, les sites de liaison (distincts or communs) se trouvent, entre autres, dans les canaux ouverts des blocs 1 ou 2 bordés par les SRS (voir table 3.10). La liaison au substrat y est favorisée par les interactions  $\pi$ - $\pi$  avec les acides aminés portes des canaux.

Les acides aminés clés qui influencent la liaison au substrat, la coopérativité, et la régiosélectivité, ont été d'abord identifiés par mutagenèse dirigée, basée sur un modèle par homologie tridimensionnelle du 3A4. Ils incluent F108, S119, I120, L211, D214, I301, F304, A305, T309, A370 et L373 [Sevrioukova & Poulos 2013a]. Ces acides aminés appartiennent aux SRS, lesquels bordent les canaux calculés par CCCPP (voir table 3.10). Ces études ont suggéré que les sites de liaison au substrat et à l'effecteur sont adjacents et se trouvent dans une grande cavité. Les premières déterminations des structures du 3A4 aux rayons X [Yano *et al.* 2004, Williams *et al.* 2004], ont montré que la plupart des acides aminés prédits se trouvent, soit dans le site actif, soit dans le canal d'accès (voir figure 2 dans [Sevrioukova & Poulos 2013a]). En effet, ces acides aminés clés se situeraient dans les sites de liaison ou dans les sites effecteurs, sur les structures secondaires (boucle BC, boucle FF', hélice I, boucle K-feuillet  $\beta_1$ -brin 4), entourant les canaux 2a, 2f et 2e, au niveau des deux régions adjacentes, le site actif et le corps des canaux.

Nous relierons l'implication des interactions  $\pi$  des acides aminés portes dans la déformation des structures étudiées par CCCPP, avec leur rôle dans la coopérativité reportée dans la littérature. La liaison coopérative au diazépam a été observée, via son interaction avec F304, dans le conformère C de 1W0E mais pas de 1TQN, cela en raison du positionnement différent de R212. En effet, dans 1TQN, R212 (dont la chaîne latérale de R212 est face au site actif), interagit fortement avec F304 et contrecarre l'impact de l'effecteur en diminuant l'accessibilité de l'effecteur à F304 (car gêné par R212). Au contraire dans 1W0E, l'interaction avec l'effecteur stabilise F304 et ses environs, spécialement F213, et favorise l'orientation du substrat actif par diminution de la distance entre le site d'oxydation du substrat et l'espèce réactive de l'hème [Fishelovitch *et al.* 2007]. Contrairement à 1TQN, F304 est accessible et borde le canal 2a au niveau du site actif dans 1W0E (car R212 est orienté vers l'extérieur), ce qui permet à F304 d'avoir une interaction  $\pi$  avec le ligand, et donc de jouer un rôle clé dans la liaison coopérative. Le site périphérique au-dessus du cluster Phe, défini uniquement dans le 3A4, auquel la progestérone est liée dans la structure cristalline [Denisov *et al.* 2009], représente un pôle d'interactions  $\pi$  cruciales dans la coopérativité. La double couche lipidique a été trouvée comme ayant un rôle stabilisant des portes aromatiques ouvertes, dans une simulation de la membrane liée au 2C9 [Cojocar *et al.* 2011].

## 4 PRESENTATION DE L'ARTICLE [Benkaidali *et al.* 2014]

Cet article, reproduit sur les pages suivantes, présente une revue des principales méthodes géométriques de calculs des canaux, puis résume les principales caractéristiques de CCCPP (dont la comparaison des modèles sphériques et cylindriques des ligands), ainsi que les résultats obtenus sur les canaux des isoformes 1A2 et 3A4.

*©Reproduced by permission of Oxford University Press,  
licence number 3555320676232, Jan 24, 2015.*

## Computing cavities, channels, pores and pockets in proteins from non-spherical ligands models

Lydia Benkaidali<sup>1,2</sup>, François André<sup>3</sup>, Boubekour Maouche<sup>4</sup>, Pridi Siregar<sup>5</sup>, Mohamed Benyettou<sup>6</sup>, François Maurel<sup>1</sup> and Michel Petitjean<sup>7,\*</sup>

<sup>1</sup>ITODYS, UMR 7086, CNRS, Université Denis Diderot, Paris 7, <sup>2</sup>Programme Doctoral International, Université Pierre & Marie Curie, Paris 6, <sup>3</sup>CEA/DSV/IBiTec-S/SB2SM, CNRS, UMR 8221, Saclay, France, <sup>4</sup>LPCTCI, Université des Sciences et de la Technologie Houari Boumediène, Algiers, Algeria, <sup>5</sup>IBC, Integrative BioComputing SARL, Place du Granier, 35135 Rennes-Chantepie, France, <sup>6</sup>LAMOSI, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, Oran, Algeria and <sup>7</sup>MTi, UMR-S 973, INSERM, University Denis Diderot, Paris 7, France

Associate Editor: Anna Tramontano

### ABSTRACT

**Motivation:** Identifying protein cavities, channels and pockets accessible to ligands is a major step to predict potential protein–ligands complexes. It is also essential for preparation of protein–ligand docking experiments in the context of enzymatic activity mechanism and structure-based drug design.

**Results:** We introduce a new method, implemented in a program named CCCPP, which computes the void parts of the proteins, i.e. cavities, channels and pockets. The present approach is a variant of the alpha shapes method, with the advantage of taking into account the size and the shape of the ligand. We show that the widely used spherical model of ligands is most of the time inadequate and that cylindrical shapes are more realistic. The analysis of the void parts of the protein is done via a network of channels depending on the ligand. The performance of CCCPP is tested with known substrates of cytochromes P450 (CYP) 1A2 and 3A4 involved in xenobiotics metabolism. The test results indicate that CCCPP is able to find pathways to the buried hemic P450 active site even for high molecular weight CYP 3A4 substrates such as two ketoconazoles together, an experimentally observed situation.

**Availability:** Free binaries are available through a software repository at <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>

**Contact:** michel.petitjean@univ-paris-diderot.fr

Received on July 24, 2013; revised on October 30, 2013; accepted on November 3, 2013

### 1 INTRODUCTION

Protein concavities play an important role for biological functions at the molecular level such as ligand binding. Knowledge of protein–ligand binding sites and their access routes contributes to optimize virtual screening of chemical libraries and then to propose new potential drugs. The concavity containing the active site may be located at various places in the protein. When it is at the surface of the protein, the concavity is called a pocket, and it is called a cavity when the active site is buried inside the protein, in which case the route to the active site is called a channel.

In any case, defining the shape of the protein needs to locate the boundary between the protein and its void parts. Many algorithms were proposed in the literature to solve this non-trivial problem. In a recent review (Pérot *et al.*, 2010), 20 geometric algorithms were reported, plus 10 algorithms based either on genomic information or on energy calculations. These latter being devoted to protein–ligand complexes energy calculation by docking (see Sperandio *et al.*, 2006 for a review), they are outside the scope of this article. We report in Table 1 a non-exhaustive list of 35 geometric algorithms and their variants.

There are several ways to classify them. The main criteria are: (i) the computation is dependent or not on probes spheres assumed to represent ligands, (ii) the computation is based either on a rectangular grid or on a Delaunay triangulation or on none of them. It has to be noticed that building the Delaunay triangulation of  $N$  input vertices in the 3D space consists practically in partitioning the smallest convex polyhedron enclosing the  $N$  input vertices into non-overlapping adjacent tetrahedra with endpoints edges at input vertices, such that none interior of the spheres circumscribed to these tetrahedra contains any of the input vertices (Edelsbrunner, 1987). The dual structure of the Delaunay triangulation is the Voronoi diagram (Edelsbrunner, 1987), which, roughly speaking, is the union of the  $N$  polyhedral cells defined from the set of the planes mid-perpendicular to the sides of these tetrahedra. To each of the input vertices  $K$ ,  $1 \leq K \leq N$ , is associated its polyhedral cell (possibly unbounded) which is the domain containing all points of the space closest to  $K$  than to any of the  $N - 1$  other vertices. This important property explains why Voronoi tessellations were used to visualize the shape of proteins, such as in Voro3D (Dupuis, 2003; Dupuis *et al.*, 2005).

Grid methods offer several drawbacks. The orientation of the 3D grid is arbitrary, and most of the time an increase of the precision by 10 needs to increase the number of nodes by  $10^3$ , thus increasing computing times by 1000. It is the case, when the measures of cavity volumes are estimated from the number of nodes or from the number of voxels (i.e. the 3D analogs of pixels). When the grid nodes are in a graph in which paths are sought to identify channels, the number of paths to be enumerated increases exponentially with the number of nodes. To overcome this problem, the user can define an arbitrary upper bound

\*To whom correspondence should be addressed.

**Table 1.** Main geometric algorithms for pockets or channels detection

Name	References	Method
Cavity search	Ho and Marshall (1990)	Gridded slices and voxels
POCKET	Levitt and Banaszak (1992)	Grid (marching cubes); probe sphere
VOIDOO	Kleywegt and Jones (1994)	Grid and voxels; probe sphere
SURFNET	Laskowski (1995)	Grid; fitting spheres into the spaces between atoms
Masuya and Doi	Masuya and Doi (1995)	Grid; probe sphere; cavity boundaries are spherical surfaces that are parts of the probe sphere
LIGSITE	Hendlich <i>et al.</i> (1997)	Grid; variant of POCKET; visualization from VRML files
Variant in LIGSITE	Stahl <i>et al.</i> (2000)	Grid; clusters of cavity points; pocket atoms were defined as being the protein grid points closest to any surface point
LigandFit	Venkatachalam <i>et al.</i> , 2003	Grid; nodes of cavities boundaries are got with a probe cube (orientation fixed by the grid)
Travel Depth	Coleman and Sharp (2006)	Grid; channels are defined from shortest path in grid to convex hull boundary
CAVER	Petřek <i>et al.</i> (2006)	Grid; minimal cost paths channels; the contribution to the cost function at each node is arbitrarily set to the inverse of a geometrically interpretable quantity in order to avoid a meaningless maximal cost path search
PocketPicker	Weisel <i>et al.</i> (2007)	Grid; pockets are clusters of grid points; an integer buriedness index is defined from the presence or absence of a protein atom at fixed distance along 30 search directions
PocketDepth	Kalidas and Chandra (2008)	Grid; shortest Euclidean travelling paths in grid to given protein atoms are identified; pockets are clusters of site points
PoreWalker	Pellegrini-Calace <i>et al.</i> (2009)	Grid; computation of an initial pore axis from the protein secondary structure (variable diameter from test spheres at various locations), then the axis is relocated and the process is iterated; the algorithm is restricted to some classes of transmembrane channels proteins
McVol	Till and Ullmann (2010)	Grid; filling probe spheres; volumes are Monte-Carlo measured (counting inner points)
POCASA	Yu <i>et al.</i> (2010)	Gridded slices; pockets are sets of free points between a probe sphere and protein surface
VICE	Tripathi and Kellogg (2010)	Grid (with integer arithmetic); pockets are lists of protein atoms; paths are lists of voxels
T-RRT	Jaillet <i>et al.</i> (2010), Cortés <i>et al.</i> (2011)	Grid; Monte-Carlo path planning through voxels
PROPORES	Lee and Helms (2012)	Grid and voxels; hybrid variant of POCKET and SURFNET; shortest paths in voxels with arbitrary cost functions
David	David (1988)	Voronoi tessellation is used to estimate void volumes; these latter are divided in two three size classes; clefts are continuous sets in the class of the intermediate size
FindSurf	Lewis (1989)	Voronoi tessellation altered with dummy tiles; the method is used in complement of David's one; clefts are lists of atoms
CAST, CASTp	Edelsbrunner and Mücke (1992, 1994), Edelsbrunner <i>et al.</i> (1998)	Original alpha shapes, based on deletions in the Delaunay triangulation of the protein atoms with the help a probe sphere of user fixed radius alpha (hence the name 'alpha shape')
APROPOS	Peters <i>et al.</i> (1996)	Alpha shapes, list of atoms in clusters by difference between two alpha shapes
MOLE	Petřek <i>et al.</i> (2007)	Variant of CAVER based on Delaunay triangulation; minimal cost paths channels; the cost function is a variant of the one in CAVER, and is still the inverse of a geometrically interpretable quantity
GeometricPotential	Xie and Bourne (2007)	Delaunay triangulation of $C_{\alpha}$ atoms, removal of tetrahedra based on arbitrary cut-off values
MolAxis	Yaffe <i>et al.</i> (2008a, b)	Channels are alpha shapes based minimal cost pathways; the name comes from the use of a medial axis (here, it is a collection of points having more than one closest point on the van der Waals surface of the protein)
SplitPocket	Tseng and Li (2009), Tseng <i>et al.</i> (2009)	Alpha shape based on weighted Delaunay triangulation; pockets are sets of empty Delaunay triangles with at least one acute triangle
Fpocket	Le Guilloux <i>et al.</i> (2009), Schmidtke <i>et al.</i> (2010)	Pockets are lists of atoms deduced from Delaunay triangulation followed by filling spheres clustering

(continued)

Table 1. Continued

Name	References	Method
UnionBall	Mach and Koehl (2011)	Enhanced version of AlphaVol (part of the original alpha shape package); the calculation of volumes of spheres unions are based on the Voronoi diagram (Edelsbrunner, 1995)
CAVER 3.0	Chovancova <i>et al.</i> (2012)	Variant of CAVER implementing new algorithms for the calculation and clustering of pathways
MOLE2	Berka <i>et al.</i> (2012)	Variant of MOLE; needs a user defined starting point; a vertex of the Voronoi diagram is removed if a sphere with interior threshold radius cannot pass through any of the tetrahedron sides; channels are defined from shortest paths detection
HOLE	Smart <i>et al.</i> (1993, 1996, 1997)	Monte-Carlo simulated annealing exploration of the space available to a sphere; at each step the maximal sphere radius is computed
PASS	Brady and Stouten (2000)	Layers of non-overlapping filling spheres are generated in cavities
SCREEN	Nayal and Honig (2006)	An envelope surface is computed with the help of a 5 Å radius probe sphere, allowing to assign depth values to vertices on the protein surface; the clusters of vertices below a depth threshold define the surfaces of the cavities
PHECOM	Kawabata and Go (2007)	Two probe spheres radii are used; small probes not overlapping large ones define pockets
Binding Response	Zhong and MacKerell (2007)	Overlapping filling spheres are clustered (two steps); clusters define pockets

Grid means 3D rectangular or cubic mesh of nodes. Voxels are 3D analogs of pixels. Alpha shapes are substructures of the Delaunay triangulation

of the shortest paths, but in this case the major part of the network of channels is lost.

At the opposite of a grid of virtual nodes, the Delaunay triangulation of the  $N$  atomic centers is more natural, and efficient algorithms exist, the main one being based on a 4D convex hull (Preparata and Shamos, 1985; Edelsbrunner, 1987). It is why the seminal concept of alpha shape appears in most modern pockets and channels calculation algorithms, as seen in Table 1. This concept was first introduced in the plane by Edelsbrunner *et al.* (1983), and then was generalized to higher dimensional spaces and proposed for molecular applications by Edelsbrunner and Mücke (1992). Alpha shapes are a generalization of the convex hull of the  $N$  points, this latter being the smallest convex polyhedron enclosing the  $N$  points. A spherical probe sphere of radius alpha erases from the polyhedral hull any point at distance less than alpha to the sphere center for all positions where the sphere does not enclose any of the  $N$  points. The resulting object is called an alpha hull. Substituting straight edges for the circular ones and triangles for the spherical caps, the final object is the alpha shape of the  $N$  points set. When alpha tends to zero, the alpha shape reduces to the  $N$  points. When alpha is large enough, the alpha shape is the convex polyhedral hull of the points. For intermediate alpha values, the alpha shape is not convex and can be used as a model of the protein shape. The original variant of the alpha shape concept that we introduce is described in the next section.

## 2 METHODS

### 2.1 Approach

Our approach differs from the existing ones based on alpha shapes variants at three levels: (i) we are interested in the structure of the network of

channels, i.e. the complement of the protein alpha shape to its convex hull. (ii) we considered several geometric models for the ligand, not only spheres and (iii) the solid representing a ligand is used to flag the triangular faces as 'can be passed' or 'cannot be passed', rather than erasing edges, triangles and tetrahedra.

The first level is of crucial importance when it is necessary to identify potential access channels to a buried active site, as in cytochromes P450 (Guengerich, 2005): the shape of the channels must be analyzed, not the shape of the protein. The second level is detailed in Section 2.2, in which it is shown that the spherical model of ligands is most of the time inadequate. The third level can be understood as follows: to travel inside the polyhedral protein hull, the ligand must pass through some sequence of triangular faces, each face separating two adjacent tetrahedra, just as triangular doors can be open or closed in a labyrinth of tetrahedral rooms: see Section 2.3. The final goal is to exhibit the full network of paths through the labyrinth, or from the exterior of the labyrinth to the chamber neighboring the hidden active site.

### 2.2 Ligands shapes analysis

In order to analyze the shapes of the ligands, we report in Table 2 the statistics for 21 geometric parameters on the dataset of 70 substrates of the cytochrome P450 (CYP) 3A4 used by Meslamani *et al.* (2009), containing from 16 to 186 atoms, hydrogens included. The CYP 3A4 substrates are known to offer a wide structural diversity (Rendic, 2002), so it is why we considered that the statistics on these 70 substrates are meaningful.

There are several ways to see in Table 2 that the spherical model of ligands is unrealistic. The first way is to look at the sphericity coefficient  $G$ , which is equal to 1 for a sphere and is less than 1 for other solids. It is the case of our dataset, for which  $G$  is always  $<0.5$ , but  $G$  reflects the shape of the van der Waals solid and thus does not include internal void cavity volumes although the need to compute them was pointed out (Sonavane and Chakrabarti, 2008). Also,  $G$  depends on the atomic radii values and there is no agreement about these values (see, for example, some suggested set of values in Gavezzotti, 1983; Richards, 1985; Scott and

**Table 2.** Statistics on 21 geometric parameters (physical units: Å, Å<sup>2</sup>, Å<sup>3</sup>)

Parameter	Minimum	Maximum	Mean	Standard deviation
$S_W$ (van der Waals surface <sup>a</sup> )	147.4	1324.6	387.8	185.9
$V_W$ (van der Waals volume <sup>a</sup> )	115.5	1162.5	334.5	166.8
$D$ (diameter <sup>b</sup> )	6.9	28.9	13.0	4.1
$R$ (radius <sup>c</sup> )	3.7	14.5	6.6	2.0
$V_S$ <sup>d</sup>	201.3	12673.8	1582.6	1985.7
$V_{ch}$ (convex hull volume)	$1.10^{-2}$	1282.9	216.1	208.8
$H_h$ (half height of the MHEC <sup>e</sup> )	$4.10^{-4}$	5.5	2.3	0.9
$R_h$ (radius of the MHEC <sup>e</sup> )	3.6	14.4	6.5	2.0
$V_H$ (volume of the MHEC <sup>e</sup> )	$4.10^{-2}$	3340.8	732.4	677.1
$H_r$ (half height of the MREC <sup>f</sup> )	2.2	7.2	3.7	1.0
$R_r$ (radius of the MREC <sup>f</sup> )	3.3	14.2	6.2	2.1
$V_R$ (volume of the MREC <sup>f</sup> )	191.0	4829.4	1091.3	1006.4
$G$ (sphericity coefficient <sup>g</sup> )	0.066	0.471	0.246	0.077
$I_v$ (volumic shape coefficient <sup>h</sup> )	$4.10^{-5}$	0.404	0.166	0.080
$I_g$ (geometric shape index <sup>i</sup> )	0.816	1.000	0.972	0.045
$H_h/R$	$8.10^{-4}$	0.649	0.364	0.129
$R_h/R$	0.876	1.000	0.985	0.018
$H_r/R$	0.257	0.830	0.592	0.141
$R_r/R$	0.786	1.000	0.947	0.045
$V_{ch}/V_H$	0.134	0.589	0.304	0.077
$V_{ch}/V_R$	$5.10^{-5}$	0.418	0.206	0.081

All computations were performed with the ASV and RAD1 freewares, available on the same software repository than CCCPP. <sup>a</sup> $S_W$  and  $V_W$  were calculated with the analytical method (Petitjean, 1994, 2013), implemented in ASV. <sup>b</sup>Length of the largest atom pair. <sup>c</sup>Radius of the smallest enclosing sphere;  $2R \geq D \geq 2R\sqrt{6/3}$  (Petitjean, 1992). <sup>d</sup>Volume of the smallest enclosing sphere. <sup>e</sup>MHEC, minimal height enclosing cylinder, calculated with the algorithm of Brandenburg and Theobald (2006), implemented in RAD1. <sup>f</sup>MREC, minimal radius enclosing cylinder, calculated with the algorithm of Petitjean (2012), implemented in RAD1. <sup>g</sup> $G = 36\pi V_W^2/S_W^3$ ;  $G$  takes values in  $[0; 1]$ . <sup>h</sup>Ratio  $V_{ch}/V_S$  of the volume of the convex hull to the volume of its smallest enclosing sphere. <sup>i</sup> $I_g = (D - R)/R$ ;  $I_g \leq 1$  (Petitjean, 1992).

Scheraga, 1966; Zefirov and Zorkii, 1989). In fact these values depend on how they are defined (see Bondi, 1964, for a discussion). Thus we neglect the atomic spheres and we look at other parameters. The geometric shape coefficient  $I_g = (D - R)/R$  (Petitjean, 1992) relates the diameter of the hull to its radius, but it varies few in this context. Assuming that the shape is convex, the best shape model is the convex hull of the atomic centers, i.e. the smallest convex polyhedron containing these centers: by definition, any other enclosing convex solid will contain this polyhedron. The observed distribution of the ratio  $I_v$  of volume of the convex hull to the volume of the smallest enclosing sphere exhibit a mean value  $<0.17$ , which is quite low, thus proving the inadequation of the sphere model. However it is much more difficult to test if a convex polyhedron can pass through pathways without colliding a set of punctual or spherical obstacles rather than testing if a sphere can pass these obstacles without colliding.

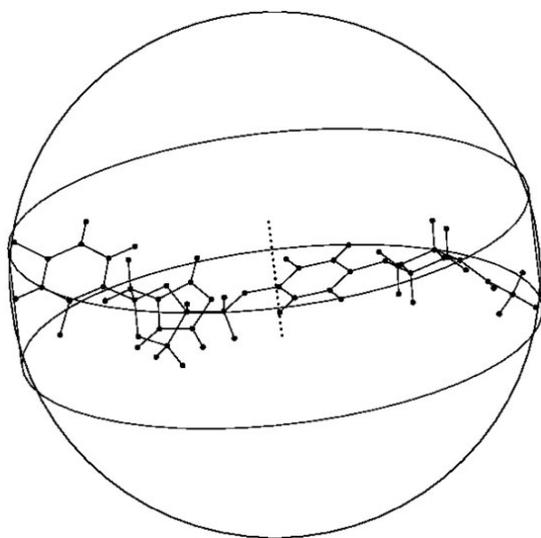
It is why the cylindrical shape was devised for molecules (Petitjean, 2012). A cylinder has a convex shape defined by only two parameters: radius and height, although a sphere has only one, a rectangular 3D box has three ones and a polyhedron has many ones. However cylinders are easier to handle than rectangular boxes for collision tests (see Section 2.3). Since there are two parameters, there is an infinity of ways to combine them to define the minimality criterion for a cylindrical solid (e.g. minimize its volume or its surface, or else), but we retained only two of them: the minimal height cylinder which is relevant for rather flat ligands, and the minimal radius cylinder which is relevant for rather elongated ligands. Intuitively, the former reflects atomic centers distributions for which the smallest eigenvalue of the inertia matrix (i.e.  $N$  times the covariance matrix) is small compared to the two other eigenvalues, although the latter reflects atomic centers distributions for which two eigenvalues are small compared to the largest one. We see in Table 2 that the radius of the

minimal height cylinder and the height of the minimal radius cylinder remain close to 1, although the respective minimized height and minimized radius of these cylinders are significantly  $<1$ . Then the distributions of the hull volume to the cylinder volume offer values significantly  $<1$ , and of the same magnitude than the sphericity coefficient  $G$  and than the volumic shape coefficient  $I_v$ , thus confirming the pertinence of the cylindrical model over the spherical model. It is of major importance to understand that, although a ligand conformer has a unique shape, it does not preclude that this unique shape can be modeled through several ways, such as the minimal height enclosing cylinder, the minimal radius enclosing cylinder, the smallest enclosing sphere (even if that latter encloses a large void volume) and so on. Furthermore, it is recalled that most ligands are flexible, and that is a second reason to insist that several realistic shape models can be simultaneously useful. However, the data of Table 2 show that cylindrical shapes are more realistic than spherical shapes, as already mentioned in that section and illustrated in Figure 1.

### 2.3 Geometric tests of passage

In the basic alpha-shape approach, a tetrahedral cell with its circumscribed sphere of radius alpha can contain a spherical probe ligand of radius alpha (protein atomic spheres are neglected). Assuming that the circumcenter is in the tetrahedra (it is most of the time the case), the spherical ligand cannot exit from the tetrahedra because each radius of the four circles circumscribed to the triangular faces is smaller than alpha. Thus the knowledge of the standard protein alpha shape does not suffice to conclude that the ligand can travel in a channel: a ligand is ensured to travel in a channel if and only if it passes through its associated sequence of triangular faces, and that is a major originality of our approach.

To implement this approach we consider the graph of the Delaunay triangulation, in which the  $N$  nodes of the graph are the tetrahedra and



**Fig. 1.** The ketoconazole, a known CYP 3A4 substrate, modeled by its minimal enclosing sphere ( $R = 10.1 \text{ \AA}$ ) and by its minimal height enclosing cylinder (half height:  $H_h = 2.4 \text{ \AA}$ ;  $R_h = 10.0 \text{ \AA}$ )

the edges of the graph are the triangular faces, i.e. two adjacent tetrahedra are two nodes joined by a non-directed edge. This graph, which underlies the Voronoi diagram structure, is called here the ‘facial graph’ and must not be confused with the one where the nodes are the atoms and the edges are the sides of the tetrahedra. Each node of the facial graph has at most four neighbors. Optionnally, a dummy node representing the exterior of the convex hull can be added to the graph, in which case each of the  $N$  nodes has exactly four neighbors. Then it suffices to perform a loop on the edges (i.e. the triangles) of the facial graph and flag the ones which cannot be passed by the ligand as deleted edges. The resulting network of channels is a subgraph of the facial graph, not necessarily connected.

The most difficult task is to decide whether or not an object (the ligand) can pass through a triangle. It is hard in the case of a non-convex object, and even for convex objects such as a polyhedral hull or a straight box it is not trivial to do. In fact the problem is to define the criterion for passage rather than to compute if yes or no the passage is possible once its criterion is defined. Thus we followed a simplified approach: we considered punctual atoms and then we defined a passage test through a triangle for the two cylindrical models under the assumption that we can neglect collisions with atoms not at the triangle vertices.

For cylinders with a ratio radius/height sufficiently close to zero, the smallest planar projected section of the cylinder is its circular section because its 2D image can be contained in any of its other planar projected sections. The axis of the cylinder is optimally set orthogonal to the plane of the triangle and we are left to a 2D problem. Then, we flag the passage test as being successful in two cases (a) or (b): (a) the triangle is acute and the radius of the circular section (i.e. the one of the cylinder) is smaller than the circumradius of the triangle, or (b) the triangle is obtuse and the circular section of the cylinder does not contain any of the triangle vertices when the center of this circular section is constrained to lie in the triangle. In the case (a), the largest possible cylinder radius is the circumradius, and in case (b) it is the distance from the vertex at the opposite of the largest side to the point at the intersection of the largest side and of the mid-perpendicular of the second largest side. In both cases (a) and (b), the center of the circular section must pass inside the triangle, not outside (the passage of a punctual solid through a 3D triangle is always defined).

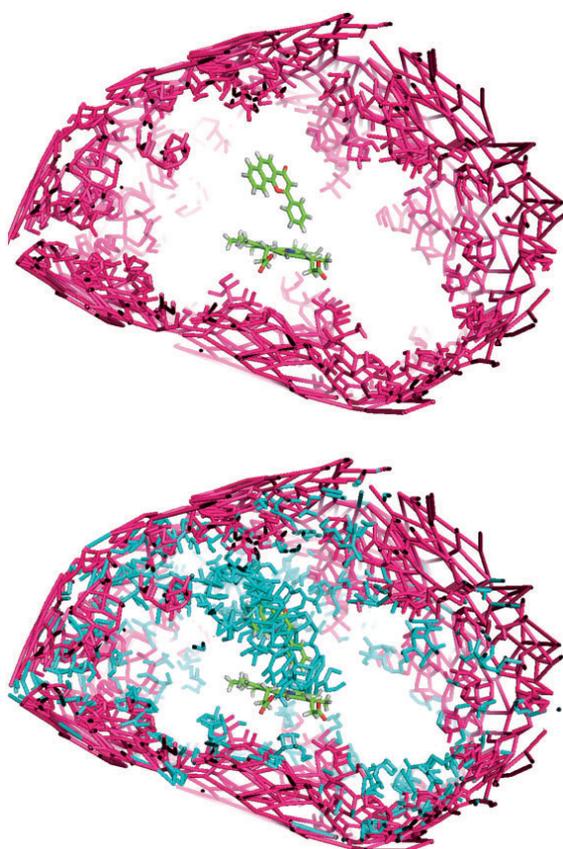
It should be noticed that the tests (a) and (b) are rigorous in the case of the largest section of a spherical ligand, for which the condition of a lack of collision with a fourth vertex can be relaxed.

For cylinders with a ratio height/radius sufficiently close to zero, the axis of the cylinder is optimally set parallel to the plane of the triangle and we are left to a 2D problem. The cylinder height is also the distance between the two closest parallel planes enclosing the ligand. Here, still under the assumption that we can neglect collisions with atoms not at the triangle vertices, we flag the passage test as being successful when the cylinder height is smaller than the largest height of the triangle (or than the second largest side of the triangle when it is obtuse). This passage test can be applied with the same limitations as above to any ligand shape provided that we know the distance between the two closest slabs enclosing the ligand, this latter distance being equal to the height of the smallest height cylinder defined in Section 2.2.

### 3 RESULTS AND DISCUSSION

To exemplify how works CCCPP, we first considered the CYP 1A2, which is known to accept rather flat substrates (Sridhar *et al.*, 2011; Zaretzki *et al.*, 2012). The data were extracted from the crystallized complex of the 1A2 with  $\alpha$ -naphthoflavone, PDB code 2HI4. The flatness, i.e. the thickness of the ligand, is defined as  $2H_h$  (see Table 2). Figure 2 shows that ligands of thickness  $4.5 \text{ \AA}$  and higher can access to surface pockets although no pathway is found to the buried hemic site. Lowering this critical value from  $4.5 \text{ \AA}$  to  $4.45 \text{ \AA}$ , the network of channels available to the ligand appears in front of the distal face of the heme, and CCCPP indicates that the network connects the exterior of the enzyme to the active site. When visualized in superimposition of the channel computed for a  $4.45 \text{ \AA}$  critical value, the flat ligand of the complex ( $0.32 \text{ \AA}$ ) appears indeed in this predicted channel, in front of the distal face of the heme. A remarkable result is that all parts of the central area of the access channel become unavailable quasi simultaneously for any ligand of thickness slightly increasing from  $0.05 \text{ \AA}$  above the critical value of  $4.5 \text{ \AA}$  although the surface pockets are still available for ligands of thickness  $4.5 \text{ \AA}$ . Another result is that the network of channels does not reduce to a small number of chains of nodes: there is NOT a small number of channels, and counting them is meaningless, at the opposite to what seems intuitively suggested by some other channels modeling approaches (see Table 1). CCCPP provides relevant information on this complex network rather than outputting a misleading small number of paths, and the full network structure can be stored both in a text file and in a molecular file to be reused or displayed by any molecular viewer.

The CYP 3A4 is known to accept a much wider diversity of substrates than the 1A2 (Rendic, 2002), some of them being very large. The case of the ketoconazole is of special interest. Its radius, as defined in Table 2, is  $10.1$ , and its thickness is  $4.8 \text{ \AA}$ . A conformational study showed that only a limited number of conformers suffices to handle flexibility (Benkaidali *et al.*, 2012), and the effect on the radius and on the thickness can be neglected to compute the network of channels. A crystallized complex containing two ketoconazoles is available, PDB code 2V0M. There are four chains A, B, C, D in the PDB file. Only surface residues are missing in chains A and D, so that no reconstruction is needed for these chains in our context. The 3D similarity between chains A and D was checked with the SDM algorithm (Petitjean, 1998) after 3D alignment of the two backbones. Only the  $C_\alpha$  of

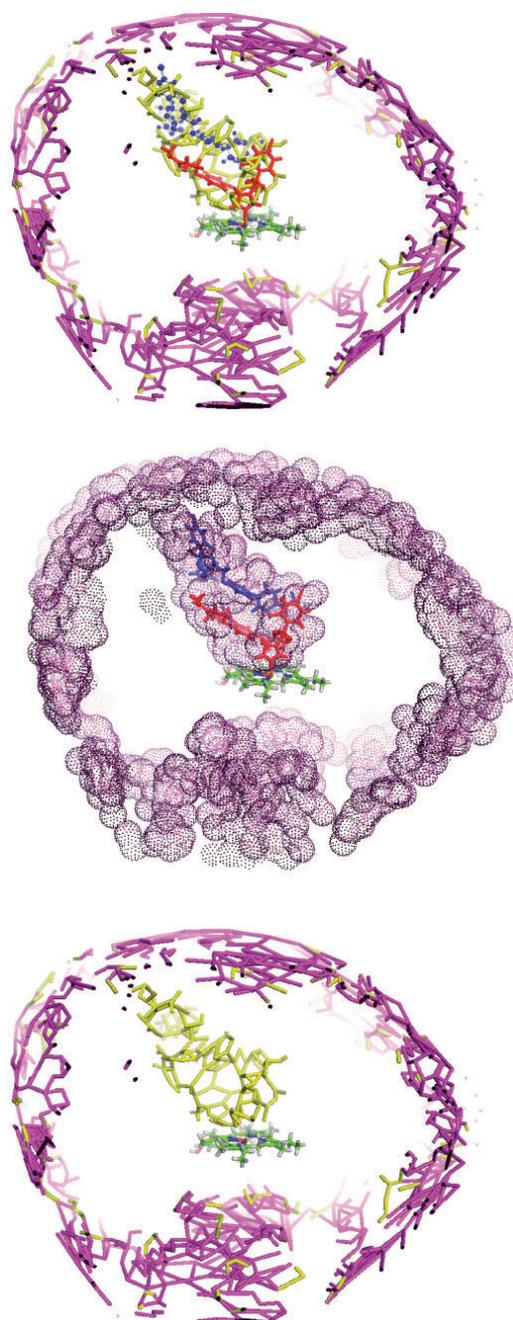


**Fig. 2.** The facial graph of the CYP 1A2 channels network available for ligands of, respectively, 4.50 Å (top) and 4.45 Å (bottom) thickness, visualized with PyMOL. For clarity, the channels are superimposed with the ligand ( $\alpha$ -naphthoflavone) of the CYP 1A2 complex, pdb code 2HI4. Edges indicate the pathways in the network, except for the ligand and for the heme, where edges indicate chemical bonds

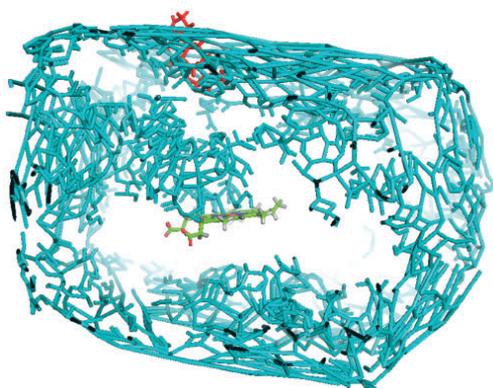
ARG 212 was outside the maximal common 3D motif, thus we neglected this difference and we considered only the chain A. Using a passage test based on a spherical model (it is an option in CCCPP), CCCPP did not find any access to the heme. However CCCPP indeed found an access for the cylindrical model. Figure 3 shows that the two ketoconazoles are indeed in the predicted channels, one in front of the distal face of the heme, and the other one just above in a dead end part of the network.

CCCPP outputs many surface pockets. It is of interest to observe that they can be used. Figure 4 shows such a situation for the complex of the CYP 3A4 with progesterone, PDB code 1W0F.

CCCPP can predict channels from the 3D structure of the enzyme without the help of a complex. We used two 3A4 structures crystallized without ligand, 1TQN and 1W0E. Due to missing residues, we reconstruct the mid of the sequence with HYDRO\_PDB (Azuara *et al.*, 2006). Figure 5 shows the results of CCCPP for two 3A4 structures found in the PDB, codes



**Fig. 3.** The facial graph of the CYP 3A4 channels network available for the ketoconazole, retrieved from PDB data, code 2V0M, visualized with PyMOL with the two ketoconazoles (top). Other display mode with virtual spheres centered on the facial graph nodes (middle). The facial graph without the ketoconazoles (main channel part in yellow) (bottom)



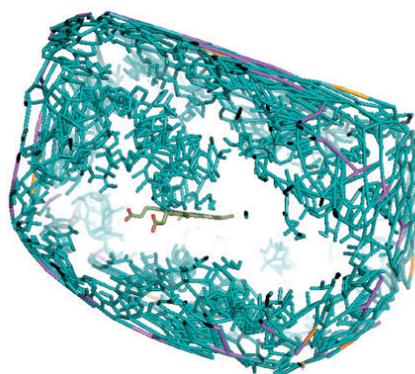
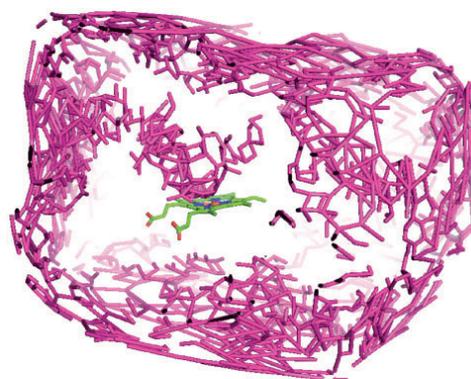
**Fig. 4.** The facial graph of the CYP 3A4 channels network available for the progesterone, retrieved from PDB data, code 1W0F, visualized with PyMOL. The progesterone lies on a pocket surface on the top

1TQN and 1W0E. Both the initial spatial structures and the reconstructed ones slightly differ, but the central part of the channels network is visible in front of the distal face of the heme. The critical thickness above which the access to the heme disappears was computed at 6.25 Å (1TQN) and at 5.25 Å (1W0E). The same values were got without reconstruction.

The critical thickness has been computed for other 3A4 complexes, such as with metyrapone and ritonavir, PDB codes 1W0G and 3NXU, respectively. For 3NXU, there are two chains A and B. For these letters, the SDM algorithm did not found significant differences between the backbones: all the 457  $C_{\alpha}$  were in the common 3D motif. The critical thickness was 5.0 Å for 1W0G and 6.0 Å for 3NXU. All critical thickness values we computed for the 3A4 ranged in the interval 5–6 Å. The differences are interpretable in terms of malleability of the enzyme: some CYPs have been observed to adopt different open and closed conformations when bound to different ligands (Yu *et al.*, 2013).

No objective numerical criterion exists to compare channels computed from different algorithms. That fact is mainly due to the deep differences in the nature of the output produced by the softwares, thus introducing a major difficulty to compare channels. For instance, let us compare the channels of the CYP 3A4 computed by Cojocar *et al.* (2007) with CAVER, cited in Table 1. The main differences are:

- CCCPP produced channels depending on the shape and size of the ligand, i.e. the routes sterically available to a small ligand were more numerous than the routes available to a large ligand, although CAVER produced ‘universal’ channels, i.e. independent of the ligand.
- CCCPP concluded about the sterical ability of a given ligand to pass or not through the channels, although CAVER did not.
- Owing to the examples presented in this section for the CYP 3A4, CCCPP found a complex topology of the network of channels, all of them being in intersection, although three 3A4 channels are mentioned by Cojocar *et al.* (2007).



**Fig. 5.** The facial graph of the CYP 3A4 channels network retrieved from PDB data, codes 1TQN (top) and 1W0E (bottom), visualized with PyMOL

Since these CAVER channels (visually, sequences of intersecting spheres) were described from their closeness (not numerical) to specific loops and helices of the CYP 3A4, it is impossible to compute some degree of similarity they offer with our single network of channels got for a specific ligand. Nevertheless, on the basis of a visual checking, the egress of the CAVER channels were found to be seemingly correct by CCCPP, although the network we found was not indicated by CAVER. CYP 3A4 channels were briefly evoked by Berka *et al.* (2012) as an example of use of a MOLE2, a variant of MOLE, this latter being considered as returning similar results to CAVER (Petřek *et al.*, 2007). The authors of MOLE2 found four channels but did not perform channels comparisons with the three CAVER channels mentioned above, although CAVER and MOLE are algorithms having much in common. Anyway, we insist that the user must restrict channels comparisons to results produced by a common software, or at least produced by softwares based on similar algorithms. CCCPP computes only voids, not binding sites. Numerical comparisons of void calculations need at least that there exist computer representations of the voids as numerical objects. Even though these computerized objects are unambiguously defined, most of the times they have deeply different meanings because they are issued from different methods.

Comparing objects of deeply different nature is not always possible because it needs to define a numerical criterion to compare these objects. It is stressed that the objects differ in nature, not by the values of their measured parameters: there is no list of parameters common to the objects. In the case of CCCPP, the computerized objects are non-connected graphs, in which each node of the graph is associated to a tetrahedra (but not to an atom) of the Delaunay triangulation of the protein. Actually none of the methods cited in Table 1 offers a computer representation compatible with the one of CCCPP (thus CCCPP offers indeed a new insight on the voids), even with those based on Delaunay triangulations (e.g. CASTp, MOLE), thus precluding extensive numerical comparisons with the help of data banks. This is a major difference with the computation of binding sites, because these latter can be checked from experimental values stored in data banks.

The examples given in this section are for illustration of our methodology: a deeper investigation of the CYP450 channels is outside the scope of this article. Reviews on specific members of this enzyme's family are available in Wade *et al.* (2004) and in Cojocaru *et al.* (2007). The computing times were around 2 min on linux and macosx workstations. Nearly a half of this time is devoted to compute the Delaunay triangulation, and most of the rest of the time is used to generate the output molecular file containing the network for display. Once the triangulation done, finding pathways for a ligand from the exterior of the protein to an end point in the center of the enzyme is done in less than 1 s. Also, we point out that CCCPP did not predict binding pockets. Computing pockets can be a preliminary step to the computation of binding pockets, but the latter problem is out of the scope of that paper [see Chen *et al.* (2011) for a recent review].

#### 4 CONCLUSION

We presented a new algorithm to compute cavities, channels and pockets in proteins, and we developed a software called CCCPP which implements this algorithm and outputs the full network of cavities. The void parts are bounded by the convex hull of the atoms of the protein. The structure of the network is stored in a text file and is written in a molecular file to be displayed with any molecular viewer. It was proved that cylindrical ligands shapes are much more realistic than spherical ligands shapes. On the basis of this unusual cylindrical model, CCCPP was successful in explaining how large molecules can be admitted in front of the buried hemic active site of CYP450s. For these latter, CCCPP indicated that there is a complex network of channels rather than a small finite number of channels. It also indicated that the final part of the network in front of the heme lies at its distal face, which is in agreement with physico chemical data. We do not claim that our methodology is superior to the tenths of available other ones. We just point out that our methodology gave rise to an alternative tool providing new information about pockets and channels in proteins. It can be used in a virtual screening context, eventually in addition to energy-based methods.

*Funding:* Programme Doctoral International 'Modélisation des Systèmes Complexes', IRD (Institut de recherche pour le développement) Paris 6, France (to L.B.) and UPMC (Université

Pierre & Marie Curie), Paris 6, France (to L.B.); Integrative BioComputing (IBC company), Rennes, France (to L.B.).

*Conflict of Interest:* none declared.

#### REFERENCES

- Azuara, C. *et al.* (2006) PDB\_Hydro: incorporating dipolar solvents with variable density in the Poisson-Boltzmann treatment of macromolecule electrostatics. *Nucleic Acids Res.*, **34**, W38–W42.
- Benkaidali, L. *et al.* (2012) How well is conformational space covered? In: Putz, M. (ed.) *Chemical Information and Computational Challenges in the 21st Century*. NOVA Science Publishers, New York, pp. 299–313.
- Berka, K. *et al.* (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.*, **40**, W222–W227.
- Bondi, A. (1964) Van der Waals volumes and radii. *J. Phys. Chem.*, **68**, 441–451.
- Brady, G.P. and Stouten, P.F. (2000) Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.*, **14**, 383–401.
- Brandenberg, R. and Theobald, T. (2006) Radii minimal projections of polytopes and constrained optimization of symmetric polynomials. *Adv. Geom.*, **6**, 71–83.
- Chen, K. *et al.* (2011) A critical comparative assessment of predictions of protein binding sites for biologically relevant organic compounds. *Structure*, **19**, 613–621.
- Chovancova, E. *et al.* (2012) CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.*, **8**, e1002708.
- Cojocaru, V. *et al.* (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta*, **1770**, 390–401.
- Coleman, R.G. and Sharp, K.A. (2006) Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J. Mol. Biol.*, **362**, 441–458.
- Cortés, J. *et al.* (2011) Encoding molecular motions in voxel maps. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, **8**, 557–563.
- David, C.W. (1988) Voronoi polyhedra as structure probes in large molecular systems—VII. Channel identification. *Comput. Chem.*, **12**, 207–208.
- Dupuis, F. (2003) Tessellations de Voronoï appliquées aux structures protéiques. PhD Thesis, Université Denis Diderot, Paris 7, France.
- Dupuis, F. *et al.* (2005) Voro3D: 3D Voronoi tessellations applied to protein structures. *Bioinformatics*, **21**, 1715–1716.
- Edelsbrunner, H. (1987) Voronoi diagrams. In: Brauer, W., Rozenberg, G. and Salomaa, A. (eds) *Algorithms in Combinatorial Geometry*. Springer-Verlag, Berlin, pp. 293–334.
- Edelsbrunner, H. (1995) The union of balls and its dual shape. *Discrete Comput. Geom.*, **13**, 415–440.
- Edelsbrunner, H. and Mücke, E.P. (1992) Three-dimensional alpha shapes. In: *Proceedings of the 1992 Boston Workshop on Volume Visualization*. ACM, pp. 75–82.
- Edelsbrunner, H. and Mücke, E.P. (1994) Three-dimensional alpha shapes. *ACM Trans. Graphics*, **13**, 43–72.
- Edelsbrunner, H. *et al.* (1983) On the shape of a set of points in the plane. *IEEE Trans. Info. Theory*, **29**, 551–559.
- Edelsbrunner, H. *et al.* (1998) On the definition and the construction of pockets in macromolecules. *Discrete Appl. Math.*, **88**, 83–102.
- Gavezzotti, A. (1983) The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *J. Am. Chem. Soc.*, **105**, 5220–5225.
- Guengerich, F.P. (2005) Human cytochrome P450 enzymes. In: Ortiz de Montellano, P.R. (ed.) *Cytochrome P450. Structure, Mechanism, and Biochemistry*. 3rd edn. Kluwer/Plenum, New York, pp. 377–530.
- Hendlich, M. *et al.* (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.*, **15**, 359–363.
- Ho, C.M. and Marshall, G.R. (1990) Cavity search: an algorithm for the isolation and display of cavity-like binding regions. *J. Comput. Aided Mol. Des.*, **4**, 337–354.
- Jaillet, L. *et al.* (2010) Sampling-based path planning on configuration-space cost-maps. *IEEE Trans. Robotics*, **26**, 635–646.
- Kalidas, Y. and Chandra, N. (2008) PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.*, **161**, 31–42.
- Kawabata, T. and Go, N. (2007) Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins*, **68**, 516–529.

- Kleywegt, G.J. and Jones, T.A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D*, **50**, 178–185.
- Laskowski, R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, **13**, 323–330.
- Lee, P.H. and Helms, V. (2012) Identifying continuous pores in protein structures with PROPORES by computational repositioning of gating residues. *Proteins*, **80**, 421–432.
- Le Guilloux et al. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, **10**, 168.
- Levitt, D.G. and Banaszak, L.J. (1992) POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graph.*, **10**, 229–234.
- Lewis, R.A. (1989) Determination of clefts in receptor structures. *J. Comput. Aided Mol. Des.*, **3**, 133–147.
- Mach, P. and Koehl, P. (2011) Geometric measures of large biomolecules: surface, volume, and pockets. *J. Comput. Chem.*, **32**, 3023–3038.
- Masuya, M. and Doi, J. (1995) Detection and geometric modeling of molecular surfaces and cavities using digital mathematical morphological operations. *J. Mol. Graph.*, **13**, 331–336.
- Meslamani et al. (2009) Assessing the geometric diversity of cytochrome P450 ligand conformers by hierarchical clustering with a stop criterion. *J. Chem. Inf. Model.*, **49**, 330–337.
- Nayal, M. and Honig, B. (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins*, **63**, 892–906.
- Pellegrini-Calace et al. (2009) PoreWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comput. Biol.*, **5**, e1000440.
- Pérot, S. et al. (2010) Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discovery Today*, **15**, 656–667.
- Peters, K.P. et al. (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.*, **256**, 201–213.
- Petitjean, M. (1992) Applications of the Radius-Diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.*, **32**, 331–337.
- Petitjean, M. (1994) On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects. *J. Comput. Chem.*, **15**, 507–523.
- Petitjean, M. (1998) Interactive maximal common 3D substructure searching with the combined SDM/RMS algorithm. *Comp. Chem.*, **22**, 463–465.
- Petitjean, M. (2012) About the algebraic solutions of smallest enclosing cylinders problems. *Appl. Alg. Eng. Comm. Comp.*, **23**, 151–164.
- Petitjean, M. (2013) Spheres unions and antesections and some of their applications in molecular modeling. In: Mucherino, A. et al. (ed.) *Distance Geometry: Theory, Methods, and Applications*. Springer, New York, pp. 61–83.
- Petrěk, M. et al. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**, 316.
- Petrěk, M. et al. (2007) MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, **15**, 1357–1363.
- Preparata, F.P. and Shamos, M.I. (1985) Convex hulls: basic algorithms. In: *Computational geometry*. Springer-Verlag, Berlin, pp. 95–149.
- Rendic, S. (2002) Summary of information on human CYP enzymes: human P450 metabolism data. *Drug. Metab. Reviews*, **34**, 83–448.
- Richards, F.M. (1985) Calculation of molecular volumes and areas for structures of known geometry. *Meth. Enzymol.*, **115**, 440–464.
- Schmidtke, P. et al. (2010) Fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.*, **38**, W582–W589.
- Scott, R.A. and Scheraga, H.A. (1966) Conformational analysis of macromolecules. III. Helical structures of polyglycine and poly-L-alanine. *J. Chem. Phys.*, **45**, 2091–2101.
- Smart, O.S. et al. (1993) The pore dimensions of gramicidin A. *Biophys J.*, **65**, 2455–2460.
- Smart, O.S. et al. (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics*, **14**, 354–360.
- Smart, O.S. et al. (1997) A novel method for structure-based prediction of ion channel conductance properties. *Biophys J.*, **72**, 1109–1126.
- Sridhar, J. et al. (2011) QSAR models of cytochrome P450 enzyme 1A2 inhibitors using CoMFA, CoMSIA and HQSAR. *SAR QSAR Environ. Res.*, **22**, 681–697.
- Sonavane, S. and Chakrabarti, P. (2008) Cavities and atomic packing in protein structures and interfaces. *PLoS Comput. Biol.*, **4**, e1000188.
- Sperandio, O. et al. (2006) Receptor-based computational screening of compound databases: the main docking-scoring engines. *Curr. Prot. Peptide Sci.*, **7**, 369–393.
- Stahl, M. et al. (2000) Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network. *Protein Eng.*, **13**, 83–88.
- Till, M.S. and Ullmann, G.M. (2010) McVol - A program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm. *J. Mol. Model.*, **16**, 419–429.
- Tripathi, A. and Kellogg, G.E. (2010) A novel and efficient tool for locating and characterizing protein cavities and binding sites. *Proteins*, **78**, 825–842.
- Tseng, Y.Y. and Li, W.-H. (2009) Identification of protein functional surfaces by the concept of a split pocket. *Proteins*, **76**, 959–976.
- Tseng, Y.Y. et al. (2009) SplitPocket: identification of protein functional surfaces and characterization of their spatial patterns. *Nucleic Acids Res.*, **37**, W384–389.
- Venkatachalam, C.M. et al. (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.*, **21**, 289–307.
- Wade, R.C. et al. (2004) A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.*, **98**, 1175–1182.
- Weisel, M. et al. (2007) PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Central J.*, **1**, 7.
- Yaffe, E. et al. (2008a) MolAxis: efficient and accurate identification of channels in macromolecules. *Proteins*, **73**, 72–86.
- Yaffe, E. et al. (2008b) MolAxis: a server for identification of channels in macromolecules. *Nucleic Acids Res.*, **36**, W210–W215.
- Xie, L. and Bourne, P.E. (2007) A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinformatics*, **8**(Suppl. 4), S9.
- Yu, J. et al. (2010) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics*, **26**, 46–52.
- Yu, X. et al. (2013) Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnol. Appl. Biochem.*, **60**, 134–145.
- Zaretski, J. et al. (2012) RS-Predictor models augmented with SMARTCyp reactivities: Robust metabolic regioselectivity predictions for nine CYP isozymes. *J. Chem. Inf. Model.*, **52**, 1637–1659.
- Zefirov, Y.V. and Zorkii, P.M. (1989) Van der Waals radii and their application in chemistry. *Russ. Chem. Rev.*, **58**, 421–440.
- Zhong, S. and MacKerell, A.D. (2007) Binding response: a descriptor for selecting ligand binding site on protein surfaces. *J. Chem. Inf. Model.*, **47**, 2303–2315.

## 5 CONCLUSION ET PERSPECTIVES

### 5.1 Conclusion

Au vu des résultats récents sur les relations structure-fonction du CYP 3A4, il est tout à fait possible de développer des analogues du ritonavir et d'autres types d'inhibiteurs offrant de meilleures propriétés physico-chimiques et pharmacochimiques. La détermination des premières structures cristallines du 3A4 avait montré que la plupart des acides aminés participent au site actif et au canal d'accès du substrat ([Williams *et al.* 2004, Yano *et al.* 2004]). Mais notre étude a permis de proposer un scénario du mouvement de ces acides aminés, et de l'ouverture du bloc 2. Les acides aminés identifiés dans la littérature comme étant impliqués dans le mécanisme d'action du 3A4 ont été retrouvés par CCCPP comme bordant les canaux et jouant un rôle dans l'ouverture/fermeture de ces canaux.

Jusqu'à-là, dans la littérature, on a parlé d'une grande cavité du 3A4 lui conférant la caractéristique d'interaction avec des ligands volumineux et l'implication dans les interactions médicamenteuses. Avec cette étude des canaux, nous ajoutons que le 3A4 peut ouvrir le bloc 2 très élargi (canaux 2f et 2S). Donc l'accueil par le 3A4 de molécules volumineuses ou multiples ne s'explique pas uniquement par la largeur du site actif, mais également par les canaux d'accès. L'ouverture du bloc 2 se fait sans le bloc 1 avec une capacité d'élargissement dépassant le bloc 1, ou bien s'ouvre en plus du plus du bloc 1, intervenant dans la coopérativité ou l'allostérie, chaque bloc accueillant une molécule.

La coopérativité du 3A4 s'expliquerait par l'intervention des canaux du bloc 2. La coopérativité dans le transport à travers les canaux s'effectuerait par fixation d'une molécule effectrice positive favorisant l'ouverture du canal, et inversement, pour un effecteur négatif, le processus d'ouverture sera défavorisé. La coopérativité dans la réactivité serait due à une molécule effectrice favorisant la réactivité avec une autre molécule se situant dans le site actif (exemple : par maintien stérique de la molécule dans le site actif), ou inversement. Il pourrait y avoir ouverture d'autres canaux via la membrane.

Une connaissance mécanistique des complexes d'interaction avec le 3A4 peut aider à développer des médicaments nouveaux ou moins toxiques. Le but ultime de notre étude est de mettre en place un outil de prédiction de la sélectivité des isoenzymes P450 : quelles molécules interagissent (substrat ou inhibiteur), avec quelle isoforme, autrement dit comment se fait leur entrée dans le canal, puis, pour une molécule réagissant avec une isoforme donnée, quelle est sa position dans le site actif, ce qui permettrait de repérer le site d'oxydation de la molécule.

Nous concluons que la sélectivité se fait au niveau des bouches d'entrée des canaux. L'interaction d'une molécule avec les acides aminés bordant les bouches des canaux détermine son entrée. Une fois la molécule entrée, elle peut atteindre le site actif : si la molécule est petite, elle peut pivoter et changer d'orientation au sein du site actif selon son affinité avec les acides aminés du site actif, et si la molécule est volumineuse elle conserve son orientation dans le site actif. Un docking rapide mesurant les interactions des ligands au niveau des bouches des canaux permettrait d'éclaircir cette sélectivité des isoformes envers les ligands. Les structures cristallines ne représentent que la photographie d'un

état conformationnel donné, et ne capturent pas la multitude d'états conformationnels pouvant exister durant le processus de migration du ligand.

## 5.2 Perspectives

Les perspectives de poursuite du projet sont les suivantes :

- Simulation des structures avec la membrane plasmique, pour une vision plus réaliste des ouvertures/fermetures des canaux.
- Docking au niveau des bouches d'entrée des canaux, et docking itératif le long des canaux.
- Etude par des méthodes quantiques de l'interaction des molécules avec les acides aminés de la bouche.
- Etendre l'étude aux autres isoformes
- Prise en compte de la participation des molécules d'eau dans la traversée des ligands : stabilisation, intermédiaire de réactivité (donneur d'H par exemple), etc.

Il est apparu que la principale discrimination des ligands se fait au niveau des bouches des canaux. Afin de prédire la position du ligand dans le site actif et ses sites d'oxydation, il serait intéressant d'étudier sa position au niveau de la bouche (et non au niveau du site actif comme cela s'est fait jusqu'à aujourd'hui), par différentes méthodes de modélisation moléculaire (docking, QSAR, chimie quantique, etc). L'étude de l'interaction des ligands avec les bouches des canaux permettrait de déterminer sa conformation d'entrée, qui est supposée se conserver le long de son trajet dans le canal jusqu'au site actif. Pour les molécules volumineuses, il n'est pas raisonnable de supposer un changement d'orientation de la molécule une fois entrée dans la protéine. La partie du ligand la plus affine avec la bouche du canal est celle qui sera exposée à l'hème, et sera celle qui sera attaquée. Il est à noter que pour une même molécule, plusieurs régions peuvent être affines à des degrés différents, ce qui influencerait sur la probabilité d'accès selon une forme ou une autre. Ceci expliquerait les sites d'oxydations primaires et secondaires (du plus probable aux moins probable). Ces études de modélisation peuvent être complétées par du criblage biochimique. Il faudrait également prendre en compte la flexibilité des espèces en interaction. La prise en compte la flexibilité de la protéine pourrait s'effectuer en élaborant des tests de passage plus sophistiqués du ligand dans chaque triangle généré par la triangulation de Delaunay de la protéine.

Une fois établi géométriquement le réseau de canaux pour une famille donnée de substrats, il faut des critères énergétiques : docking dynamique ou flexible prenant en compte les parois et les paramètres de réactivité éventuels obtenus par des méthodes QM/MM, établissement d'un filtre énergétique rapide prédisant les substrats improbables des CYPs dans un contexte de criblage virtuel. La déformabilité et la flexibilité peuvent être prises en compte, entre autres par diverses simulations de dynamique moléculaire.

La suite du projet comprendrait plusieurs aspects originaux:

- Extension des calculs à diverses chimères des CYPs. En utilisant le concept d'évolution moléculaire dirigée des protéines, utilisation d'outils d'ingénierie des protéines basés sur des approches combinatoires et évolutives permettant d'exprimer des variants d'un CYP particulier. Ces variants sont proches structuralement, mais la palette de substrats reconnus diffère, ce qui en fait un modèle d'étude original pour valider notre approche de modélisation des canaux.

- Simulations de dynamique moléculaire intégrant les informations obtenues sur les canaux, ce qui n'est pas le cas des diverses études de simulations publiées sur le cheminement de substrats (généralement en mode sortie du site actif vers l'extérieur). Dans le cas présent, nous déterminons d'abord les seuls canaux accessibles au substrat considéré (souvent un seul est identifié), et nous disposons d'un point d'entrée compatible.
- Production d'un logiciel intégrant la visualisation des parois des canaux, ce que ne font pas les logiciels classiques tels que ALVIS, CAVER, MOLE.

Les structures disponibles dans la base PDB peuvent comporter des parties manquantes, qui peuvent être reconstruites par des approches robustes de modélisation par homologie. Les prédictions sur les substrats reconnus faites *in-silico* pourront être vérifiées *in-vitro*, en fonction de la disponibilité de l'isoforme recombinante et des substrats.

La flexibilité enzymatique peut être considérée à deux niveaux : les déformations du squelette et la flexibilité des chaînes latérales. Les déformations du squelette peuvent être évaluées en générant un nombre limité de géométries perturbées de conformations du squelette. Les mouvements des chaînes latérales peuvent être considérées via divers modèles, dont le modèle rigide, le modèle rigide après suppression de la chaînes latérales au-delà du carbone  $\beta$ , etc.

Dans une précédente étude, qui visait à reconstruire un cytochrome P450 bactérien, il est indiqué que les blocs structuraux conservés chez les CYPs sont une bonne base pour modéliser les CYPs de structure inconnue, y compris dans le cas d'une faible similarité de séquence avec des structures connues [Jean *et al.* 1997]. Depuis cette étude, de nombreuses structures cristallographiques de CYPs ont été résolues, et la caractérisation des blocs structuraux conservés a été raffinée [Nguyen 2007]. Ce travail est basé sur des outils ayant permis la découverte de sous-structures similaires entre deux structures [Carpentier *et al.* 2005] ou davantage [Pisanti *et al.* 2006]. De tels blocs structuraux peuvent être vus comme des signatures de la famille des CYPs, dans la mesure où les profils de séquences extraits de ces blocs permettent d'extraire des séquences de CYPs dans des bases de séquences génériques comme TrEMBL, avec une bonne sélectivité (tout en se rappelant que les similarités de séquences des CYPs peuvent être faibles, jusqu'à 18% d'identité).

L'analyse en modes normaux et l'analyse en composantes principales des trajectoires de dynamique moléculaire (*essential dynamics*) fournissent une vision des mouvements globaux des protéines et de leur déformabilité. L'énergie de déformation associée aux déplacements des atomes pour les modes normaux de basse fréquence (qui sont les plus pertinents), montre un gradient continu de flexibilité décroissant du centre du site actif vers la surface externe de l'enzyme. Les régions les plus flexibles incluent les SRS (sites de reconnaissance du substrat). L'analyse en *essential dynamics* des trajectoires de simulations de dynamique moléculaire de durées variables (de 2 à 14 ns), suggère que l'enzyme se déplace principalement par diffusion le long d'une large barrière de potentiel. Il a été montré que les différences observées entre structures connues dans diverses conformations (avec ou sans substrats) suivent des mouvements d'ensemble déduits de calculs de modes normaux [Krebs *et al.* 2002]. De plus, les variations structurales observées entre les membres d'une superfamille SCOP semblent suivre les mouvements des modes normaux [Leo-Macias *et al.* 2005]. Il serait utile de confirmer ce comportement pour les CYPs. Nous pensons que les blocs structuraux similaires peuvent être analysés par les modes normaux. L'idée est de décrire la dynamique des CYPs en termes de mouvements relatifs des blocs, et d'inférer l'ouverture potentielle des canaux à partir du glissement de

ces blocs. Donc nous devons définir plus précisément les blocs structuraux en utilisant davantage de structures, y compris des structures déformées obtenues après calculs des modes normaux.

Pour les chaînes latérales, nous proposons d'utiliser la méthode SCit [Tufféry *et al.* 1997], qui utilise un squelette dépendant d'une bibliothèque de rotamères. Cette bibliothèque s'appuie sur l'analyse de la conformation des chaînes latérales basée sur la description du squelette par un alphabet structural utilisant un modèle de Markov caché [Camproux *et al.* 2004].

En cas d'échec de la modélisation de la flexibilité enzymatique (exemple, temps de calculs théoriques), on pourrait se contenter d'un nombre limité de modèles rigides (de même pour le ligand).

On peut distinguer deux causes d'échec dans l'approche *in-silico* proposée : (a) les substrats connus n'accèdent pas au site actif, et (b) seules les molécules volumineuses sont marquées comme incapables d'accéder au site actif. Dans le cas (a), nous proposons de relaxer les contraintes stériques du modèle jusqu'à ce que les substrats connus soient établis comme ligands potentiels. Dans le cas (b), nous proposons d'ajouter des contraintes non stériques, telles que des fonctions de pénalité associées aux interactions électrostatiques.

Finalement, rappelons que nous ne négligeons pas le fait que l'approche strictement géométrique soit insuffisante pour effectuer un criblage virtuel efficace.

## ANNEXE : PRODUCTION SCIENTIFIQUE

Publications dans des revues internationales à comité de lecture :

- Benkaidali, L.; André, F.; Maouche, B.; Siregar, P.; Benyettou, M.; Maurel, F.; Petitjean, M. Computing cavities, channels, pores and pockets in proteins from non spherical ligands models. *Bioinformatics* **2014**, *30*[6], 792–800.  
[doi 10.1093/bioinformatics/btt644](https://doi.org/10.1093/bioinformatics/btt644)

Chapitres d'ouvrage :

- Benkaidali, L.; Mansouri, K.; Tuffery, P.; André, F.; Petitjean, M. How well is conformational space covered? In: *Chemical Information and Computational Challenges in the 21st Century*. Putz, M. (Ed.). Nova Science, New York, **2012**, chap. 13, pp. 299–313.  
[https://www.novapublishers.com/catalog/product\\_info.php?products\\_id=22003](https://www.novapublishers.com/catalog/product_info.php?products_id=22003)

Communications orales **invitées** à des conférences internationales :

- Benkaidali, L. *2-eme Conférence Internationale Energies, Matériaux et environnement*. 2–3 décembre 2012, Université Djilali Bounaama Khemis Miliana, Algérie.  
[http://www.univ-km.dz:8087/index.php?option=com\\_content&view=article&id=115:2eme-conference-internationale-l-energies-materiaux-et-environnement-r-&catid=1:present](http://www.univ-km.dz:8087/index.php?option=com_content&view=article&id=115:2eme-conference-internationale-l-energies-materiaux-et-environnement-r-&catid=1:present)

Communications par poster à des conférences internationales :

- Benkaidali, L.; André, F.; Maouche, B.; Benyettou, M.; Maurel, F.; Petitjean, M. *In-silico prediction of P450 cytochromes substrates through virtual screening and geometric modeling of active site access channels*. JCTC10, 10<sup>èmes</sup> Journées de Chimie Théorique et Computationnelle, USTO Mohamed Boudiaf, 18–19 novembre **2012**, Oran, Algérie.  
<https://sites.google.com/site/jctc10oran2012/>

Communications orales à des conférences nationales :

- Benkaidali, L.; André, F.; Maurel, F.; Petitjean, M. *In-silico prediction of P450 cytochromes substrates through virtual screening and geometric modeling of active site access channels*. Abstract des 6<sup>èmes</sup> journées de la Société Française de Chémoinformatique (SFCi), pp. 31–32. Inria Nancy Grand Est - LORIA, 10–11 octobre **2013**, Nancy.  
(Prix de la meilleure présentation orale décerné par la SFCi à L. Benkaidali)

- Benkaidali, L. *Conférence sur la modélisation Mathématique et Informatique des Systèmes Complexes (COMISCO)*. IRD-UPMC, novembre 2013.  
<http://www.ummisco.ird.fr/pdi/spip.php?rubrique47>

□ Communications par poster à des conférences nationales :

- Benkaidali, L.; André, F.; Maurel, F.; Petitjean, M. *In-silico prediction of P450 cytochromes substrates through virtual screening and geometric modeling of active site access channels*. XVIII<sup>ème</sup> congrès du GGMM (Groupe de Graphisme et Modélisation Moléculaire), 21–23 mai **2013**, Saint-Pierre d'Oléron.  
<https://ggmm2013.wordpress.com/>

□ Séminaires nationaux :

- Benkaidali, L.; Scolari, V. *Rôle de la formation du complexe nucléotidique (CRP-cAMP-ADN) dans la régulation de la fonction de l'ADN: Nouvelle approche d'étude des poches d'interactions combinant modélisation géométrique et docking*. Séminaire IRD-UPMC, octobre 2012.
- Benkaidali, L. *8-èmes journées de modélisation de Paris ENSCP*, mai 2013.  
<http://molsim.info/journees2014/programme2013.pdf>

## BIBLIOGRAPHIE

- [Abraham *et al.* 2002] Abraham, M.H.; Ibrahim, A.; Zissimos, A.M.; Zhao, Y.H.; Comer, J.; Reynolds, D.P. Application of hydrogen bonding calculations in property based drug design. *Drug Discov. Today* **2002**, *7*[20], 1056–1063. doi [10.1016/S1359-6446\(02\)02478-9](https://doi.org/10.1016/S1359-6446(02)02478-9)
- [Ahlström & Zamora 2008] Ahlström, M.M.; Zamora, I. Characterization of type II ligands in CYP2C9 and CYP3A4. *J. Med. Chem.* **2008**, *51*[6], 1755–1763. doi [10.1021/jm701121y](https://doi.org/10.1021/jm701121y)
- [Attias & Petitjean 1993] Attias, R.; Petitjean, M. Statistical analysis of atom topological neighborhoods and multivariate representations of a large chemical file. *J. Chem. Inf. Comput. Sci.* **1993**, *33*[5], 649–656. doi [10.1021/ci00015a001](https://doi.org/10.1021/ci00015a001)
- [Axelrod 1955] Axelrod, J. The enzymatic deamination of amphetamine (Benzedrine). *J. Biol. Chem.* **1955**, *214*[2], 753–763. PMID [14381413](https://pubmed.ncbi.nlm.nih.gov/14381413/)
- [Afzellus *et al.* 2007] Afzellus, L.; Hasselgren Arnby, C.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev.* **2007**, *39*[1], 61–86. doi [10.1080/03602530600969374](https://doi.org/10.1080/03602530600969374)
- [Al-Gailany *et al.* 1978] Al-Gailany, K.A.S.; Houston, J.B.; Bridges, J.W. The role of substrate lipophilicity in determining type 1 microsomal p450 binding characteristics. *Biochem. Pharmacol.* **1978**, *27*[5], 783–788. doi [10.1016/0006-2952\(78\)90521-X](https://doi.org/10.1016/0006-2952(78)90521-X)
- [André *et al.* 2004] André, F.; Delaforge, M.; Loiseau, N. A method for performing restrained dynamics docking of one or multiple substrates on multi-specific enzymes. Patent CEA-CNRS WO/2004/038655. <http://www.google.com/patents/WO2004038655A3>
- [Arroya-Mañez *et al.* 2011] Arroya-Mañez, P.; Bikiel, D.E.; Boechi, L.; Capece, L.; Di Lella, S.; Estrin, D.A.; Marti, M.A.; Moreno, D.M.; Nadra, A.D.; Petruk, A.A. Protein dynamics and ligand migration interplay as studied by computer simulation. *Biochim. Biophys. Acta* **2011**, *1814*[8], 1054–1064. doi [10.1016/j.bbapap.2010.08.005](https://doi.org/10.1016/j.bbapap.2010.08.005)
- [Atkins 2005] Atkins, W.M. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 291–310. doi [10.1146/annurev.pharmtox.45.120403.100004](https://doi.org/10.1146/annurev.pharmtox.45.120403.100004)
- [Atkins *et al.* 2001] Atkins, W.M.; Wang, R.W.; Lu, A.Y. Allosteric behavior in cytochrome p450-dependent in vitro drug-drug interactions: a prospective based on conformational dynamics. *Chem. Res. Toxicol.* **2001**, *14*[4], 338–347. doi [10.1021/tx0002132](https://doi.org/10.1021/tx0002132)
- [Austin *et al.* 2002] Austin, R.P.; Barton, P.; Cockroft, S.L.; Wenlock, M.C.; Riley, R.J. The influence of nonspecific microsomal binding on apparent intrinsic clearance, and its prediction from physicochemical properties. *Drug Metab. Dispos.* **2002**, *30*[12], 1497–1503. doi [10.1124/dmd.30.12.1497](https://doi.org/10.1124/dmd.30.12.1497)

- [Baas *et al.* 2004] Baas, B.J.; Denisov, I.G.; Sligar, S.G. Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment. *Arch. Biochem. Biophys.* **2004**, *430*[2], 218–228. doi [10.1016/j.abb.2004.07.003](https://doi.org/10.1016/j.abb.2004.07.003)
- [Backes *et al.* 1998] Backes, W.L.; Batie, C.J.; Cawley, G.F. Interactions among P450 enzymes when combined in reconstituted systems: formation of a 2B4-1A2 complex with a high affinity for NADPH-cytochrome P450 reductase. *Biochemistry* **1998**, *37*[37], 12852–12859. doi [10.1021/bi980674a](https://doi.org/10.1021/bi980674a)
- [Baylon *et al.* 2013] Baylon, J.L.; Lenov, I.L.; Sligar, S.G.; Tajkhorshid, E. Characterizing the membrane-bound state of cytochrome P450 3A4: Structure, depth of insertion, and orientation. *J. Am. Chem. Soc.* **2013**, *135*[23], 8542–8551. doi [10.1021/ja4003525](https://doi.org/10.1021/ja4003525)
- [Benkaidali *et al.* 2014] Benkaidali, L.; André, F.; Maouche, B.; Siregar, P.; Benyettou, M.; Maurel, F.; Petitjean, M. Computing cavities, channels, pores and pockets in proteins from non spherical ligands models. *Bioinformatics* **2014**, *30*[6], 792–800. doi [10.1093/bioinformatics/btt644](https://doi.org/10.1093/bioinformatics/btt644)
- [Benkaidali *et al.* 2012] Benkaidali, L.; Mansouri, K.; Tuffery, P.; André, F.; Petitjean, M. How well is conformational space covered? In: *Chemical Information and Computational Challenges in the 21st Century*. Putz, M. (Ed.). Nova Science, New York, **2012**, chap. 13, pp. 299–313. [https://www.novapublishers.com/catalog/product\\_info.php?products\\_id=22003](https://www.novapublishers.com/catalog/product_info.php?products_id=22003)
- [Berka *et al.* 2012] Berka, K.; Hanák, O.; Sehnal, D.; Banáš, P.; Navrátilová, V.; Jaiswal, D.; Ionescu, C.M.; Vařeková, R.S.; Koča, J.; Otyepka, M. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.* **2012**, *40*[W1], W222–W227. doi [10.1093/nar/gks363](https://doi.org/10.1093/nar/gks363)
- [Berka *et al.* 2011] Berka, K.; Hendrychová, T.; Anzenbacher, P.; Otyepka, M. Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J. Phys. Chem. A* **2011**, *115*[41], 11248–11255. doi [10.1021/jp204488j](https://doi.org/10.1021/jp204488j)
- [Berka *et al.* 2013] Berka, K.; Paloncýová, M.; Anzenbacher, P.; Otyepka, M. Behavior of human cytochromes P450 on lipid membranes. *J. Phys. Chem. B* **2013**, *117*[39], 11556–11564. doi [10.1021/jp4059559](https://doi.org/10.1021/jp4059559)
- [Black 1992] Black, S.D. Membrane topology of the mammalian P450 cytochromes. *FASEB J.* **1992**, *6*[2], 680–685. PMID [1537456](https://pubmed.ncbi.nlm.nih.gov/1537456/)
- [Black *et al.* 1994] Black, S.D.; Martin, S.T.; Smith, C.A. Membrane topology of liver microsomal cytochrome P450 2B4 determined via monoclonal antibodies directed to the halt-transfer signal. *Biochemistry* **1994**, *33*[22], 6945–6951. doi [10.1021/bi00188a025](https://doi.org/10.1021/bi00188a025)
- [Boda *et al.* 2014] Boda, D.; Kovács, R.; Gillespie, D.; Kristóf, T. Selective transport through a model calcium channel studied by local equilibrium Monte Carlo simulations coupled to the Nernst–Planck equation. *J. Mol. Liq.* **2014**, *189*, 100–112. doi [10.1016/j.molliq.2013.03.015](https://doi.org/10.1016/j.molliq.2013.03.015)
- [Bohme *et al.* 2007] Bohme Leite, T.; Gomes, D.; Miteva, M.A.; Chomilier, J.; Villoutreix, B.O.; Tufféry, P. Frog: a FFree Online druG 3D Conformation Generator. *Nucleic Acids Res.* **2007**, *35*[suppl. 2], W568–W572. doi [10.1093/nar/gkm289](https://doi.org/10.1093/nar/gkm289)

- [Bouvier *et al.* 2008] Bouvier, D.; Hamon, V.; Perrin, L.; André, F. Malleability of cytochrome P450 3A4 from a theoretical perspective. *Eur. Biophys. J.* **2007**, *36*[suppl. 1], S51–S248. doi [10.1007/s00249-007-0178-7](https://doi.org/10.1007/s00249-007-0178-7) (abstract P-204 (p.S104) of the 6<sup>th</sup> European Biophysics Congress, London, UK, July 2007).
- [Bränden *et al.* 2014] Bränden, G.; Sjögren, T.; Schnecke, V.; Xue, Y. Structure-based ligand design to overcome CYP inhibition in drug discovery projects. *Drug Discov. Today* **2014**, *19*[7], 905–911. doi [10.1016/j.drudis.2014.03.012](https://doi.org/10.1016/j.drudis.2014.03.012)
- [Bren & Oostenbrink 2012] Bren, U.; Oostenbrink, C. Cytochrome P450 3A4 inhibition by ketoconazole: Tackling the problem of ligand cooperativity using molecular dynamics simulations and free-energy calculations. *J. Chem. Inf. Model.* **2012**, *52*[6], 1573–1582. doi [10.1021/ci300118x](https://doi.org/10.1021/ci300118x)
- [Bridges *et al.* 1998] Bridges, A.; Gruenke, L.; Chang, Y.T.; Vakser, I.A.; Loew, G.; Waskell, L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b<sub>5</sub> and cytochrome P450 reductase. *J. Biol. Chem.* **1998**, *273*[27], 17036–17049. doi [10.1074/jbc.273.27.17036](https://doi.org/10.1074/jbc.273.27.17036)
- [Brown *et al.* 2008] Brown, C.M.; Reisfeld, B.; Mayeno, A.M. Cytochromes P450: a structure-based summary of biotransformations using representative substrates. *Drug Metab. Rev.* **2008**, *40*[1], 1–100. doi [10.1080/03602530802309742](https://doi.org/10.1080/03602530802309742)
- [Buening *et al.* 1978] Buening, M.K.; Fortner, J.G.; Kappas, A.; Conney, A.H. 7,8-Benzoflavone stimulates the metabolic activation of aflatoxin B<sub>1</sub> to mutagens by human liver. *Biochem. Biophys. Res. Comm.* **1978**, *82*[1], 348–355. doi [10.1016/0006-291X\(78\)90616-2](https://doi.org/10.1016/0006-291X(78)90616-2)
- [Caldwell 2000] Caldwell, G.W. Compound optimization in early- and late-phase drug discovery: acceptable pharmacokinetic properties utilizing combined physicochemical, in vitro and in vivo screens. *Curr. Opin. Drug Discov. Devel.* **2000**, *3*[1], 30–41. PMID [19649835](https://pubmed.ncbi.nlm.nih.gov/19649835/)
- [Cameron *et al.* 2005] Cameron, M.D.; Wen, B.; Allen, K.E.; Roberts, A.G.; Schuman, J.T.; Campbell, A.P.; Kunze, K.L.; Nelson, S.D. Cooperative binding of midazolam with testosterone and alpha-naphthoflavone within the CYP3A4 active site: a NMR T1 paramagnetic relaxation study. *Biochemistry* **2005**, *44* [43], 14143–14151. doi [10.1021/bi051689t](https://doi.org/10.1021/bi051689t)
- [Camproux *et al.* 2004] Camproux, A.C.; Gautier, R.; Tufféry, P. A hidden markov model derived structural alphabet for proteins. *J. Mol. Biol.* **2004**, *339*[3], 591–605. doi [10.1016/j.jmb.2004.04.005](https://doi.org/10.1016/j.jmb.2004.04.005)
- [Carpentier *et al.* 2005] Carpentier, M.; Brouillet, S.; Pothier, J. YAKUSA: a fast structural databases scanning method. *Prot. Struct. Funct. Bioinf.* **2005**, *61*[1], 137–151. doi [10.1002/prot.20517](https://doi.org/10.1002/prot.20517)
- [Chovancova *et al.* 2012] Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Dambosky, J. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **2012**, *8*[10], e1002708. doi [10.1371/journal.pcbi.100270](https://doi.org/10.1371/journal.pcbi.100270)
- [Clark 2001] Clark, D.E. Prediction of intestinal absorption and blood-brain barrier penetration by computational methods. *Comb. Chem. High Throughput Screen.* **2001**, *4*[6], 477–496. doi [10.2174/1386207013330850](https://doi.org/10.2174/1386207013330850)

- [Clark & Pickett 2000] Clark, D.E.; Pickett, S.D. Computational methods for the prediction of 'drug-likeness'. *Drug Discov. Today* **2000**, *5*[2], 49–58. doi [10.1016/S1359-6446\(99\)01451-8](https://doi.org/10.1016/S1359-6446(99)01451-8)
- [Claude 1943] Claude, A. The constitution of protoplasm. *Science* **1943**, *97*[2525], 451–456. doi [10.1126/science.97.2525.451](https://doi.org/10.1126/science.97.2525.451)
- [Cojocaru *et al.* 2011] Cojocaru, V.; Balali-Mood, K.; Sansom, M.S.P.; Wade, R.C. Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLoS Comput. Biol.* **2011**, *7*[8], e1002152. doi [10.1371/journal.pcbi.1002152](https://doi.org/10.1371/journal.pcbi.1002152)
- [Cojocaru *et al.* 2007] Cojocaru, V.; Winn, P.J.; Wade, R.C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* **2007**, *1770*[3], 390–401. doi [10.1016/j.bbagen.2006.07.005](https://doi.org/10.1016/j.bbagen.2006.07.005)
- [Cojocaru *et al.* 2012] Cojocaru, V.; Winn, P.J.; Wade, R.C. Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr. Drug Metab.* **2012**, *13*[2], 143–154. doi [10.2174/138920012798918462](https://doi.org/10.2174/138920012798918462)
- [Coon 2005] Coon, M.J. Cytochrome P450: Nature's most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 1–25. doi [10.1146/annurev.pharmtox.45.120403.100030](https://doi.org/10.1146/annurev.pharmtox.45.120403.100030)
- [Cortès *et al.* 2005] Cortès, J.; Siméon, T.; Ruiz de Angulo, V.; Guieysse, D.; Remaud-Siméon, M.; Tran, V. A path planning approach for computing large amplitude motions of flexible molecules. *Bioinformatics* **2005** *21*[suppl. 1], i116–i125. doi [10.1093/bioinformatics/bti1017](https://doi.org/10.1093/bioinformatics/bti1017)
- [Cosme & Johnson 2000] Cosme, J.; Johnson, E.F. Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme. Mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding. *J. Biol. Chem.* **2000**, *275*[4], 2545–2553. doi [10.1074/jbc.275.4.2545](https://doi.org/10.1074/jbc.275.4.2545)
- [Cronin *et al.* 2001] Cronin, M.T.D.; Manga, N.; Seward, J.R.; Sinks, G.D.; Schultz, T.W. Parametrization of electrophilicity for the prediction of the toxicity of aromatic compounds. *Chem. Res. Toxicol.* **2001**, *14*[11], 1498–1505. doi [10.1021/tx015502k](https://doi.org/10.1021/tx015502k)
- [Cupp-Vickery *et al.* 2001] Cupp-Vickery, J.R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. *J. Mol. Biol.* **2001**, *311*[1], 101–110. doi [10.1006/jmbi.2001.4803](https://doi.org/10.1006/jmbi.2001.4803)
- [Cupp-Vickery & Poulos 1995] Cupp-Vickery, J.R.; Poulos, T.L. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Struct. Biol.* **1995**, *2*[2], 144–153. doi [10.1038/nsb0295-144](https://doi.org/10.1038/nsb0295-144)
- [Davis *et al.* 2009] Davis, J.L.; Papich, M.G.; Heit, M.C. Antifungal and antiviral drugs. In: *Veterinary Pharmacology and Therapeutics*. Riviere, J.E.; Papich, M.G. (Eds.). 9<sup>th</sup> ed., Wiley-Blackwell, New-York, **2009**, sect. 9, chap. 39. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0813820618.html>
- [Davydov *et al.* 2003] Davydov, D.R.; Halpert, J.R.; Renaud, J.P.; Hui Bon Hoa, G. Conformational heterogeneity of cytochrome P450 3A4 revealed by high pressure spectroscopy. *Biochem. Biophys. Res. Commun.* **2003**, *312*[1], 121–130. doi [10.1016/j.bbrc.2003.09.247](https://doi.org/10.1016/j.bbrc.2003.09.247)

- [de Groot & Elkins 2002] de Groot, M.J.; Ekins, S. Pharmacophore modeling of cytochromes P450. *Adv. Drug Deliv. Rev.* **2002**, *54*[3], 367–383. doi [10.1016/S0169-409X\(02\)00009-1](https://doi.org/10.1016/S0169-409X(02)00009-1)
- [Delaunay 1934] Delaunay, B. Sur la sphère vide. A la mémoire de Georges Voronoï. *Bull. Acad. Sci. URSS, Série VII* **1934**, *6*, 793–800. <http://mi.mathnet.ru/eng/izv4937>
- [Denisov *et al.* 2007] Denisov, I.G.; Baas, B.J.; Grinkova, Y.V.; Sligar, S.G. Cooperativity in cytochrome P450 3A4: linkages in substrate binding, spin state, uncoupling, and product formation. *J. Biol. Chem.* **2007**, *282*[10], 7066–7076. doi [10.1074/jbc.M609589200](https://doi.org/10.1074/jbc.M609589200)
- [Denisov *et al.* 2009] Denisov, I.G.; Frank, D.J.; Sligar, S.G. Cooperative properties of cytochromes P450. *Pharmacol. Ther.* **2009**, *124*[2], 151–167. doi [10.1016/j.pharmthera.2009.05.011](https://doi.org/10.1016/j.pharmthera.2009.05.011)
- [Denisov *et al.* 2012] Denisov, I.G.; Shih, A.Y.; Sligar, S.G. Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.* **2012**, *108*, 150–158. doi [10.1016/j.jinorgbio.2011.11.026](https://doi.org/10.1016/j.jinorgbio.2011.11.026)
- [Dijkstra 1959] Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*[1], 269–271. doi [10.1007/BF01386390](https://doi.org/10.1007/BF01386390)
- [Domanski *et al.* 2001] Domanski, T.L.; He, Y.A.; Khan, K.K.; Roussel, F.; Wang, Q.; Halpert, J.R. Phenylalanine and tryptophan scanning mutagenesis of CYP3A4 substrate recognition site residues and effect on substrate oxidation and cooperativity. *Biochemistry* **2001**, *40*[34], 10150–10160. doi [10.1021/bi010758a](https://doi.org/10.1021/bi010758a)
- [Dubois *et al.* 1987] Dubois, J.E.; Panaye, A.; Attias, R. DARC system: notions of defined and generic substructures. Filiation and coding of FREL substructure (SS) classes. *J. Chem. Inf. Comput. Sci.* **1987**, *27*[2], 74–82. doi [10.1021/ci00054a007](https://doi.org/10.1021/ci00054a007)
- [Durrant *et al.* 2011] Durrant, J.D.; de Oliveira, C.A.F.; McCammon, J.A. POVME: an algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* **2011**, *29*[5], 773–776. doi [10.1016/j.jmgm.2010.10.007](https://doi.org/10.1016/j.jmgm.2010.10.007)
- [Edelsbrunner 1987] Edelsbrunner, H. Voronoi diagrams. *Algorithms in combinatorial geometry*, Springer-Verlag, Berlin, **1987**, chap. 13, pp. 293–334. doi [10.1007/978-3-642-61568-9](https://doi.org/10.1007/978-3-642-61568-9)
- [Edelsbrunner & Mücke 1992] Edelsbrunner, H.; Mücke, E.P. Three-dimensional alpha shapes. *Proceedings of the 1992 Boston workshop on Volume Visualization*, VolVis92, Boston, MA. ACM, New York, **1992**, pp. 75–82. doi [10.1145/147130.147153](https://doi.org/10.1145/147130.147153)
- [Edelsbrunner & Mücke 1994] Edelsbrunner, H.; Mücke, E.P. Three-dimensional alpha shapes. *ACM Trans. Graphics*, **1994**, *13*[1], 43–72. doi [10.1145/174462.156635](https://doi.org/10.1145/174462.156635)
- [Ekroos & Sjögren 2006] Ekroos, M.; Sjögren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. USA* **2006**, *103*[37], 13682–13687. doi [10.1073/pnas.0603236103](https://doi.org/10.1073/pnas.0603236103)
- [Elkins *et al.* 2001] Ekins, S.; de Groot, M.J.; Jones, J.P. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. *Drug Metab. Dispos.* **2001**, *29*[7], 936–944. PMID [11408357](https://pubmed.ncbi.nlm.nih.gov/11408357/)

- [Estabrook 2003] Estabrook, R.W. A passion for P450s (remembrances of the early history of research on cytochrome P450). *Drug Metab. Dispos.* **2003**, *31*[12], 1461–1473. doi [10.1124/dmd.31.12.1461](https://doi.org/10.1124/dmd.31.12.1461)
- [Estabrook 2005] Estabrook, R.W. Steroid hydroxylations: A paradigm for cytochrome P450 catalyzed mammalian monooxygenation reactions. *Biochem. Biophys. Res. Commun.* **2005**, *338*[1], 290–298. doi [10.1016/j.bbrc.2005.08.168](https://doi.org/10.1016/j.bbrc.2005.08.168)
- [Fa *et al.* 2015] Fa, B.; Cong, S.; Wang, J. Pi-pi stacking mediated cooperative mechanism for human cytochrome P450 3A4. *Molecules.* **2015**, *20*[5], 7558–7573. doi [10.3390/molecules20057558](https://doi.org/10.3390/molecules20057558)
- [Farooq & Roberts 2010] Farooq, Y.; Roberts, G.C. Kinetics of electron transfer between NADPH-cytochrome P450 reductase and cytochrome P450 3A4. *Biochem. J.* **2010**, *432*[3], 485–493. doi [10.1042/BJ20100744](https://doi.org/10.1042/BJ20100744)
- [Fernando *et al.* 2006] Fernando, H.; Halpert, J.R.; Davydov, D.R. Resolution of multiple substrate binding sites in cytochrome P450 3A4: the stoichiometry of the enzyme-substrate complexes probed by FRET and Job's titration. *Biochemistry* **2006**, *45*[13], 4199–4209. doi [10.1021/bi052491b](https://doi.org/10.1021/bi052491b)
- [Ferrari *et al.* 1991] Ferrari, S.; Leemann, T.; Dayer, P. The role of lipophilicity in the inhibition of polymorphic cytochrome P450IID6 oxidation by beta-blocking agents in vitro. *Life Sci.* **1991**, *48*[23], 2259–2265. doi [10.1016/0024-3205\(91\)90341-8](https://doi.org/10.1016/0024-3205(91)90341-8)
- [Fishelovitch *et al.* 2007] Fishelovitch, D.; Hazan, C.; Shaik, S.; Wolfson, H.J.; Nussinov, R. Structural dynamics of the cooperative binding of organic molecules in the human cytochrome P450 3A4. *J. Am. Chem. Soc.* **2007**, *129*[6], 1602–1611. doi [10.1021/ja066007j](https://doi.org/10.1021/ja066007j)
- [Fishelovitch *et al.* 2009] Fishelovitch, D.; Shaik, S.; Wolfson, H.J.; Nussinov, R. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: Involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B* **2009**, *113*[39], 13018–13025. doi [10.1021/jp810386z](https://doi.org/10.1021/jp810386z)
- [Fishelovitch *et al.* 2010] Fishelovitch, D.; Shaik, S.; Wolfson, H.J.; Nussinov, R. How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? *J. Phys. Chem. B* **2010**, *114*[17], 5964–5970. doi [10.1021/jp101894k](https://doi.org/10.1021/jp101894k)
- [Furnham *et al.* 2014] Furnham, N.; Holliday, G.L.; de Beer, T.A.; Jacobsen, J.O.; Pearson, W.R.; Thornton, J.M. The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* **2014**, *42*[D1], D485–D489. doi [10.1093/nar/gkt1243](https://doi.org/10.1093/nar/gkt1243)
- [Galetin *et al.* 2003] Galetin, A.; Clarke, S.E.; Houston, J.B. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab. Dispos.* **2003**, *31*[9], 1108–1116. doi [10.1124/dmd.31.9.1108](https://doi.org/10.1124/dmd.31.9.1108)
- [Gao & Hansch 1996] Gao, H.; Hansch, C. QSAR of P450 oxidation: on the value of comparing  $K_{cat}$  and  $K_m$  with  $K_{cat}/K_m$ . *Drug Metab. Rev.* **1996**, *28*[4], 513–526. doi [10.3109/03602539608994017](https://doi.org/10.3109/03602539608994017)
- [Garfinkel 1958] Garfinkel, D. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Arch. Biochem. Biophys.* **1958**, *77*[2], 493–509. PMID [13584011](https://pubmed.ncbi.nlm.nih.gov/13584011/)

- [Goel *et al.* 2008] Goel, S; Çömezoglu, S.N.; Cohen, M.; Perrin, L.; André, F.; Jayabalan, D.; Iacono, L.; Comprelli, A.; Ly, V.T.; Zhang, D.; Xu, C.; Humphreys, W.G.; McDaid, H.; Goldberg, G.; Horwitz, S.B.; Mani, S. The effect of ketoconazole on the pharmacokinetics and pharmacodynamics of ixabepilone, a first in class epothilone B analog in late phase clinical development. *Clin. Cancer Res.* **2008**, *14*[9], 2701–2709. doi [10.1158/1078-0432.CCR-07-4151](https://doi.org/10.1158/1078-0432.CCR-07-4151)
- [Gorski *et al.* 1994] Gorski, J.C.; Hall, S.D.; Jones, D.R.; VandenBranden, M.; Wrighton, S.A. Regioselective biotransformation of midazolam by members of the human cytochrome P450 3A (CYP3A) subfamily. *Biochem. Pharmacol.* **1994**, *47*[9], 1643–1653. doi [10.1016/0006-2952\(94\)90543-6](https://doi.org/10.1016/0006-2952(94)90543-6)
- [Gotoh 1992] Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **1992**, *267*[1], 83–90. PMID [1730627](https://pubmed.ncbi.nlm.nih.gov/1730627/)
- [Graham & Peterson 1999] Graham, S.E.; Peterson, J.A. How similar are P450s and what can their differences teach us? *Arch. Biochem. Biophys.* **1999**, *369*[1], 24–29. doi [10.1006/abbi.1999.1350](https://doi.org/10.1006/abbi.1999.1350)
- [Graham-Lorence *et al.* 1995] Graham-Lorence, S.; Amarneh, B.; White, R.E.; Peterson, J.A.; Simpson, E.R. A three-dimensional model of aromatase cytochrome P450. *Protein Sci.* **1995**, *4*[6], 1065–1080. doi [10.1002/pro.5560040605](https://doi.org/10.1002/pro.5560040605)
- [Graham-Lorence & Peterson 1995] Graham-Lorence, S.E.; Peterson, J.A. The molecular structure of P450S: The conserved and the variable elements. *Adv. Mol. Cell Biol.* **1995**, *14*, 57–79. doi [10.1016/S1569-2558\(08\)60340-9](https://doi.org/10.1016/S1569-2558(08)60340-9)
- [Graham-Lorence *et al.* 1997] Graham-Lorence, S.; Truan, G.; Peterson, J.A.; Falck, J.R.; Wei, S.; Helvig, C.; Capdevila, J.H. An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14*S*,15*R*)-arachidonic acid epoxygenase. *J. Biol. Chem.* **1997**, *272*[2], 1127–1135. doi [10.1074/jbc.272.2.1127](https://doi.org/10.1074/jbc.272.2.1127)
- [Guengerich 1995] Guengerich, F.P. Cytochromes P450 of human liver. Classification and activity profiles of the major enzymes. In: *Advances in drug metabolism in man*. Pacifici, G.M.; Fracchia, G.N. (Eds). European Commission, Luxembourg, **1995** chap. 7, section II, pp. 179–231. <http://bookshop.europa.eu/en/advances-in-drug-metabolism-in-man-pbCDNA15439/?CatalogCategoryID=X74KABstxngAAAEjdpEY4e5L>
- [Guengerich 1999] Guengerich, F.P. Cytochrome P-450 3A4: regulation and role in drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1–17. doi [10.1146/annurev.pharmtox.39.1.1](https://doi.org/10.1146/annurev.pharmtox.39.1.1)
- [Guengerich 2003] Guengerich, F.P. Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions. *Arch. Biochem. Biophys.* **2003**, *409*[1], 59–71. doi [10.1016/S0003-9861\(02\)00415-0](https://doi.org/10.1016/S0003-9861(02)00415-0)
- [Guengerich 2006] Guengerich, F.P. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J.* **2006**, *8*[1], E101–E111. doi [10.1208/aapsj080112](https://doi.org/10.1208/aapsj080112)
- [Guengerich 2013] Guengerich, F.P. New trends in cytochrome P450 research at the half-century mark. *J. Biol. Chem.* **2013**, *288*[24], 17063–17064. doi [10.1074/jbc.R113.466821](https://doi.org/10.1074/jbc.R113.466821)

- [Guengerich *et al.* 2004] Guengerich, F.P.; Krauser, J.A.; Johnson, W.W. Rate-limiting steps in oxidations catalyzed by rabbit cytochrome P450 1A2. *Biochemistry* **2004**, *43*[33], 10775–10788. doi [10.1021/bi0491393](https://doi.org/10.1021/bi0491393)
- [Guengerich & MacDonald 1990] Guengerich, F.P.; MacDonald, T.L. Mechanisms of cytochrome P-450 catalysis. *FASEB J.* **1990**, *4*[8], 2453–2459. PMID [2185971](https://pubmed.ncbi.nlm.nih.gov/2185971/)
- [Gund 1977] Gund, P. Three-dimensional pharmacophoric pattern searching. In: *Progress in Molecular and Subcellular Biology*, vol. 5, Hahn, F.E.; Kersten, H.; Kersten, W.; Szybalski, W. (Eds.), **1977**, pp. 117–143, Springer-Verlag. doi [10.1007/978-3-642-66626-1](https://doi.org/10.1007/978-3-642-66626-1)
- [Haines *et al.* 2001] Haines, D.C.; Tomchick, D.R.; Machius, M.; Peterson, J.A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001**, *40*[45], 13456–13465. doi [10.1021/bi011197q](https://doi.org/10.1021/bi011197q)
- [Hann & Oprea 2004] Hann, M.M.; Oprea, T.I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* **2004**, *8*[3], 255–263. doi [10.1016/j.cbpa.2004.04.003](https://doi.org/10.1016/j.cbpa.2004.04.003)
- [Hansch 1969] Hansch, C. A quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*[8], 232–239. doi [10.1021/ar50020a002](https://doi.org/10.1021/ar50020a002)
- [Hansch *et al.* 2004] Hansch, C.; Mekapati, S.B.; Kurup, A.; Verma, R.P. QSAR of cytochrome P450. *Drug Metab. Rev.* **2004**, *36*[1], 105–156. doi [10.1081/DMR-120028428](https://doi.org/10.1081/DMR-120028428)
- [Hansch & Zhang 1993] Hansch, C.; Zhang, L. Quantitative structure-activity relationships of cytochrome P450. *Drug Metab. Rev.* **1993**, *25*[1-2], 1–48. doi [10.3109/03602539308993972](https://doi.org/10.3109/03602539308993972)
- [Harlow & Halpert 1997] Harlow, G.R.; Halpert, J.R. Alanine-scanning mutagenesis of a putative substrate recognition site in human cytochrome P450 3A4. Role of residues 210 and 211 in flavonoid activation and substrate specificity. *J. Biol. Chem.* **1997**, *272*[9], 5396–5402. doi [10.1074/jbc.272.9.5396](https://doi.org/10.1074/jbc.272.9.5396)
- [Harlow & Halpert 1998] Harlow, G.R.; Halpert, J.R. Analysis of human cytochrome P450 3A4 cooperativity: Construction and characterization of a site-directed mutant that displays hyperbolic steroid hydroxylation kinetics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*[12], 6636–6641. PMID [9618464](https://pubmed.ncbi.nlm.nih.gov/9618464/)
- [Hasemann *et al.* 1995] Hasemann, C.A.; Kurumbail, R.G.; Boddupalli, S.S.; Peterson, J.A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **1995**, *3*[1], 41–62. doi [10.1016/S0969-2126\(01\)00134-4](https://doi.org/10.1016/S0969-2126(01)00134-4)
- [Hasemann *et al.* 1994] Hasemann, C.A.; Ravichandran, K.G.; Peterson, J.A.; Deisenhofer, J. Crystal structure and refinement of cytochrome P450<sub>terp</sub> at 2.3 Å resolution. *J. Mol. Biol.* **1994**, *236*[4], 1169–1185. doi [10.1016/0022-2836\(94\)90019-1](https://doi.org/10.1016/0022-2836(94)90019-1)
- [Hayaishi *et al.* 1955] Hayaishi, O.; Katagiri, M.; Rothberg, S. Mechanism of the pyrocatechase reaction. *J. Am. Chem. Soc.* **1955**, *77*[20], 5450–5451. doi [10.1021/ja01625a095](https://doi.org/10.1021/ja01625a095)
- [Hayano *et al.* 1955] Hayano, M.; Lindberg, M.C.; Dorfman, R.I.; Hancock, J.E.H.; von E. Doering, W. On the mechanism of the C-11 $\beta$ -hydroxylation of steroids;

- A study with H<sub>2</sub>O<sup>18</sup> and O<sub>2</sub><sup>18</sup>. *Arch. Biochem. Biophys.* **1955**, 59[2], 529–532. doi [10.1016/0003-9861\(55\)90520-8](https://doi.org/10.1016/0003-9861(55)90520-8)
- [He *et al.* 2003] He, Y.A.; Roussel, F.; Halpert, J.R. Analysis of homotropic and heterotropic cooperativity of diazepam oxidation by CYP3A4 using site-directed mutagenesis and kinetic modeling. *Arch. Biochem. Biophys.* **2003**, 409[1], 92–101. doi [10.1016/S0003-9861\(02\)00484-8](https://doi.org/10.1016/S0003-9861(02)00484-8)
- [Headlam *et al.* 2003] Headlam, M.J.; Wilce, M.C.; Tuckey, R.C. The F-G loop region of cytochrome P450sc (CYP11A1) interacts with the phospholipid membrane. *Biochim. Biophys. Acta* **2003**, 1617[1-2], 96–108. doi [10.1016/j.bbamem.2003.09.007](https://doi.org/10.1016/j.bbamem.2003.09.007). Erratum: *Biochim Biophys Acta.* **2004**, 1661[2], 219. doi [10.1016/j.bbamem.2004.01.004](https://doi.org/10.1016/j.bbamem.2004.01.004)
- [Hendlich *et al.* 1997] Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: Automatic and efficient detection of potential small molecule binding sites in proteins. *J. Mol. Gr. Model.* **1997**, 15[6], 359–363. doi [10.1016/S1093-3263\(98\)00002-3](https://doi.org/10.1016/S1093-3263(98)00002-3)
- [Hendrychova *et al.* 2012] Hendrychova, T.; Berka K, Navratilova V, Anzenbacher P, Otyepka M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.* **2012**, 13[2], 177–189. doi [10.2174/138920012798918408](https://doi.org/10.2174/138920012798918408)
- [Hutzler & Tracy 2002] Hutzler, J.M.; Tracy, T.S. Atypical kinetic profiles in drug metabolism reactions. *Drug Metab. Dispos.* **2002**, 30[4], 355–362. doi [10.1124/dmd.30.4.355](https://doi.org/10.1124/dmd.30.4.355)
- [Ishigami *et al.* 2001] Ishigami, M.; Honda, T.; Takasaki, W.; Ikeda, T.; Komai, T.; Ito, K.; Sugiyama, Y. A comparison of the effects of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors on the CYP3A4-dependent oxidation of mexazolam in vitro. *Drug Metab. Dispos.* **2001**, 29[3], 282–288. PMID [11181496](https://pubmed.ncbi.nlm.nih.gov/11181496/)
- [Im *et al.* 2000] Im, W.; Seefeld, S.; Roux, B. A grand canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels. *Biophys. J.* **2000**, 79[2], 788–801. doi [10.1016/S0006-3495\(00\)76336-3](https://doi.org/10.1016/S0006-3495(00)76336-3)
- [Isin & Guengerich 2006] Isin, E.M.; Guengerich, F.P. Kinetics and thermodynamics of ligand binding by cytochrome P450 3A4. *J. Biol. Chem.* **2006**, 281[14], 9127–9136. doi [10.1074/jbc.M511375200](https://doi.org/10.1074/jbc.M511375200)
- [Isin & Guengerich 2007] Isin, E.M.; Guengerich, F.P. Multiple sequential steps involved in the binding of inhibitors to cytochrome P450 3A4. *J. Biol. Chem.* **2007**, 282[9], 6863–6874. doi [10.1074/jbc.M610346200](https://doi.org/10.1074/jbc.M610346200)
- [Isin & Guengerich 2008] Isin, E.M.; Guengerich, F.P. Substrate binding to cytochromes P450. *Anal. Bioanal. Chem.* **2008**, 392[6], 1019–1030. doi [10.1007/s00216-008-2244-0](https://doi.org/10.1007/s00216-008-2244-0)
- [Jankowsky *et al.* 2011] Jankowski, C.K.; Chiasson, J.B.; Dako, E.; Doucet, K.; Surette, M.E.; André, F.; Delaforge, M. Modeling of cytochrome P450 (Cyt P450, CYP) channels. *Spectroscopy* **2011**, 25[2], 63–87. doi [10.3233/SPE-2011-0499](https://doi.org/10.3233/SPE-2011-0499)
- [Jean *et al.* 1997] Jean, P.; Pothier, J.; Dansette, P.M.; Mansuy, D.; Viari, A. Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450. *Prot. Struct. Funct. Gen.* **1997**, 28[3], 388–404. doi [10.1002/\(SICI\)1097-0134\(199707\)28:3<388::AID-PROT9>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-0134(199707)28:3<388::AID-PROT9>3.0.CO;2-8)

- [Johnson *et al.* 2002] Johnson, E.F.; Wester ,M.R.; Stout ,C.D. The structure of microsomal cytochrome P450 2C5: a steroid and drug metabolizing enzyme. *Endocr. Res.* **2002**, *28*[4], 435–441. doi [10.1081/ERC-120016820](https://doi.org/10.1081/ERC-120016820)
- [Jones *et al.* 1996] Jones, B.C.; Hawksworth, G.; Horne, V.A.; Newlands, A.; Morsman, J.; Tute, M.S.; Smith, D.A. Putative active site template model for cytochrome P4502C9 (tolbutamide hydroxylase). *Drug Metab. Dispos.* **1996**, *24*[2], 260–266. PMID [8742240](https://pubmed.ncbi.nlm.nih.gov/8742240/)
- [Kalra 2007] Kalra, B.S. Cytochrome P450 enzyme isoforms and their therapeutic implications: An update. *Indian J. Med. Sci.* **2007**, *61*[2], 102–116. <http://www.indianjmedsci.org/text.asp?2007/61/2/102/30351>
- [Kaur *et al.* 2015] Kaur, P.; Chamberlin, R.; Sevrioukova, I.F.; Poulos, T.L. Structure-based inhibitor design for evaluation of a CYP3A4 pharmacophore model. *J. Med. Chem.* **2015**. doi [10.1021/acs.jmedchem.5b01146](https://doi.org/10.1021/acs.jmedchem.5b01146)
- [Kennedy 1997] Kennedy, T. Managing the drug discovery/development interface. *Drug Discov. Today* **1997**, *2*[10], 436–444. doi [10.1016/S1359-6446\(97\)01099-4](https://doi.org/10.1016/S1359-6446(97)01099-4)
- [Kenworthy *et al.* 1999] Kenworthy, K.E.; Bloomer, J.C.; Clarke, S.E.; Houston, J.B. CYP3A4 drug interactions: correlation of 10 *in vitro* probe substrates. *J. Clin. Pharmacol.* **1999**, *48*[5], 716–727. doi [10.1046/j.1365-2125.1999.00073.x](https://doi.org/10.1046/j.1365-2125.1999.00073.x)
- [Kenworthy *et al.* 2001] Kenworthy, K.E.; Clarke S.E.; Andrews, J.; Houston, J.B. Multisite kinetic models for CYP3A4: Simultaneous activation and inhibition of diazepam and testosterone metabolism. *Drug Metab. Dispos.* **2001**, *29*[12], 1644–1651. PMID [11717184](https://pubmed.ncbi.nlm.nih.gov/11717184/)
- [Kerns & Di 2003] Kerns, E.H.; Di, L. Pharmaceutical profiling in drug discovery. *Drug Discov. Today* **2003**, *8*[7], 316–323. doi [10.1016/S1359-6446\(03\)02649-7](https://doi.org/10.1016/S1359-6446(03)02649-7)
- [Kim & Guengerich 2005] Kim, D.; Guengerich, F.P. Cytochrome P450 activation of arylamines and heterocyclic amines. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 27–49. doi [10.1146/annurev.pharmtox.45.120403.100010](https://doi.org/10.1146/annurev.pharmtox.45.120403.100010)
- [Kim *et al.* 2006] Kim, K.H.; Isin, E.M.; Yun, C.H.; Kim, D.H.; Guengerich, F.P. Kinetic deuterium isotope effects for 7-alkoxycoumarin O-dealkylation reactions catalyzed by human cytochromes P450 and in liver microsomes. Rate-limiting C-H bond breaking in cytochrome P450 1A2 substrate oxidation. *FEBS J.* **2006**, *273*[10], 2223–2231. doi [10.1111/j.1742-4658.2006.05235.x](https://doi.org/10.1111/j.1742-4658.2006.05235.x)
- [Kingsley & Lill 2014a] Kingsley, L.J.; Lill, M.A. IterTunnel; a method for predicting and evaluating ligand EgressTunnels in proteins with buried active sites. *J. Cheminform.* **2014**, *6*[suppl. 1], P62. doi [10.1186/1758-2946-6-S1-P62](https://doi.org/10.1186/1758-2946-6-S1-P62)
- [Kingsley & Lill 2014b] Kingsley, L.J.; Lill, M.A. Including ligand-induced protein flexibility into protein tunnel prediction. *J. Comput. Chem.* **2014**, *35*[24], 1748–1756. doi [10.1002/jcc.23680](https://doi.org/10.1002/jcc.23680)
- [Kingsley & Lill 2014c] Kingsley, L.J.; Lill, M.A. Ensemble generation and the influence of protein flexibility on geometric tunnel prediction in cytochrome P450 enzymes. *PLOS One* **2014**, *9*[6], e99408. doi [10.1371/journal.pone.0099408](https://doi.org/10.1371/journal.pone.0099408)
- [Kingsley & Lill 2015] Kingsley, L.J.; Lill, M.A. Substrate tunnels in enzymes: Structure–function relationships and computational methodology. *Prot. Struct. Funct. Bioinf.* **2015**, *83*[4], 599–611. doi [10.1002/prot.24772](https://doi.org/10.1002/prot.24772)

- [Kirillova *et al.* 2008] Kirillova, S.; Cortés, J.; Stefaniu, A.; Siméon, T. An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins. *Prot. Struct. Funct. Bioinf.* **2008**, *70*[1], 131–143. doi [10.1002/prot.21570](https://doi.org/10.1002/prot.21570)
- [Kleywegt & Jones 1994] Kleywegt, G.J.; Jones, T.A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1994**, *50*[2], 178–185. doi [10.1107/S0907444993009989](https://doi.org/10.1107/S0907444993009989)
- [Klingenberg 1958] Klingenberg, M. Pigments of rat liver microsomes. *Arch. Biochem. Biophys.* **1958**, *75*[2], 376–386. PMID [13534720](https://pubmed.ncbi.nlm.nih.gov/13534720/)
- [Koley *et al.* 1995] Koley, A.P.; Buters, J.T.M.; Robinson, R.C.; Markowitz, A.; Friedman, F.K. CO binding kinetics of human cytochrome P450 3A4: Specific interaction of substrates with kinetically distinguishable conformers. *J. Biol. Chem.* **1995**, *270*[10], 5014–5018. doi [10.1074/jbc.270.10.5014](https://doi.org/10.1074/jbc.270.10.5014)
- [Koley *et al.* 1997] Koley, A.P.; Buters, J.T.M.; Robinson, R.C.; Markowitz, A.; Friedman, F.K. Differential mechanisms of cytochrome P450 inhibition and activation by  $\alpha$ -naphthoflavone. *J. Biol. Chem.* **1997**, *272*[6], 3149–3152. doi [10.1074/jbc.272.6.3149](https://doi.org/10.1074/jbc.272.6.3149)
- [Krebs *et al.* 2002] Krebs, W.G.; Alexandrov V.; Wilson, C.A.; Echols, N.; Yu, H.; Gerstein, M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Prot. Struct. Funct. Gen.* **2002**, *48*[4], 682–695. doi [10.1002/prot.10168](https://doi.org/10.1002/prot.10168)
- [Krishnamoorthy *et al.* 2010] Krishnamoorthy, N.; Gajendrarao, P.; Thangapandian, S.; Lee, Y.; Lee, K.W. Probing possible egress channels for multiple ligands in human CYP3A4: A molecular modeling study. *J. Mol. Model.* **2010**, *16*[4], 607–614. doi [10.1007/s00894-009-0571-5](https://doi.org/10.1007/s00894-009-0571-5)
- [Kronbach *et al.* 1989] Kronbach, T.; Mathys, D.; Umeno, M.; Gonzalez, F.J.; Meyer, U.A. Oxidation of midazolam and triazolam by human liver cytochrome P450III A4. *Mol. Pharmacol.* **1989**, *36*[1], 89–96. PMID [2787473](https://pubmed.ncbi.nlm.nih.gov/2787473/)
- [Kubinyi 1997] Kubinyi, H. QSAR and 3D QSAR in drug design. Part 2: applications and problems. *Drug Discov. Today* **1997**, *2*[12], 538–546. doi [10.1016/S1359-6446\(97\)01084-2](https://doi.org/10.1016/S1359-6446(97)01084-2)
- [Lafite *et al.* 2007] Lafite, P.; André, F.; Zeldin, D.C.; Dansette, P.M.; Mansuy, D. Unusual regioselectivity and active site topology of human cytochrome P450 2J2. *Biochemistry* **2007**, *46*[36], 10237–10247. doi [10.1021/bi700876a](https://doi.org/10.1021/bi700876a)
- [Lee *et al.* 2003] Lee, D.S.; Yamada, A.; Sugimoto, H.; Matsunaga, I.; Ogura, H.; Ichihara, K.; Adachi, S.; Park, S.Y.; Shiro, Y. Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from *Bacillus subtilis*. Crystallographic, spectroscopic, and mutational studies. *J. Biol. Chem.* **2003**, *278*[11], 9761–9767. doi [10.1074/jbc.M211575200](https://doi.org/10.1074/jbc.M211575200)
- [Lee *et al.* 1994] Lee., H.S.; Jin, C.; Park, J.; Kim, D.H. Modulation of cytochrome P450 activities by 7,8-benzoflavone and its metabolites. *Biochem. Mol. Biol. Int.* **1994**, *34*[3], 483–491. PMID [7833826](https://pubmed.ncbi.nlm.nih.gov/7833826/)
- [Leo *et al.* 1971] Leo, A.; Hansch, C.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71* [6], 525–616. doi [10.1021/cr60274a001](https://doi.org/10.1021/cr60274a001)

- [Leo-Macias *et al.* 2005] Leo-Macias, A.; Lopez-Romero, P.; Lupyan, D.; Zerbino, D.; Ortiz, A.R. An analysis of core deformations in protein superfamilies. *Biophys J.* **2005**, *88*[2], 1291–1299. doi [10.1529/biophysj.104.052449](https://doi.org/10.1529/biophysj.104.052449)
- [Lewis 1986] Lewis, D.F.V. Physical methods in the study of the active site geometry of cytochromes P-450. *Drug Metab. Rev.* **1986**, *17*[1-2], 1–66. doi [10.3109/03602538608998285](https://doi.org/10.3109/03602538608998285)
- [Lewis 2003a] Lewis, D.F.V. Human cytochromes P450 associated with the Phase 1 metabolism of drugs and other xenobiotics: a compilation of substrates and inhibitors of the CYP1, CYP2 and CYP3 families. *Curr. Med. Chem.* **2003**, *10*[19], 1955–1972. doi [10.2174/0929867033456855](https://doi.org/10.2174/0929867033456855)
- [Lewis 2003b] Lewis, D.F.V. On the estimation of binding affinity ( $\Delta G_{\text{bind}}$ ) for human P450 substrates (based on  $K_m$  and  $K_{\Delta}$  values). *Curr. Drug Metab.* **2003**, *4*[5], 331–340. doi [10.2174/1389200033489343](https://doi.org/10.2174/1389200033489343)
- [Lewis 2004] Lewis, D.F.V. Hydrogen bonding in human P450-substrate interactions: a major contribution to binding affinity. *Scientific World J.* **2004**, *4*, 1074–1082. doi [10.1100/tsw.2004.210](https://doi.org/10.1100/tsw.2004.210)
- [Lewis & Broughton 2002] Lewis, D.F.V.; Broughton, H.B. Molecular binding interactions: their estimation and rationalization of QSARs in terms of theoretically derived parameters. *Scientific World J.*, **2002**, *2*, 1776–1802. doi [10.1100/tsw.2002.343](https://doi.org/10.1100/tsw.2002.343)
- [Lewis & Dickins 2002] Lewis, D.F.V.; Dickins, M. Substrate SARs in human P450s. *Drug Discov. Today* **2002**, *7*[17], 918–925. doi [10.1016/S1359-6446\(02\)02412-1](https://doi.org/10.1016/S1359-6446(02)02412-1)
- [Lewis & Dickins 2003] Lewis, D.F.V.; Dickins, M. Baseline lipophilicity relationships in human cytochromes P450 associated with drug metabolism. *Drug Metab. Rev.* **2003**, *35*[1], 1–18. doi [10.1081/DMR-120026877](https://doi.org/10.1081/DMR-120026877)
- [Lewis & Ito 2010] Lewis, D.F.V.; Ito, Y. Human CYPs involved in drug metabolism: structures, substrates and binding affinities. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*[6], 661–674. doi [10.1517/17425251003674380](https://doi.org/10.1517/17425251003674380)
- [Lewis *et al.* 2004] Lewis, D.F.V.; Jacobs, M.N.; Dickins, M. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discov. Today* **2004**, *9*[12], 530–537. doi [10.1016/S1359-6446\(04\)03115-0](https://doi.org/10.1016/S1359-6446(04)03115-0)
- [Lewis & Lake 1996] Lewis, D.F.V.; Lake, B.G. Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: rationalization of CYP1A substrate specificity in terms of active site amino acid residues. *Xenobiotica* **1996**, *26*[7], 723–753. doi [10.3109/00498259609046745](https://doi.org/10.3109/00498259609046745)
- [Lewis *et al.* 1995] Lewis, D.F.V.; Lake, B.G.; Parke, D.V. Molecular orbital-generated QSARs in a homologous series of alkoxyresorufins and studies of their interactive docking with P450s. *Xenobiotica* **1995**, *25*[12], 1355–1369. doi [10.3109/00498259509061923](https://doi.org/10.3109/00498259509061923)
- [Li *et al.* 1995] Li, A.P.; Kaminski, D.L.; Rasmussen, A. Substrates of human hepatic cytochrome P450 3A4. *Toxicology* **1995**, *104*[1-3], 1–8. doi [10.1016/0300-483X\(95\)03155-9](https://doi.org/10.1016/0300-483X(95)03155-9)
- [Li & Poulos 1997] Li, H.; Poulos, T.L. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Struct. Biol.* **1997**, *4*[2], 140–146. doi [10.1038/nsb0297-140](https://doi.org/10.1038/nsb0297-140)

- [Li *et al.* 2011b] Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. *Prot. Struct. Funct. Bioinf.* **2011**, *79*[1], 271–281. doi [10.1002/prot.22880](https://doi.org/10.1002/prot.22880)
- [Liang *et al.* 1998] Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*[9], 1884–1897. doi [10.1002/pro.5560070905](https://doi.org/10.1002/pro.5560070905)
- [Lichtarge *et al.* 1996] Lichtarge, O.; Bourne, H.R.; Cohen, F.E. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **1996**, *257*[2], 342–358. doi [10.1006/jmbi.1996.0167](https://doi.org/10.1006/jmbi.1996.0167)
- [Liew *et al.* 2010] Liew, C.W.; Sharff, A.; Kotaka, M.; Kong, R.; Sun, H.; Qureshi, I.; Bricogne, G.; Liang, Z.X.; Lescar, J. Induced-fit upon ligand binding revealed by crystal structures of the hot-dog fold thioesterase in dynamycin biosynthesis. *J. Mol. Biol.* **2010**, *404*[2], 291–306. doi [10.1016/j.jmb.2010.09.041](https://doi.org/10.1016/j.jmb.2010.09.041)
- [Lipinski 2000] Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharm. Tox. Methods* **2000**, *44*[1], 235–249. doi [10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6)
- [Lipinski *et al.* 1997] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. *Adv. Drug Deliv. Rev.* **1997**, *23*[1-3], 3–26. doi [10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
- [Lipper 1999] Lipper, R.A. E Pluribus product. *Modern Drug Discov.* **1999**, *2*[1], 55–60. <http://pubs.acs.org/subscribe/archive/mdd/>
- [Livingstone 2000] Livingstone, D.J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* **2000**, *40*[2], 195–209. doi [10.1021/ci990162i](https://doi.org/10.1021/ci990162i)
- [Lüdemann *et al.* 1997] Lüdemann, S.K.; Carugo O.; Wade, R.C. Substrate access to cytochrome P450cam: A comparison of a thermal motion pathway analysis with molecular dynamics simulation data. *J. Mol. Model.* **1997**, *3*[8], 369–374. doi [10.1007/s008940050053](https://doi.org/10.1007/s008940050053)
- [Lüdemann *et al.* 2000a] Lüdemann, S.K.; Lounnas, V.; Wade, R.C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* **2000**, *303*[5], 797–811. doi [10.1006/jmbi.2000.4154](https://doi.org/10.1006/jmbi.2000.4154)
- [Lüdemann *et al.* 2000b] Lüdemann, S.K.; Lounnas, V.; Wade, R.C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.* **2000**, *303*[5], 813–830. doi [10.1006/jmbi.2000.4155](https://doi.org/10.1006/jmbi.2000.4155)
- [Mancy *et al.* 1995] Mancy, A.; Broto, P.; Dijols, S.; Dansette, P.M.; Mansuy, D. The substrate binding site of human liver cytochrome P450 2C9: an approach using designed tienilic acid derivatives and molecular modeling. *Biochemistry* **1995**, *34*[33], 10365–10375. doi [10.1021/bi00033a007](https://doi.org/10.1021/bi00033a007)
- [Mason *et al.* 1955] Mason, H.S.; Fowlks, W.L.; Peterson, E. Oxygen transfer and electron transport by the phenolase complex. *J. Am. Chem. Soc.* **1955**, *77*[10], 2914–2915. doi [10.1021/ja01615a088](https://doi.org/10.1021/ja01615a088)

- [Meslamani *et al.* 2009] Meslamani, J.E.; André, F.; Petitjean, M. Assessing the geometric diversity of cytochrome P450 ligand conformers by hierarchical clustering with a stop criterion. *J. Chem. Inf. Model.* **2009**, *49*[2], 330–337. doi [10.1021/ci800275k](https://doi.org/10.1021/ci800275k)
- [Miteva *et al.* 2010] Miteva, M.A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*[suppl. 2], W622–W627. doi [0.1093/nar/gkq325](https://doi.org/0.1093/nar/gkq325)
- [Modi 2003] Modi, S. Computational approaches to the understanding of ADMET properties and problems. *Drug Discov. Today* **2003**, *8*[14], 621–623. doi [10.1016/S1359-6446\(03\)02739-9](https://doi.org/10.1016/S1359-6446(03)02739-9)
- [Munro *et al.* 2013] Munro, A.W.; Girvan, H.M.; Mason, A.E.; Dunford, A.J.; McLean, K.J. What makes a P450 tick? *Trends Biochem. Sci.* **2013**, *38*[3], 140–150. doi [10.1016/j.tibs.2012.11.006](https://doi.org/10.1016/j.tibs.2012.11.006)
- [Munro *et al.* 2007] Munro, A.W.; Girvan, H.M.; McLean, K.J. Variations on a (t)heme–novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat. Prod. Rep.* **2007**, *24*[3], 585–609. doi [10.1039/B604190F](https://doi.org/10.1039/B604190F)
- [Mustafa *et al.* 2014] Mustafa, G.; Yu, X.; Wade, R.C. Structure and dynamics of human drug-metabolizing cytochrome P450 enzymes. In: *Drug Metabolism Prediction*. Kirchmair, J. (Ed.). Vol. 63, Wiley, New York, **2014**, sect. 4, pp. 77–102.
- [Nagano & Poulos 2005] Nagano, S.; Poulos, T.L. Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. *J. Biol. Chem.* **2005**, *280*[36], 31659–31663. doi [10.1074/jbc.M505261200](https://doi.org/10.1074/jbc.M505261200)
- [Nebert & Dalton 2006] Nebert, D.W.; Dalton, T.P. The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nature Rev. Cancer* **2006**, *6*[12], 947–960. doi [10.1038/nrc2015](https://doi.org/10.1038/nrc2015)
- [Nguyen 2007] Nguyen, T.A. Relations structure-fonction dans la superfamille des Cytochromes P450. Etudes bioinformatiques. *Thèse de doctorat en informatique*, Université Paris Diderot, 29 octobre **2007** <https://tel.archives-ouvertes.fr/tel-00447215>
- [Nguyen *et al.* 2008] Nguyen, T.A.; Tychopoulos, M.; Bichat, F.; Zimmermann, C.; Flinois, J.P.; Diry, M.; Ahlberg, E.; Delaforge, M.; Corcos, L.; Beaune, P.; Dansette, P.; André, F.; de Waziers, I. Improvement of cyclophosphamide activation by CYP2B6 mutants: from in silico to ex vivo. *Mol. Pharmacol.* **2008**, *73*[4], 1122–1133. doi [10.1124/mol.107.042861](https://doi.org/10.1124/mol.107.042861)
- [Nikolova & Jaworska 2003] Nikolova, N.; Jaworska, J. Approaches to measure chemical similarity – a review. *QSAR Comb. Sci* **2003**, *22*[9-10], 1006–1026. doi [10.1002/qsar.200330831](https://doi.org/10.1002/qsar.200330831)
- [Oliver *et al.* 1997] Oliver, C.F.; Modi, S.; Primrose, W.U.; Lian, L.Y.; Roberts, G.C.K. Engineering the substrate specificity of *Bacillus megaterium* cytochrome P-450 BM3: hydroxylation of alkyl trimethylammonium compounds. *Biochem. J.* **1997**, *327*[2], 537–544. doi [10.1042/bj3270537](https://doi.org/10.1042/bj3270537)
- [Omura 2011] Omura, T. Recollection of the early years of the research on cytochrome P450. *Proc. Jpn. Acad. Ser. B* **2011**, *87*[10], 617–640. doi [10.2183/pjab.87.617](https://doi.org/10.2183/pjab.87.617)

- [Omura & Sato 1962] Omura, T.; Sato, R. A new cytochrome in liver microsomes. *J. Biol. Chem.* **1962**, *237*[4], PC1375–PC1376. PMID 14482007
- [Omura & Sato 1964] Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **1964**, *239*[7], 2370–2378. PMID 14209971
- [Ozols *et al.* 1985] Ozols, J.; Heinemann, F.S.; Johnson, E.F. The complete amino acid sequence of a constitutive form of liver microsomal cytochrome P-450. *J. Biol. Chem.* **1985**, *260*[9], 5427–5434. PMID 3988762
- [Palade & Siekevitz] Palade, G.E.; Siekevitz, P. Liver microsomes, an integrated morphological and biochemical study. *J. Biochem. Biophys. Cytol.* **1956**, *2*[2], 171–200 (plus 13 pages: images). PMID 13319380
- [Palma *et al.* 1994] Palma, P.N.; Moura, I.; LeGall, J.; Van Beeumen, J.; Wampler, J.E.; Moura, J.J. Evidence for a ternary complex formed between flavodoxin and cytochrome c3: 1H-NMR and molecular modeling studies. *Biochemistry* **1994**, *33*[21], 6394–6407. doi 10.1021/bi00187a003
- [Park *et al.* 1997] Park, S.Y.; Shimizu, H.; Adachi, S.; Nakagawa, A.; Tanaka, I.; Nakahara, K.; Shoun, H.; Obayashi, E.; Nakamura, H.; Iizuka, T.; and Shiro, Y. Crystal structure of nitric oxide reductase from denitrifying fungus *Fusarium oxysporum*. *Nature Struct. Biol.* **1997**, *4*[10], 827–832. doi 10.1038/nsb1097-827
- [Park *et al.* 2002] Park, S.Y.; Yamane, K.; Adachi, S.; Shiro, Y.; Weiss, K.E.; Maves, S.A.; Sligar, S.G. Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus*: high resolution structure and functional properties. *J. Inorg. Biochem.* **2002**, *91*[4], 491–501. doi 10.1016/S0162-0134(02)00446-4
- [Pearson *et al.* 2006] Pearson, J.T.; Hill, J.J.; Swank, J.; Isoherranen, N.; Kunze, K.L.; Atkins, W.M. Surface plasmon resonance analysis of antifungal azoles binding to CYP3A4 with kinetic resolution of multiple binding orientations. *Biochemistry* **2006**, *45*[20], 6341–6353. doi 10.1021/bi0600042
- [Pérot *et al.* 2010] Pérot, S.; Sperandio, O.; Miteva, M.A.; Camproux, A.C.; Villoutreix, B.O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* **2010**, *15*[15/16], 656–667. doi 10.1016/j.drudis.2010.05.015
- [Peterson *et al.* 1976] Peterson, J.A.; Ebel, R.E.; O’Keeffe, D.H.; Matsubara, T.; Estabrook, R.W. Temperature dependence of cytochrome P-450 reduction. A model for NADPH-cytochrome P-450 reductase:cytochrome P-450 interaction. *J. Biol. Chem.* **1976**, *251*[13], 4010–4016. PMID 819436
- [Peterson & Graham 1998] Peterson, J.A., Graham, S.E. A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure* **1998**, *6*[9], 1079–1085. doi 10.1016/S0969-2126(98)00109-9
- [Petitjean 1998] Petitjean M. Interactive maximal common 3D substructure searching with the combined SDM/RMS algorithm. *Comp. Chem.* **1998**, *22*[6], 463–465. doi 10.1016/S0097-8485(98)00017-5
- [Petitjean 1999] Petitjean M. On the Root Mean Square Quantitative Chirality and Quantitative Symmetry Measures. *J. Math. Phys.* **1999**, *40*[9], 4587–4595. doi 10.1063/1.532988

- [Petitjean 2002] Petitjean M. Chiral mixtures. *J. Math. Phys.* **2002**, *43*[8], 4147–4157. doi [10.1063/1.1484559](https://doi.org/10.1063/1.1484559)
- [Petitjean 2012] Petitjean, M. About the algebraic solutions of smallest enclosing cylinders problems. *Appl. Alg. Eng. Comm. Comp.* **2012**, *23*[3-4], 151–164. doi [10.1007/s00200-012-0171-y](https://doi.org/10.1007/s00200-012-0171-y)
- [Petřek *et al.* 2007] Petřek, M.; Košinová, P.; Koča, J.; Otyepka, M. MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **2007**, *15*[11], 1357–1363. doi [10.1016/j.str.2007.10.007](https://doi.org/10.1016/j.str.2007.10.007)
- [Petřek *et al.* 2006] Petřek, M.; Otyepka, M.; Banáš, P.; Košinová, P.; Koča, J.; Damborský, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **2006**, *7*[6], 316. doi [10.1186/1471-2105-7-316](https://doi.org/10.1186/1471-2105-7-316)
- [Pikuleva *et al.* 1999] Pikuleva, I.A.; Cao, C.; Waterman, M.R. An additional electrostatic interaction between adrenodoxin and P450c27 (CYP27A1) results in tighter binding than between adrenodoxin and P450scc (CYP11A1) *J. Biol. Chem.* **1999**, *274*[4], 2045–2052. doi [10.1074/jbc.274.4.2045](https://doi.org/10.1074/jbc.274.4.2045)
- [Pisanti *et al.* 2006] Pisanti, N.; Soldano, H.; Carpentier, M; Pothier, J. Implicit and explicit representation of approximated motifs. Iliopoulos, C.; Park, K.; Steinhöfel, K. (Eds.) In: *Algorithms in bioinformatics*, King's College, London, **2006**, *Texts in Algorithms*, vol. 6. <http://www.collegepublications.co.uk/algorithemics/?000006>
- [Podust *et al.* 2004] Podust, L.M.; Yermalitskaya, L.V.; Lepesheva, G.I.; Podust, V.N.; Dalmasso, E.A.; Waterman, M.R. Estriol bound and ligand-free structures of sterol 14 $\alpha$ -demethylase. *Structure* **2004**, *12*[11], 1937–1945. doi [10.1016/j.str.2004.08.009](https://doi.org/10.1016/j.str.2004.08.009)
- [Poulos *et al.* 1985] Poulos, T.L.; Finzel, B.C.; Gunsalus, I.C.; Wagner, G.C.; Kraut, J. The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450. *J. Biol. Chem.* **1985**, *260*[30], 16122–16130. PMID [4066706](https://pubmed.ncbi.nlm.nih.gov/4066706/)
- [Poulos *et al.* 1986] Poulos, T.L.; Finzel, B.C.; Howard, A.J. Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450. *Biochemistry* **1986**, *25*[18], 5314–5322. doi [10.1021/bi00366a049](https://doi.org/10.1021/bi00366a049)
- [Poulos *et al.* 1987] Poulos, T.L., Finzel B.C., Howard A.J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **1987**, *195*[3], 687–700. doi [10.1016/0022-2836\(87\)90190-2](https://doi.org/10.1016/0022-2836(87)90190-2)
- [Poulos & Johnson 2005] Poulos, T.L.; Johnson, E.F. Structures of cytochrome P450 enzymes. In: *Cytochrome P450, Structure, Mechanism, and Biochemistry*. Ortiz de Montellano, P.R. (Ed.). 3<sup>rd</sup> ed., Kluwer/Plenum, New York, **2005**, chap. 3, pp. 87–114. doi [10.1007/b139087](https://doi.org/10.1007/b139087)
- [Pravda *et al.* 2014] Pravda, L.; Berka, K.; Svobodová Vařeková, R.; Sehnal, D.; Banáš, P.; Laskowski, R.A.; Koča, J.; Otyepka, M Anatomy of enzyme channels. *BMC Bioinformatics* **2014**, *15*[1], 379. doi [10.1186/s12859-014-0379-x](https://doi.org/10.1186/s12859-014-0379-x)
- [Prokop *et al.* 2012] Prokop, Z.; Gora, A.; Brezovsky, J.; Chaloupkova, R.; Stepankova, V.; Damborsky, J. Engineering of protein tunnels: the keyhole-lock-key model for catalysis by enzymes with buried active sites. In: *Protein Engineering Handbook*, vol. 3. Lutz, S, Bornscheuer, U.T. (Eds.). Wiley-VCH, Weinheim, **2012**. pp. 421–464. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-3527331239.html>

- [Raag *et al.* 1991] Raag, R.; Martinis, S.A.; Sligar, S.G.; Poulos, T.L. Crystal structure of the cytochrome P-450<sub>CAM</sub> active site mutant Thr252Ala. *Biochemistry* **1991**, *30*[48], 11420–11429. doi [10.1021/bi00112a008](https://doi.org/10.1021/bi00112a008)
- [Ramos-Nino *et al.* 1996] Ramos-Nino, M.E.; Clifford, M.N.; Adams, M.R. Quantitative structure-activity relationship for the effect of benzoic acids, cinnamic acids and benzaldehydes on *Listeria monocytogenes*. *J. Appl. Bacteriol.* **1996**, *80*[3], 303–310. doi [10.1111/j.1365-2672.1996.tb03224.x](https://doi.org/10.1111/j.1365-2672.1996.tb03224.x)
- [Ravichandran *et al.* 1993] Ravichandran, K.G.; Boddupalli, S.S.; Hasemann, C.A.; Peterson, J.A.; Deisenhofer, J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **1993**, *261*[5122], 731–736. doi [10.1126/science.8342039](https://doi.org/10.1126/science.8342039)
- [Rendic 2002] Rendic, S. Summary of information on human CYP enzymes: Human P450 metabolism data. *Drug Metab. Rev.* **2002**, *34*[1-2], 83–448. doi [10.1081/DMR-120001392](https://doi.org/10.1081/DMR-120001392)
- [Rendic & Guengerich 2012] Rendic, S.; Guengerich, F.P. Contributions of human enzymes in carcinogen metabolism. *Chem. Res. Toxicol.* **2012**, *25*[7], 1316–1383. doi [10.1021/tx300132k](https://doi.org/10.1021/tx300132k)
- [Riley *et al.* 2001] Riley, R.J.; Parker, A.J.; Trigg, S.; Manners, C.N. Development of a generalized, quantitative physicochemical model of CYP3A4 inhibition for use in early drug discovery. *Pharm. Res.* **2001**, *18*[5], 652–655. doi [10.1023/A:1011085411050](https://doi.org/10.1023/A:1011085411050)
- [Roberts *et al.* 2005] Roberts, A.G.; Campbell, A.P.; Atkins, W.M. The thermodynamic landscape of testosterone binding to cytochrome P450 3A4: ligand binding and spin state equilibria. *Biochemistry* **2005**, *44*[4], 1353–1366. doi [10.1021/bi0481390](https://doi.org/10.1021/bi0481390)
- [Rodrigues & Rushmore 2004] Rodrigues, A.D.; Rushmore, T.H. Cytochrome P450 pharmacogenetics in drug development: in vitro studies and clinical consequences. *Curr. Drug Metab.* **2002**, *3*[3], 289–309. doi [10.2174/1389200023337522](https://doi.org/10.2174/1389200023337522). Erratum in *Curr. Drug Metab.* **2002**, *3*[5], 559.
- [Rowland *et al.* 2006] Rowland, P.; Blaney, F.E.; Smyth, M.G.; Jones, J.J.; Leydon, V.R.; Oxbrow, A.K.; Lewis, C.J.; Tennant, M.G.; Modi, S.; Eggleston, D.S.; Chenery, R.J.; Bridges, A.M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281*[11], 7614–7622. doi [10.1074/jbc.M511232200](https://doi.org/10.1074/jbc.M511232200)
- [Rui *et al.* 2006] Rui, L.; Pochapsky, S.S.; Pochapsky, T.C. Comparison of the complexes formed by cytochrome P450<sub>cam</sub> with cytochrome b<sub>5</sub> and putidaredoxin, two effectors of camphor hydroxylase activity. *Biochemistry* **2006**, *45*[12], 3887–3897. doi [10.1021/bi052318f](https://doi.org/10.1021/bi052318f)
- [Ryan & Engel 1956] Ryan, K.J.; Engel, L.L. Steroid 21-hydroxylation by adrenal microsomes and reduced triphosphopyridine nucleotide. *J. Am. Chem. Soc.* **1956**, *78*[11], 2654–2655. doi [10.1021/ja01592a099](https://doi.org/10.1021/ja01592a099)
- [Rydberg *et al.* 2007] Rydberg, P.; Rod, T.H.; Olsen, L.; Ryde, U. Dynamics of water molecules in the active-site cavity of human cytochromes P450. *J. Phys. Chem. B* **2007**, *111*[19], 5445–5457. doi [10.1021/jp070390c](https://doi.org/10.1021/jp070390c)
- [Sarver *et al.* 1997] Sarver, J.G.; White, D.; Erhardt, P.; Bachmann, K. Estimating xenobiotic half-lives in humans from rat data: influence of log P. *Environ. Health Perspect.* **1997**, *105*[11], 1204–1209. PMID [9370523](https://pubmed.ncbi.nlm.nih.gov/9370523/)

- [Schleinkofer *et al.* 2005] Schleinkofer, K.; Sudarko; Winn, P.J.; Lüdeman, S.K.; Wade, R.C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO reports* **2005**, *6*[6], 584–589. doi [10.1038/sj.embor.7400420](https://doi.org/10.1038/sj.embor.7400420)
- [Schlichting *et al.* 2000] Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A.M.; Maves, S.A.; Benson, D.E.; Sweet, R.M.; Ringe, D.; Petsko, G.A.; Sligar, S.G. The catalytic pathway of cytochrome p450cam at atomic resolution. *Science* **2000**, *287*[5458], 1615–1622. doi [10.1126/science.287.5458.1615](https://doi.org/10.1126/science.287.5458.1615)
- [Schenkman *et al.* 1967] Schenkman, J.B.; Remmer, H.; Estabrook, R.W. Spectral studies of drug interaction with hepatic microsomal cytochrome. *Mol. Pharmacol.* **1967**, *3*[2], 113–123. PMID [4382749](https://pubmed.ncbi.nlm.nih.gov/4382749/)
- [Schuler & Sligar 2007] Schuler, M.A.; Sligar, S.G. Diversities and differences in P450 systems: an introduction. In: *Metal Ions in Life Sciences*. Sigel, A.; Sigel, H.; Sigel, R.K.O. (Eds). Wiley, **2007**, Vol. 3, chap.1, pp. 1–26. doi [10.1002/9780470028155.ch1](https://doi.org/10.1002/9780470028155.ch1)
- [Schwab *et al.* 1988] Schwab, G.E.; Raucy, J.L.; Johnson, E.F. Modulation of rabbit and human hepatic cytochrome P-450-catalyzed steroid hydroxylations by alpha-naphthoflavone. *Mol. Pharmacol.* **1988**, *33*[5], 493–499. PMID [3367901](https://pubmed.ncbi.nlm.nih.gov/3367901/)
- [Scott & Halpert 2005] Scott, E.E.; Halpert, J.R. Structures of cytochrome P450 3A4. *Trends Biochem. Sci.* **2005**, *30*[1], 5–7. doi [10.1016/j.tibs.2004.11.004](https://doi.org/10.1016/j.tibs.2004.11.004)
- [Scott *et al.* 2003] Scott, E.E.; He, Y.A.; Wester, M.R.; White, M.A.; Chin, C.C.; Halpert, J.R.; Johnson, E.F.; Stout, C.D. An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc. Natl. Acad. Sci. USA* **2003**, *100*[23], 13196–13201. doi [10.1073/pnas.2133986100](https://doi.org/10.1073/pnas.2133986100)
- [Scott *et al.* 2004] Scott, E.E.; White, M.A.; He, Y.A.; Johnson, E.F.; Stout, C.D.; Halpert, J.R. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J. Biol. Chem.* **2004**, *279*[26], 27294–27301. doi [10.1074/jbc.M403349200](https://doi.org/10.1074/jbc.M403349200)
- [Sehnal *et al.* 2013] Sehnal, D.; Svobodová Vařeková, R.; Berka, K.; Pravda, L.; Navrátilová, V.; Banáš, P.; Ionescu, C.M.; Otyepka, M.; Koča, J. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform.* **2013**, *5*, 39. doi [10.1186/1758-2946-5-39](https://doi.org/10.1186/1758-2946-5-39)
- [Serizawa & Matsuoka 1991] Serizawa, N.; Matsuoka, T. A two component-type cytochrome P-450 monooxygenase system in a prokaryote that catalyzes hydroxylation of ML-236B to pravastatin, a tissue-selective inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Biochim. Biophys. Acta* **1991**, *1084*[1], 35–40. doi [10.1016/0005-2760\(91\)90052-J](https://doi.org/10.1016/0005-2760(91)90052-J)
- [Sevrioukova & Poulos 2010] Sevrioukova, I.F.; Poulos, T.L. Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci. USA* **2010**, *107*[43], 18422–18427. doi [10.1073/pnas.1010693107](https://doi.org/10.1073/pnas.1010693107)
- [Sevrioukova & Poulos 2012a] Sevrioukova, I.F.; Poulos, T.L. Interaction of human cytochrome P4503A4 with ritonavir analogs. *Arch. Biochem. Biophys.* **2012**, *520*[2], 108–116. doi [10.1016/j.abb.2012.02.018](https://doi.org/10.1016/j.abb.2012.02.018)

- [Sevrioukova & Poulos 2012b] Sevrioukova, I.F.; Poulos, T.L. Structural and mechanistic insights into the interaction of cytochrome P4503A4 with bromoergocryptine, a type I ligand. *J. Biol. Chem.* **2012**, *287*[5], 3510–3517. doi [10.1074/jbc.M111.317081](https://doi.org/10.1074/jbc.M111.317081)
- [Sevrioukova & Poulos 2013a] Sevrioukova, I.F.; Poulos, T.L. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Trans.* **2013**, *42*[9], 3116–3126. doi [10.1039/C2DT31833D](https://doi.org/10.1039/C2DT31833D)
- [Sevrioukova & Poulos 2013b] Sevrioukova, I.F.; Poulos, T.L. Dissecting cytochrome P450 3A4-ligand interactions using ritonavir analogues. *Biochemistry* **2013**, *52*[26], 4474–4481. doi [10.1021/bi4005396](https://doi.org/10.1021/bi4005396)
- [Sevrioukova & Poulos 2013c] Sevrioukova, I.F.; Poulos, T.L. Pyridine-substituted desoxyritonavir is a more potent inhibitor of cytochrome P450 3A4 than ritonavir. *J. Med. Chem.* **2013**, *56*[9], 3733–3741. doi [10.1021/jm400288z](https://doi.org/10.1021/jm400288z)
- [Sevrioukova & Poulos 2015] Sevrioukova, I.F.; Poulos, T.L. Anion-dependent stimulation of CYP3A4 monooxygenase. *Biochemistry* **2015**, *54*[26], 4083–4096. doi [10.1021/acs.biochem.5b00510](https://doi.org/10.1021/acs.biochem.5b00510)
- [Sgrignani *et al.* 2012] Sgrignani, J.; Magistrato, A. Influence of the membrane lipophilic environment on the structure and on the substrate access/egress routes of the human aromatase enzyme. A computational study. *J. Chem. Inf. Model.* **2012**, *52*[2], 1595–1606. doi [10.1021/ci300151h](https://doi.org/10.1021/ci300151h)
- [Shen *et al.* 2012] Shen, Z.; Cheng, F.; Xu, Y.; Fu, J.; Xiao, W.; Shen, J.; Liu, G.; Li, W.; Tang, Y. Investigation of indazole unbinding pathways in CYP2E1 by molecular dynamics simulations. *PLOS One* **2012**, *7*[3], e33500. doi [10.1371/journal.pone.0033500](https://doi.org/10.1371/journal.pone.0033500). Erratum: *PLOS One* **2012**, *7*[9]. doi [10.1371/annotation/86d9f6df-7175-467c-a9ff-94eac53af128](https://doi.org/10.1371/annotation/86d9f6df-7175-467c-a9ff-94eac53af128)
- [Shen & Strobel 1993] Shen, S.; Strobel, H.W. Role of lysine and arginine residues of cytochrome P450 in the interaction between cytochrome P4502B1 and NADPH-cytochrome P450 reductase. *Arch. Biochem. Biophys.* **1993**, *304*[1], 257–265. doi [10.1006/abbi.1993.1347](https://doi.org/10.1006/abbi.1993.1347)
- [Shou *et al.* 1994] Shou, M.; Grogan, J.; Mancewicz, J.A.; Krausz, K.W.; Gonzalez, F.J.; Gelboin, H.V.; Korzekwa, K.R. Activation of CYP3A4: Evidence for the simultaneous binding of two substrates in a cytochrome P450 active site. *Biochemistry* **1994**, *33*[21], 6450–6455. doi [10.1021/bi00187a009](https://doi.org/10.1021/bi00187a009)
- [Sibson *et al.* 1978] Sibson, R. Locally equiangular triangulations. *Comput. J.* **1978**, *21*[3], 243–245. doi [10.1093/comjnl/21.3.243](https://doi.org/10.1093/comjnl/21.3.243)
- [Sinko 1999] Sinko, P.J. Drug selection in early drug development: screening for acceptable pharmacokinetic properties using combined in vitro and computational approaches. *Curr. Opin. Drug Discov. Devel.* **1999**, *2*[1], 42–48. PMID [19649916](https://pubmed.ncbi.nlm.nih.gov/19649916/)
- [Skopalik *et al.* 2008] Skopalik, J.; Anzenbacher, P.; Otyepka, M. Flexibility of human cytochromes P450: Molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J. Phys. Chem. B.* **2008**, *112*[27], 8165–8173. doi [10.1021/jp800311c](https://doi.org/10.1021/jp800311c)
- [Sligar *et al.* 1984] Sligar, S.G.; Gelb, M.H.; Heimbrook, D.C. Bio-organic chemistry and cytochrome P-450-dependent catalysis. *Xenobiotica* **1984**, *14*[1/2], 63–86. doi [10.3109/00498258409151399](https://doi.org/10.3109/00498258409151399)

- [Smith 1994] Smith, D.A. Chemistry and enzymology: their use in the prediction of human drug metabolism. *Eur. J. Pharm. Sci.* **1994**, *2*[1-2], 69–71. doi [10.1016/0928-0987\(94\)90074-4](https://doi.org/10.1016/0928-0987(94)90074-4)
- [Smith *et al.* 1997a] Smith, D.A.; Ackland, M.J.; Jones, B.C. Properties of cytochrome P450 isoenzymes and their substrates. Part 1: active site characteristics. *Drug Discov. Today* **1997**, *2*[10], 406–414. doi [10.1016/S1359-6446\(97\)01081-7](https://doi.org/10.1016/S1359-6446(97)01081-7)
- [Smith *et al.* 1997b] Smith, D.A.; Ackland, M.J.; Jones, B.C. Properties of cytochrome P450 isoenzymes and their substrates. Part 2: properties of cytochrome P450 substrates. *Drug Discov. Today* **1997**, *2*[11], 479–486. doi [10.1016/S1359-6446\(97\)01081-7](https://doi.org/10.1016/S1359-6446(97)01081-7)
- [Smith & Jones 1992] Smith, D.A.; Jones, B.C. Speculations on the Substrate Structure-Activity Relationship (SSAR) of cytochrome P450 enzymes. *Biochem. Pharmacol.* **1992**, *44*[11], 2089–2098. doi [10.1016/0006-2952\(92\)90333-E](https://doi.org/10.1016/0006-2952(92)90333-E)
- [Smith *et al.* 1996] Smith, D.A.; Jones, B.C.; Walker, D.K. Design of drugs involving the concepts and theories of drug metabolism and pharmacokinetics. *Med. Res. Rev.* **1996**, *16*[3], 243–266. doi [10.1002/\(SICI\)1098-1128\(199605\)16:3<243::AID-MED2>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1098-1128(199605)16:3<243::AID-MED2>3.0.CO;2-Z)
- [Smith & van de Waterbeemd, 1999] Smith, D.A.; van de Waterbeemd, H. Pharmacokinetics and metabolism in early drug discovery. *Curr. Opin. Chem. Biol.* **1999**, *3*[4], 373–378. doi [10.1016/S1367-5931\(99\)80056-8](https://doi.org/10.1016/S1367-5931(99)80056-8)
- [Sousa *et al.* 2006] Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Protein–ligand docking: current status and future challenges. *Prot. Struct. Funct. Bioinf.* **2006**, *65*[1], 15–26. doi [10.1002/prot.21082](https://doi.org/10.1002/prot.21082)
- [Stayton & Sligar 1990] Stayton, P.S.; Sligar, S.G. The cytochrome P-450<sub>cam</sub> binding surface as defined by site-directed mutagenesis and electrostatic modeling. *Biochemistry* **1990**, *29*[32], 7381–7386. doi [10.1021/bi00484a005](https://doi.org/10.1021/bi00484a005)
- [Stierand *et al.* 2006] Stierand, K.; Maaß, P.; Rarey, M. Molecular complexes at a glance: Automated generation of two-dimensional complex diagrams. *Bioinformatics* **2006**, *22*[14], 1710–1716. doi [10.1093/bioinformatics/btl150](https://doi.org/10.1093/bioinformatics/btl150)
- [Stsiapanava *et al.* 2014] Stsiapanava, A.; Olsson, U.; Wan, M.; Kleinschmidt, T.; Rutishauser, D.; Zubarev, R.A.; Samuelsson, B.; Rinaldo-Matthis, A.; Haeggström, J.Z. Binding of Pro-Gly-Pro at the active site of leukotriene A<sub>4</sub> hydrolase/aminopeptidase and development of an epoxide hydrolase selective inhibitor. *Proc. Natl. Acad. Sci. USA* **2014**, *111*[11], 4227–4232. doi [10.1073/pnas.1402136111](https://doi.org/10.1073/pnas.1402136111)
- [Szklarz & Halpert 1997] Szklarz, G.D.; Halpert, J.R. Molecular modeling of cytochrome P450 3A4. *J. Comput.-Aided Mol. Des.* **1997**, *11* [3], 265–272.
- [Tanaka 1998] Tanaka, E. Clinically important pharmacokinetic drug-drug interactions: role of cytochrome P450 enzymes. *J. Clin. Pharmacol. Ther.* **1998**, *23*[6], 403–416. doi [10.1046/j.1365-2710.1998.00086.x](https://doi.org/10.1046/j.1365-2710.1998.00086.x)
- [Tanaka & Knox 1959] Tanaka, T.; Knox, W.E. The nature and mechanism of the tryptophan pyrrolase (peroxidase-oxidase) reaction of *Pseudomonas* and of rat liver. *J. Biol. Chem.* **1959**, *234*[5], 1162–1170. PMID [13654338](https://pubmed.ncbi.nlm.nih.gov/13654338/)

- [Teague 2003] Teague, S.J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2003**, 2[7], 527–541. doi [10.1038/nrd1129](https://doi.org/10.1038/nrd1129)
- [Terfloth *et al.* 2007] Terfloth, L.; Bienfait, B.; Gasteiger, J. Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates. *J. Chem. Inf. Model.* **2007**, 47[4], 1688–1701. doi [10.1021/ci700010t](https://doi.org/10.1021/ci700010t)
- [Testa1995] Testa, B. Drug metabolism. In: *Burger's Medicinal Chemistry and Drug Discovery*, 5<sup>th</sup> ed., vol. 1 *Principles and Practice*. pp. 129–180. Wolff, M.E. (ed.), **1995**, Wiley-Interscience. <http://www.sudoc.fr/019070241>
- [Tufféry *et al.* 1997] Tufféry, P. Etchebest, C. Hazout, S. Prediction of protein side chain conformations: A study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Eng.* **1997**, 10[4], 361–372. doi [10.1093/protein/10.4.361](https://doi.org/10.1093/protein/10.4.361)
- [Ueng *et al.* 1997] Ueng, Y.F.; Kuwabara, T.; Chun, Y.J.; Guengerich, F.P. Cooperativity in oxidations catalyzed by cytochrome P450 3A4. *Biochemistry* **1997**, 36[2], 370–381. doi [10.1021/bi962359z](https://doi.org/10.1021/bi962359z)
- [Unno *et al.* 1996] Unno, M.; Shimada, H.; Toba, Y.; Makino, R.; Ishimura, Y. Role of Arg<sup>112</sup> of cytochrome p450<sub>cam</sub> in the electron transfer from reduced putidaredoxin. Analyses with site-directed mutants. *J. Biol. Chem.* **1996**, 271[30], 17869–17874. doi [10.1074/jbc.271.30.17869](https://doi.org/10.1074/jbc.271.30.17869)
- [Urban *et al.* 2014] Urban, P.; Truan, G.; Pompon, D. High-throughput functional screening of steroid substrates with wild-type and chimeric P450 enzymes. *BioMed Res. Int.* **2014**, 2014, 764102. doi [10.1155/2014/764102](https://doi.org/10.1155/2014/764102)
- [Urban *et al.* 2015] Urban, P.; Truan, G.; Pompon, D. Access channels to the buried active site control substrate specificity in CYP1A P450 enzymes. *Biochim. Biophys. Acta* **2015**, 1850[4], 696–707. doi [10.1016/j.bbagen.2014.12.015](https://doi.org/10.1016/j.bbagen.2014.12.015)
- [van de Waterbeemd *et al.* 2001] van de Waterbeemd, H.; Smith, D.A.; Beaumont, K.; Walker, D.K. Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, 44[9], 1313–1333. doi [10.1021/jm000407e](https://doi.org/10.1021/jm000407e)
- [van de Waterbeemd & Testa 1987] van de Waterbeemd, H.; Testa, B. The parametrization of lipophilicity and other structural properties in drug design. In: *Advances in Drug Research*, **1987**, vol. 16, pp. 85–225. Testa, B. (Ed.). Academic Press, London. doi [10.1016/0165-6147\(88\)90181-2](https://doi.org/10.1016/0165-6147(88)90181-2)
- [Vermeir *et al.* 1992a] Vermeir, M.; Boens, N.; Heirwegh, K.P.M. Interaction of 7-n-alkoxycoumarins with cytochrome P450<sub>2</sub> and their partitioning into liposomal membranes. *Biochem. J.* **1992**, 284[2], 483–490. doi [10.1042/bj2840483](https://doi.org/10.1042/bj2840483)
- [Vermeir *et al.* 1992b] Vermeir, M.; Boens, N.; Heirwegh, K.P.M. Modelling of interaction of basic lipophilic ligands with cytochrome P450 reconstituted in liposomes. Determination of membrane partition coefficients of S-(–)-nicotine and N,N-diethylaniline from spectral binding studies and fluorescence quenching. *Biochim. Biophys. Acta (Biomembranes)* **1992**, 1107[1], 93–104. doi [10.1016/0005-2736\(92\)90333-H](https://doi.org/10.1016/0005-2736(92)90333-H)
- [Vidakovic *et al.* 1998] Vidakovic, M.; Sligar, S.G.; Li, H.; Poulos, T.L. Understanding the role of the essential Asp251 in cytochrome P450<sub>cam</sub> using site-directed mutagenesis, crystallography, and kinetic solvent isotope effect. *Biochemistry* **1998**, 37[26], 9211–9219. doi [10.1021/bi980189f](https://doi.org/10.1021/bi980189f)

- [von Moltke *et al.* 1998] von Moltke, L.L.; Greenblatt, D.J.; Schmider, J.; Wright, C.E.; Harmatz, J.S.; Shader, R.I. In vitro approaches to predicting drug interactions in vivo. *Biochem. Pharmacol.* **1998**, *55*[2], 113–122. doi [10.1016/S0006-2952\(97\)00239-6](https://doi.org/10.1016/S0006-2952(97)00239-6)
- [Voronoi 1908] Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs. *J. Reine Angew. Math.* **1908**, *134*, 198–287. doi [10.1515/crll.1908.134.198](https://doi.org/10.1515/crll.1908.134.198)
- [Wada & Waterman 1992] Wada, A.; Waterman, M.R. Identification by site-directed mutagenesis of two lysine residues in cholesterol side chain cleavage cytochrome P450 that are essential for adrenodoxin binding. *J. Biol. Chem.* **1992**, *267*[32], 22877–22882. PMID [1429635](https://pubmed.ncbi.nlm.nih.gov/1429635/)
- [Wade 1990] Wade, R.C. Solvation of the active site of cytochrome P450-cam. *J. Comput.-Aided Mol. Des.* **1990**, *4*[2], 199–204. PMID [2213064](https://pubmed.ncbi.nlm.nih.gov/2213064/)
- [Wade *et al.* 2004] Wade, R.C.; Winn, P.J.; Schlichting, I.; Sudarko. A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.* **2004**, *98*[7], 1175–1182. doi [10.1016/j.jinorgbio.2004.02.007](https://doi.org/10.1016/j.jinorgbio.2004.02.007)
- [Waller *et al.* 1996] Waller, C.L.; Evans, M.V.; McKinney, J.D. Modeling the cytochrome P450-mediated metabolism of chlorinated volatile organic compounds. *Drug Metab. Dispos.* **1996**, *24*[2], 203–210. PMID [8742232](https://pubmed.ncbi.nlm.nih.gov/8742232/)
- [Walker *et al.* 1995] Walker, D.K.; Jones, B.C.; Smith, D.A. Differentiation of compounds which bind to P450 isozymes CYP2C9 and CYP3A4 by physicochemical properties. *Br. J. Clin. Pharmacol.* **1995**, *40*[2], 178P. PMCID : [PMC1365181](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC1365181/)
- [Wermuth 2004] Wermuth, C.G. Selective optimization of side activities: another way for drug discovery. *J. Med. Chem.* **2004**, *47*[6], 1303–1314. doi [10.1021/jm030480f](https://doi.org/10.1021/jm030480f)
- [Wester *et al.* 2003] Wester, M.R.; Johnson E.F., Marques-Soares C., Dijols S., Dansette P.M., Mansuy D., Stout C.D. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: Evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*[31], 9335–9345. doi [10.1021/bi0345561](https://doi.org/10.1021/bi0345561)
- [White & Coon 1980] White, R.E.; Coon, M.J. Oxygen activation by cytochrome P-450. *Ann. Rev. Biochem.* **1980**, *49*, 315–356. doi [10.1146/annurev.bi.49.070180.001531](https://doi.org/10.1146/annurev.bi.49.070180.001531)
- [Wikström *et al.* 2005] Wikström, M.; Ribacka, C.; Molin, M.; Laakkonen, L.; Verkhovsky, M.; Puustinen, A. Gating of proton and water transfer in the respiratory enzyme cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* **2005**, *102*[30], 10478–10481. doi [10.1073/pnas.0502873102](https://doi.org/10.1073/pnas.0502873102)
- [Wilkinson 1996] Wilkinson, G.R. Cytochrome P4503A (CYP3A) metabolism: prediction of in vivo activity in humans. *J. Pharmacokinet. Biopharm.* **1996**, *24*[5], 475–490. doi [10.1007/BF02353475](https://doi.org/10.1007/BF02353475)
- [Williams *et al.* 2000] Williams, P.A.; Cosme, J.; Sridhar, V.; Johnson, E.F.; McRee, D.E. Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Mol. Cell* **2000**, *5*[1], 121–131. doi [10.1016/S1097-2765\(00\)80408-6](https://doi.org/10.1016/S1097-2765(00)80408-6)
- [Williams *et al.* 2004] Williams, P.A.; Cosme, J.; Vinković, D.M.; Ward, A.; Angove, H.C.; Day, P.J.; Vornrhein, C.; Tickle, I.J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*[5684], 683–686. doi [10.1126/science.1099736](https://doi.org/10.1126/science.1099736)

- [Winn *et al.* 2002] Winn, P.J.; Lüdemann, S.K.; Gauges, R.; Lounnas, V.; Wade, R.C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. USA* **2002**, *99*[8], 5361–5366. doi [10.1073/pnas.082522999](https://doi.org/10.1073/pnas.082522999)
- [Wolf 2015] Wolf, R.M. Extracting ligands from receptors by reversed targeted molecular dynamics. *J. Comput.-Aided Mol. Des.* **2015**, *29*[11], 1025–1034. doi [10.1007/s10822-015-9863-2](https://doi.org/10.1007/s10822-015-9863-2)
- [Woo *et al.* 2004] Woo, H.-J.; Dinner, A.R.; Roux, B. Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **2004**, *121*[13], 6392–6400. doi [10.1063/1.1784436](https://doi.org/10.1063/1.1784436)
- [Xue *et al.* 2001] Xue, L.; Wang, H.F.; Wang, Q.; Szklarz, G.D.; Domanski, T.L.; Halpert, J.R.; Correia, M.A.. Influence of P450 3A4 SRS-2 residues on cooperativity and/or regioselectivity of aflatoxin B<sub>1</sub> oxidation. *Chem. Res. Toxicol.* **2001**, *14*[5], 483–491.
- [Yaffe *et al.* 2008a] Yaffe, E.; Fishelovitch, D.; Wolfson, H.J.; Halperin, D.; Nussinov, R. MolAxis: Efficient and accurate identification of channels in macromolecules. *Prot. Struct. Funct. Bioinf.* **2008**, *73*[1], 72–86. doi [10.1002/prot.22052](https://doi.org/10.1002/prot.22052)
- [Yaffe *et al.* 2008b] Yaffe, E.; Fishelovitch, D.; Wolfson, H.J.; Halperin, D.; Nussinov, R. MolAxis: a server for identification of channels in macromolecules. *Nucleic Acids Res.* **2008**, *36*[suppl. 2], W210–W215. doi [10.1093/nar/gkn223](https://doi.org/10.1093/nar/gkn223)
- [Yano *et al.* 2004] Yano, J.K.; Wester, M.R.; Schoch, G.A.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.* **2004**, *279*[37], 38091–38094. doi [10.1074/jbc.C400293200](https://doi.org/10.1074/jbc.C400293200)
- [Yu *et al.* 2013] Yu, X.; Cojocaru, V.; Wade, R.C. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotech. Appl. Biochem.* **2013**, *60*[1], 134–145. doi [10.1002/bab.1074](https://doi.org/10.1002/bab.1074)
- [Yuan *et al.* 2002] Yuan, R.; Madani, S.; Wei, X.X.; Reynolds, K.; Huang, S.M. Evaluation of cytochrome P450 probe substrates commonly used by the pharmaceutical industry to study in vitro drug interactions *Drug Metab. Dispos.* **2002**, *30*[12], 1311–1319. doi [10.1124/dmd.30.12.1311](https://doi.org/10.1124/dmd.30.12.1311)
- [Zhang2007 & Bruice 2007] Zhang, X.; Bruice, T.C. Histone lysine methyltransferase SET7/9: formation of a water channel precedes each methyl transfer. *Biochemistry* **2007**, *46*[51], 14838–14844. doi [10.1021/bi7014579](https://doi.org/10.1021/bi7014579)
- [Zhao *et al.* 2005] Zhao, B.; Guengerich, F.P.; Voehler, M.; Waterman, M.R. Role of active site water molecules and substrate hydroxyl groups in oxygen activation by cytochrome P450 158A2: A new mechanism of proton transfer. *J. Biol. Chem.* **2005**, *280*[51], 42188–42197. doi [10.1074/jbc.M509220200](https://doi.org/10.1074/jbc.M509220200)
- [Zhao *et al.* 2006] Zhao, Y.; White, M.A.; Muralidhara, B.K.; Sun, L.; Halpert, J.R.; Stout, C.D.; Scott, E.E. Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: Insight into P450 conformational plasticity and membrane interaction. *J. Biol. Chem.* **2006**, *281*[9], 5973–5981. doi [10.1074/jbc.M511464200](https://doi.org/10.1074/jbc.M511464200)

[Zhou 2008] Zhou, S.F. Drugs behave as substrates, inhibitors and inducers of human cytochrome P450 3A4. *Curr. Drug Metab.* **2008**, *9*[4], 310–322. doi [10.2174/138920008784220664](https://doi.org/10.2174/138920008784220664)

**Etude et applications de nouveaux modèles géométriques des canaux d'accès au site actif de certains cytochromes P450 humains par des ligands volumineux.**

**Résumé :**

Les cytochromes P450s (CYPs) sont des hémoprotéines intervenant dans la fonction de détoxification cellulaire. Le site actif des CYPs est enfoui dans la protéine, mais accessible aux ligands par des canaux. A l'aide d'une méthode récente basée sur la triangulation de Delaunay de la protéine, et implémentée dans le logiciel CCCPP, nous avons modélisé géométriquement ces canaux pour plusieurs isoformes humaines, dont le 3A4, présent au niveau du foie humain et responsable de la métabolisation d'un nombre important de médicaments, afin de constituer un filtre stérique destiné au criblage virtuel rapide de chimiothèques. Cette approche nous a permis d'obtenir des informations sur les mécanismes d'ouverture et de fermeture des canaux, permettant d'expliquer comment des ligands volumineux peuvent accéder au site actif. Ces résultats confirment et étendent ceux de la littérature, et peuvent contribuer à l'élaboration de médicaments nouveaux ou ayant moins d'effets secondaires.

**Mots-clés :**

cytochrome P450; CYP 3A4; site actif; cavité; canal; triangulation de Delaunay

---

**Analysis and applications of new geometrical models of active site access channels of some human cytochromes P450 for large ligands.**

**Abstract:**

The cytochromes P450s (CYPs) are hemoproteins involved in the cellular detoxification function. The CYPs active site is buried inside the protein, but it can be accessed by the ligands through channels. With a recent method based upon the Delaunay triangulation of the protein, and implemented in the CCCPP software, we modeled geometrically these channels for several human isoforms, including the 3A4, located in the human liver and responsible of the metabolization of an important number of drugs, in order to build a sterical filter devoted to high throughput virtual screening of chemical libraries. This approach let us to get information on mechanisms of opening and closing of the channels, allowing to explain how large ligands can access to the active site. These results are in agreement and extend those found in the literature, and can contribute to the design of new drugs or of drugs having less side effects.

**Keywords:**

cytochrome P450; CYP 3A4; active site; cavity; channel; Delaunay triangulation