



**HAL**  
open science

# La multiplicité de transport de la P-glycoprotéine : Etudes de modélisation comparative et de docking au sein de la famille des protéines ABC

Anis Bessadok

► **To cite this version:**

Anis Bessadok. La multiplicité de transport de la P-glycoprotéine : Etudes de modélisation comparative et de docking au sein de la famille des protéines ABC. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université Pierre et Marie Curie - Paris VI, 2011. Français. NNT: . tel-00711662

**HAL Id: tel-00711662**

**<https://theses.hal.science/tel-00711662>**

Submitted on 25 Jun 2012

**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.

**THESE DE DOCTORAT DE  
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité

Bioinformatique Structurale  
(Ecole doctorale iViV 387)

Présentée par

M. Anis Bessadok

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

**La multiplicité de transport de la P-glycoprotéine :  
Etudes de modélisation comparative et de docking  
au sein de la famille des protéines ABC**

soutenue le 11 Juillet 2011

devant le jury composé de :

M. Marcel DELAFORGE	Directeur de thèse
M. Jean Michel JAULT	Rapporteur
M. Daniel PICOT	Rapporteur
M. Germain TRUGNAN	Examineur
M. Jacques CHOMILIER	Examineur
M. Philippe VAYER	Examineur
M. François ANDRE	Examineur

## Résumé

La P-glycoprotéine (P-gp) appartient à la famille des transporteurs ABC qui confèrent, aux microorganismes pathogènes et cellules tumorales humaines, par transport actif à travers la membrane plasmique, une résistance à de multiples molécules antibiotiques et anticancéreuses sans parenté structurale. Malgré les quelques structures de transporteurs ABC bactériens, la caractérisation par microscopie électronique de la P-gp, et la récente structure de P-gp de souris déposée dans la Protein Data Bank (PDB), obtenir une structure pour la P-gp humaine est d'un intérêt particulier à cause de son importance clinique. Actuellement, il n'existe pas de modèle structural d'interaction P-gp/substrat permettant d'expliquer sa multispecificité. Par modélisation par homologie, nous avons reconstruit trois structures de la Pgp humaine: une en présence et deux autres en absence de nucléotide. La liaison du nucléotide change l'accessibilité du transporteur de la face cytoplasmique vers la face extracellulaire. Ces trois états conformationnels ont été placés dans un environnement membranaire afin de révéler la localisation des mutations spécifiques altérant la liaison de la P-gp à certains médicaments. Enfin, nous avons réalisé une étude de docking sur deux structures modélisées de P-gp de Hamster dans les deux orientations. Ce docking a concerné trois molécules sans aucune ressemblance structurale (vérapamil, vinblastine et tentoxine) mais dont les fixations sont mutuellement soit compétitives, soit non-compétitives. Les meilleures poses obtenues sont compatibles avec l'existence de deux pharmacophores distincts pour la reconnaissance des drogues transportées par la P-gp.

**Mots clés :** Transporteurs ABC, P-glycoprotéine, modélisation structurale comparative, amarrage moléculaire.

**Title:** The multispecificity of transport of P-glycoprotein: Comparative modeling and docking studies within the family of ABC proteins

## Abstract

Some ATP-binding-cassette (ABC) transporters confer multidrug resistance (MDR) to pathogenic microorganisms and human tumor cells by mediating the extrusion of a variety of chemotherapeutic drugs out of the cell. One major resistance to chemotherapy has been correlated with the presence of these ABC transporters, acting as molecular « pumps » that actively transport drugs out of the cell, the most prevalent of these being P-glycoprotein (P-gp). Although the X-ray structures of two bacterial ABC proteins have been determined, and the first structures of mouse P-gp released recently in the Protein Data Bank (PDB), establishing molecular structures of human P-gp is still of particular interest because of its clinical relevance.

Using homology modeling tools, we rebuilt three structures of human P-gp in different conformations: one nucleotide-bound and two in the absence of nucleotide. Nucleotide binding swaps the accessibility of the transporter from cytoplasmic (inward) facing to extracellular (outward) facing. We placed the three modeled conformational states in a native-like membrane environment, in order to locate several known mutations that alter specificity of drug binding, in particular relatively to the membrane surroundings.

We then rebuilt two modeled structures of hamster P-gp in inward and outward conformations, and conducted docking studies of three molecules without any structural similarity (verapamil, vinblastine and tentoxin), but having competitive, non competitive and partial overlapping binding modes. The best poses obtained are consistent with the existence of two different pharmacophores for the recognition of the drug transported by P-glycoprotein.

**Keywords:** ABC transporters, comparative modeling, docking, P-glycoprotein.

**Adresse:** Laboratoire Stress Oxydant et Détoxication, URA 2096 du CNRS. CEA-Saclay, 91191 Gif-sur-Yvette cedex

# Sommaire

## Chapitre A : Introduction générale

I.	Les transporteurs ABC.....	6
I.1.	Organisation structurale des transporteurs ABC.....	8
I.2.	Les transporteurs ABC et leurs rôles dans les maladies génétiques.....	10
I.3.	Les sous-familles de transporteurs ABC.....	13
II.	ABCB1 : La P-glycoprotéine ou P-gp.....	17
II.1.	Généralités : MDR et Détoxification.....	17
II.2.	Localisation tissulaire de la P-gp.....	19
II.3.	La P-gp d'un point de vue fonctionnel.....	20
II.3.1.	La P-gp : entre transport de substrats et hydrolyse d'ATP.....	21
II.3.2.	La P-gp : un modèle de pompe atypique?.....	22
II.3.3.	La P-gp : sa multispécificité de reconnaissance.....	25
II.3.4.	La P-gp : le phénotype MDR.....	30
II.4.	La P-gp d'un point de vue structural.....	31
II.4.1.	Structure du gène ABCB1.....	31
II.4.2.	Structure de la protéine P-gp.....	31
III.	Objectifs de la thèse.....	47
IV.	Bibliographie.....	48

## Chapitre B: Etude phylogénétique

I.	Introduction.....	54
II.	Analyse des séquences protéiques.....	54
III.	Alignement des séquences protéiques:.....	56
IV.	Analyses phylogénétiques.....	58
IV.1.	Méthodologie.....	58
IV.2.	Phylogénie des transporteurs ABC humains.....	60
IV.3.	La phylogénie au service de la modélisation.....	62
V.	Bibliographie.....	65

## Chapitre C: Modélisation Moléculaire

I.	Introduction.....	67
II.	Etude structurale de transporteurs ABC humains .....	68
II.1.	Contexte scientifique .....	68
II.2.	Analyse de la topologie membranaire des transporteurs ABC .....	70
II.2.1.	ABCB1 et ABCC8 : des topologies membranaires similaires .....	70
II.2.2.	Quelle topologie pour la région linker ?.....	73
II.3.	Modélisation comparative de transporteurs ABC humains .....	77
II.3.1.	Sélection des templates et alignements de séquences .....	77
II.3.3.	Evaluation des modèles.....	82
II.4.	Etude du cycle de translocation de la P-gp .....	143
II.4.1.	Analyse des cavités et des canaux d'accès de la P-gp.....	143
II.4.2.	Comparaison de deux conformations .....	146
III.	Conclusion .....	150
IV.	Bibliographie.....	152

## Chapitre D: Docking

I.	Introduction.....	154
II.	Le processus de docking.....	155
II.1.	Présentation .....	155
II.2.	L'étape de docking .....	156
II.2.1.	La recherche systématique .....	156
II.2.2.	La recherche stochastique.....	156
II.2.3.	La recherche déterministe .....	157
II.3.	L'étape de scoring .....	157
III.	Etude de docking chez les transporteurs ABC multispécifiques .....	158
III.1.	Contexte scientifique .....	158
III.2.	Les protagonistes considérés .....	160
III.2.1.	Le récepteur .....	160
III.2.2.	Les ligands .....	161
III.3.	Résultats du docking .....	164
III.3.1.	Résultats du docking sur la P-gp de hamster .....	164
III.3.2.	Résultats du docking sur la P-gp humaine .....	177

IV. Conclusion .....	182
V. Bibliographie.....	183

### Bilan général et perspectives

I. Bilan général.....	185
II. Perspectives.....	188

### Annexes.

.....	189
-------	-----

# Introduction générale

# Chapitre A : Introduction générale

---

## I. Les transporteurs ABC

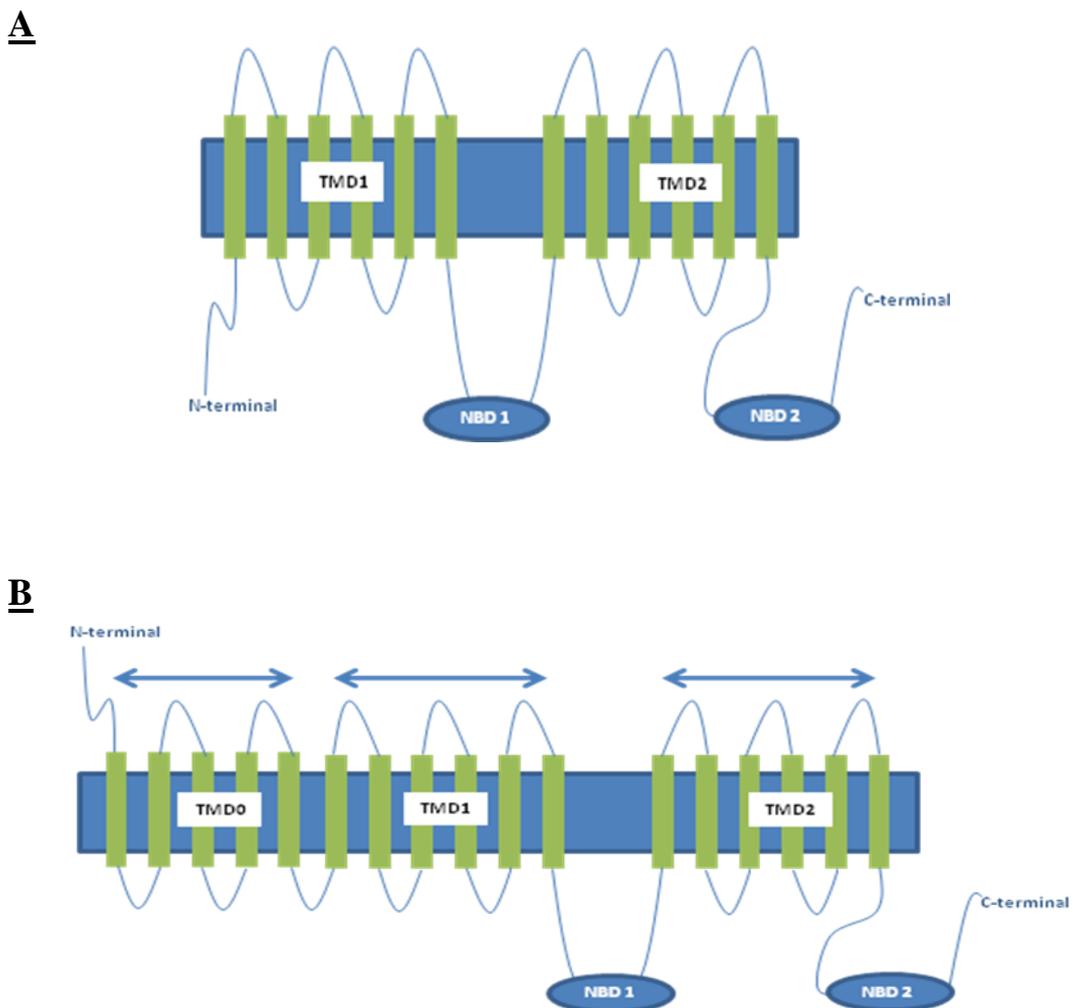
Toutes les cellules et les compartiments subcellulaires sont séparés du milieu extérieur par des membranes lipidiques. La survie cellulaire nécessite un passage régulé et sélectif de molécules spécifiques à travers ces membranes, non seulement pour acquérir les substances nutritives et excréter de nombreux produits mais aussi pour une multitude d'autres fonctions. En effet, certaines protéines membranaires sont des canaux ioniques ou des transporteurs passifs autorisant le passage de constituants selon leur gradient électrochimique ou des systèmes de transport actif permettant le transport de substances contre leur gradient de concentration. Parmi les transporteurs actifs, la famille des transporteurs à « ATP-binding cassette » (ABC) est la plus large avec des représentants chez les procaryotes et les eucaryotes [1]. Plus de 200 protéines impliquées dans le transport des substrats à travers les membranes biologiques sont des membres de la superfamille des protéines ABC [2]. Les eucaryotes ne possèdent que des systèmes d'export, alors que les procaryotes possèdent à la fois des systèmes d'export et d'import. La nature des composés transportés est extrêmement variable. A ce jour, 51 gènes ont été décrits dans cette famille chez l'Homme. 48 d'entre eux sont fonctionnels, 2 sont des pseudogènes alors que, pour un gène, la séquence de référence a été retirée des banques de données [3]. Les analyses phylogénétiques menées par Dean et al., ont permis de répartir cette famille en sept sous-familles, allant de la sous-famille A à la sous-famille G (Tableau-A-1) [4]. Il faut noter qu'une petite partie des protéines ABC est composée de protéines solubles, et ne sont donc pas des transporteurs membranaires.

Sous famille	Gènes	Alias	Locus
<b>ABCA</b>	ABCA1	ABC1	9q31.1
	ABCA2	ABC2	9q34
	ABCA3	ABC3	16p13.3
	ABCA4	ABCR	1p22.1-p21
	ABCA5	ABC13	17q24.3
	ABCA6		17q24.3
	ABCA7		19p13.3
	ABCA8		17q24
	ABCA9		17q24
	ABCA10		17q24
	ABCA11	pseudogène	4p16.3
	ABCA12		2q34
	ABCA13		7p12.3
<b>ABCB</b>	ABCB1	MDR1, P-gp	7q21.1
	ABCB2	TAP1	6p21.3
	ABCB3	TAP2	6p21.3
	ABCB4	PGY3, MDR2/3	7q21.1
	ABCB5		7p15.3
	ABCB6	MTABC3	2q36
	ABCB7	ABC7	Xq12-q13
	ABCB8	MABC1	7q36
	ABCB9		12q24
	ABCB10	MTABC2	1q42
	ABCB11	SPGP, BSEP	2q24
<b>ABCC</b>	ABCC1	MRP1	16p13.1
	ABCC2	MRP2	10q24
	ABCC3	MRP3	17q22
	ABCC4	MRP4	13q32
	ABCC5	MRP5	3q27
	ABCC6	MRP6	16p13.1
	ABCC7	CFTR	7q31.2
	ABCC8	SUR1	11p15.1
	ABCC9	SUR2	12p12.1
	ABCC10	MRP7	6p21.1
	ABCC11	MRP8	16q12.1
	ABCC12	MRP9	16q12.1
	ABCC13	PRED (pseudogène)	21q11.2
<b>ABCD</b>	ABCD1	ALD	Xq28
	ABCD2	ALD1, ALDR	12q11-q12
	ABCD3	PXMP1, PMP70	1p22-p21
	ABCD4	PMP69, P70R	14q24.3
<b>ABCE</b>	ABCE1	OABP, RNS41	4q31
<b>ABCF</b>	ABCF1	ABC50	6p21.33
	ABCF2		7q36
	ABCF3		3q27.1
<b>ABCG</b>	ABCG1	ABC8, WHITE1	21q22.3
	ABCG2	MXR, BCRP	4q22
	ABCG3	Séquence retirée	
	ABCG4	WHITE2	11q23.3
	ABCG5	Sterolin-1	2p21
	ABCG8	Sterolin-2	2p21

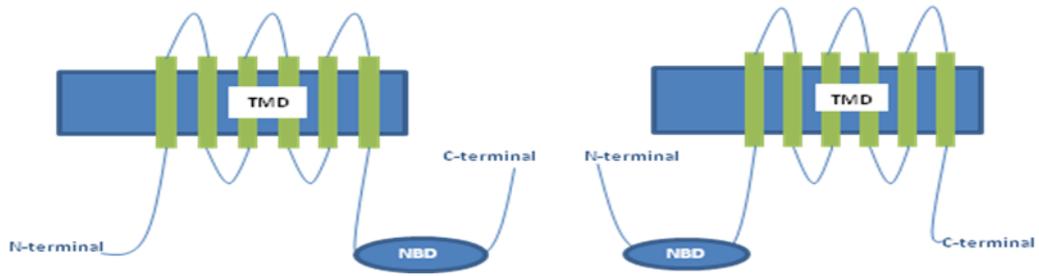
**Tableau-A-1 : Liste des sous-familles des ABC humains, leurs gènes, leurs alias et leurs loci chromosomaux [4]**

## I.1. Organisation structurale des transporteurs ABC

Les transporteurs ABC présentent une remarquable conservation dans leur séquence et dans leur organisation structurale. Une protéine est classée en tant que membre de la famille ABC lorsqu'elle possède un domaine ATP-binding cassette. Ces domaines sont distinctifs, et conservés au sein de la famille des transporteurs ABC. L'unité fonctionnelle d'un ABC transporteur entier contient deux domaines transmembranaires (TMD pour *TransMembrane Domain*) et deux domaines cytoplasmiques (NBD pour *Nucleotide-Binding Domain*). Les héli-transporteurs, comportant un domaine TMD et un domaine NBD, doivent former un homodimère ou un hétérodimère pour constituer un transporteur fonctionnel (Figure-A-1). Certains transporteurs ABC comportent un autre domaine TMD additionnel du côté N-terminal, appelé TMD0 (voir Figure-A-1).



**C**

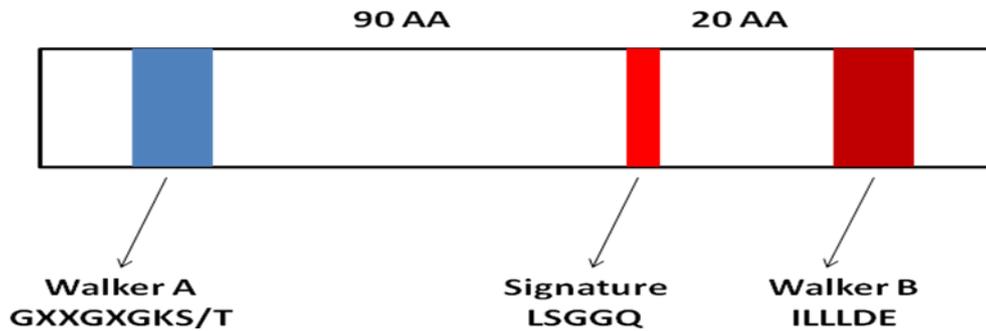


**Figure-A-1 : Organisation schématique des transporteurs ABC.** A) Un transporteur typique ABC possède 4 domaines : 2 domaines transmembranaires (en vert) et 2 domaines cytoplasmiques liant l'ATP (en bleu). Exemple : ABCB1, ABCB11, ABCC4. B) Un transporteur entier peut posséder un domaine additionnel TMD0 en amont à TMD1. Exemple : ABCC1, ABCC2, ABCC3. C) Un hémis-transporteur ABC possède 2 domaines : 1 domaine transmembranaire (en vert) et un domaine cytoplasmique (en bleu). Exemple : ABCD, ABCG

Les TMD comportent plusieurs hélices transmembranaires. Typiquement, il y a six hélices par domaine, soit un total de douze hélices pour un transporteur entier (mais il existe des exceptions, comme BtuCD). Les TMDs constituent les domaines hydrophobes d'un transporteur, et ils ont un rôle primordial dans la reconnaissance et le transport du substrat. L'enchaînement est le plus souvent N-TMD-NBD-C, mais le contraire existe aussi (exemple : ABCG).

Les domaines cytoplasmiques ou NBD sont des domaines hydrophiles conservés. Ce sont les domaines impliqués dans la liaison et l'hydrolyse de l'ATP. Les NBD de tous les membres de la superfamille des protéines ABC partagent trois motifs caractéristiques qui permettent de les identifier par analyse de séquences. Ces motifs forment la « cassette » de fixation de l'ATP (Figure-A-2) :

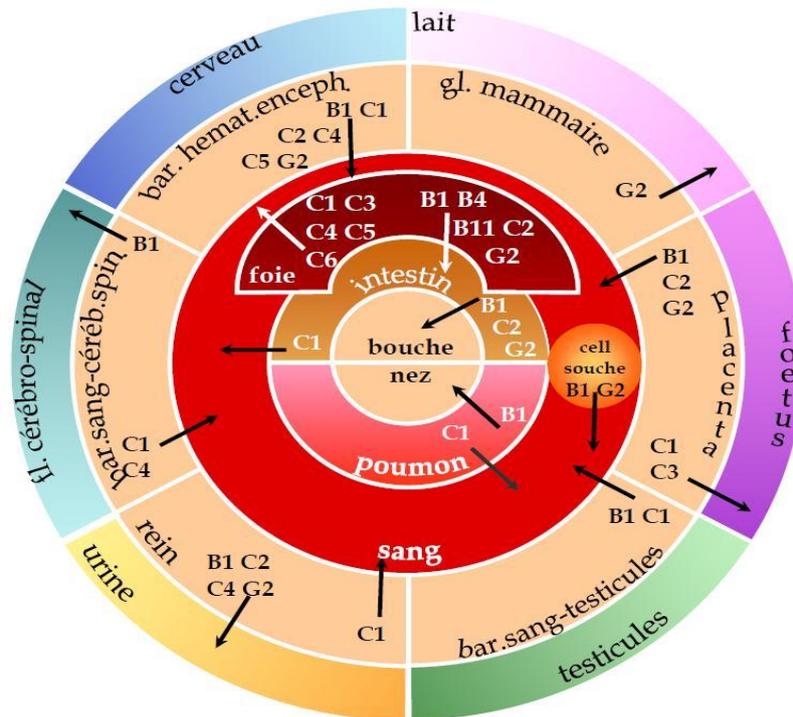
- Les motifs de Walker A et B [5]: ce sont des courtes séquences en acides aminés. Ces motifs sont conservés chez certaines protéines fixant l'ATP, en particulier les transporteurs ABC, mais ce ne sont pas les seules : il y a aussi certaines protéines kinases ainsi que des protéines de signalisation (exemple : HRAS).
- Le motif « LSGGQ » constitue la « signature » spécifique des transporteurs ABC. Sa présence entre les deux motifs Walker A et B sert à déterminer l'appartenance d'une protéine à la superfamille des transporteurs ABC.



**Figure-A-2 : Organisation des domaines NBD.** Une séquence consensus des domaines NBD d'une protéine ABC. Les motifs Walker A et B sont distants d'environ 120 acides aminés. La signature ABC se situe à une vingtaine d'acides aminés en amont du motif Walker B. Les acides aminés les plus conservés sont indiqués.

## I.2. Les transporteurs ABC et leurs rôles dans les maladies génétiques

Les transporteurs ABC sont donc des protéines membranaires qui transportent une large variété de substrats à travers les membranes intra- et extra- cellulaires. Cette fonction de transport leur permet d'assurer, chez les mammifères, un rôle important des transporteurs ABC dans la détoxification et la protection de l'organisme contre les xénobiotiques. Ce rôle prépondérant, dans les études récentes, est révélé par la distribution tissulaire des transporteurs ABC. Ils sont en effet exprimés au niveau d'importantes barrières tissulaires, telles que la muqueuse intestinale et l'endothélium qui constitue la barrière hémato-encéphalique, mais aussi d'épithéliums d'élimination, comme les membranes apicales du canicule biliaire des hépatocytes et du tubule rénal [6] (Figure-A-3).



**Figure-A-3 : Rôle pharmacologique des transporteurs ABC [6]**

A partir de la découverte initiale que la résistance multiple aux médicaments anticancéreux (MDR, pour « multidrug resistance ») était due à l'expression d'un transporteur ABC [7] (ABCB1 ou P-gp), il est devenu clair que plusieurs maladies chez l'homme résultent de mutations observées chez les transporteurs ABC. Ainsi, on associe à ce jour à 22 des 48 transporteurs connus des pathologies observées chez l'Homme [8], <http://nutrigene.4t.com/humanabc.htm>, [9]. Les mutations dans les transporteurs ABC causent ou contribuent à plusieurs désordres, mendéliens ou complexes, incluant l'adrénoleucodystrophie, la mucoviscidose, la dégénérescence rétinienne, et certaines formes d'hypercholestérolémie et de cholestase (Tableau-A-2). La maladie génétique la plus fréquente est la mucoviscidose, causée par des mutations observées chez la protéine CFTR/ABCC7.

Nom	Fonction	Maladies associées	Phénotype
ABCA1	Transport de cholestérol et de phospholipides	Maladie de Tangier	Athérosclérose Déficiency en lipoprotéines à haute densité (HDL)
ABCA2	Résistance aux médicaments	Maladie d'Alzheimer	Troubles neurodégénératifs
ABCA3	Résistance aux xénobiotiques	Détresse respiratoire néonatale	Déficiency en surfactant pulmonaires
ABCA4	Transport des rétinoïdes	Maladie de Stargadt Rétinopathie pigmentaire Dégénérescence maculaire	Baisse de l'acuité visuelle
ABCA7	Non connue	Syndrome de Sjögren	Sécheresse oculaire et buccale
ABCA12	Non connue	Ichtyose Harlequin Ichtyose lamellaire de type II	Hyperkératose (tégument crouteux), troubles respiratoires, infectieux ou alimentaires Peau sèche, accumulation des squames
ABCB1	Résistance aux médicaments Protection contre les xénobiotiques hydrophobes	Résistance multiple des cellules cancéreuses aux médicaments cytotoxiques	Echec de la chimiothérapie anticancéreuse
ABCB2	Transport des peptides	Déficiency immunitaire	
ABCB3	Transport des peptides	Déficiency immunitaire	Vascularisation nécrosante
ABCB4	Transport d'acide biliaire Transport de la phosphatidylcholine	Cholestase intrahépatique familiale progressive de type 3	Insuffisance hépatique
ABCB7	Transport de fer	Anémie sidéroblastique liée à l'X	Troubles neurologiques
ABCB11	Transport d'acides biliaires	Cholestase intrahépatique familiale progressive de type 2	Insuffisance hépatique
ABCC1	Résistance aux médicaments Protection contre les xénobiotiques hydrophobes	Résistance multiple des cellules cancéreuses aux médicaments	Echecs de la chimiothérapie anticancéreuse
ABCC2	Efflux d'anion organique	Syndrome de Dubin-Johnson	Ictère bénin chronique Hyperbilirubinémie
ABCC6	Non connue	Pseudoxanthome élastique	Pathologie héréditaire du tissu conjonctif Calcification et fragmentation des fibres élastiques (peau, rétine, parois artérielles)
ABCC7	Canal ion chlorure	Mucoviscidose	Troubles respiratoires (inflammation chronique des bronches) Troubles digestifs (pancréas, intestin)
ABCC8	Récepteur aux sulfonylurées	Hyperinsulinisme congénital de l'enfant	Hypoglycémie
ABCC9	Régulation du canal potassique à ATP	Cardiomyopathie dilatée avec tachycardie ventriculaire	Insuffisance cardiaque
ABCD1	Transport intracellulaire des acides gras à très longues chaînes	Adrénoleucodystrophie	Démyélinisation progressive du système nerveux central et insuffisance surrénale
ABCD3		Syndrome de Zellweger (type 2)	Syndrome cérébro-hépatorenal Leucodystrophie Absence de peroxyosomes
ABCG2	Résistance aux médicaments MDR	Protoporphyrémie érythropoïétique	Photosensibilité aiguë
ABCG5	Transport des stérols	Sitostérolémie	Hypercholestérolémie et artériosclérose prématurée
ABCG8	Transport des stérols	Sitostérolémie	Hypercholestérolémie et artériosclérose prématurée

**Tableau-A-2 : Maladies associées aux transporteurs ABC.** Pour chaque transporteur ABC, est indiqué sa fonction, les maladies auxquelles il a été associé, ainsi que les phénotypes observés.

Les transporteurs ABC assurent le transport de très nombreux substrats. Ces substrats peuvent être des substances physiologiques (peptides, hormones, sucres, ions...) ou des substances exogènes potentiellement toxiques (métaux lourds, contaminants...). Dans ce dernier cas, le transport de ces substances reflète le rôle clé des transporteurs ABC dans la détoxification de l'organisme chez les mammifères. Certains transporteurs ABC sont impliqués dans les phénomènes de résistance multiple aux médicaments anticancéreux (MultiDrug Resistance). C'est le cas pour certaines protéines ABC des sous-familles B (B1, P-gp), C (MRP1, MultiDrug Resistance-associated Protein 1) et G (G2, MXR/BCRP), dont l'expression au niveau des cellules cancéreuses a été associée à une fonction de détoxification par efflux actif des médicaments, pouvant mettre en échec les chimiothérapies anticancéreuses [10] (Tableau-A-3).

<b>Protéine</b>	<b>Nom</b>	<b>Médicaments Chimio-thérapeutiques expulsés par le transporteur commun</b>
<b>ABCA2</b>	ABC2	Estramustine
<b>ABCB1</b>	MDR1, P-gp	Colchicine, Doxorubicine, Etoposide, Vinblastine, Paclitaxel
<b>ABCC1</b>	MRP1	Doxorubicine, Daunomycine, Etoposide, Colchicine, Méthotrexate
<b>ABCC2</b>	MRP2	Vinblastine, Cisplatine, Doxorubicine, Méthotrexate
<b>ABCC3</b>	MRP3	Méthotrexate, Etoposide
<b>ABCC4</b>	MRP4	6-mercaptopurine et 6-thioguanine, Méthotrexate
<b>ABCC5</b>	MRP5	6-mercaptopurine et 6-thioguanine
<b>ABCC6</b>	MRP6	Etoposide
<b>ABCC11</b>	MRP8	5-fluorouracil
<b>ABCG2</b>	BCRP	Mitoxantrone, Topotécan, Doxorubicine, Daunomycine, Méthotrexate, Irinotécan, Imatinibe

**Tableau-A-3 : Les transporteurs ABC impliqués dans la résistance aux médicaments [10]**

### **I.3. Les sous-familles de transporteurs ABC**

En se basant sur leur conservation de séquences, et plus spécialement celles de leurs domaines NBD qui comportent les motifs spécifiques, les gènes ABC eucaryotes ont été répartis en sept sous-familles:

### ***La sous-famille ABCA***

La sous-famille ABCA comprend 12 transporteurs entiers. Cette sous-famille contient les plus grands gènes ABC. Ainsi, certaines protéines de cette sous-famille sont composées de plus de 2000 acides aminés et ont une masse moléculaire supérieure à 200kDa. Les membres de cette sous-famille sont impliqués dans le transport des lipides. ABCA1 est impliqué dans le transport de cholestérol vers les lipoprotéines à haute densité (HDL) naissantes, mais qui est décrit comme étant secondaire à une fonction de transfert intramembranaire de phosphatidylsérine. Des mutations au niveau de ce gène ont été associées à une maladie héréditaire, la maladie de Tangier, se traduisant par des désordres de transport du cholestérol et des HDL. [11-12]. C'est aussi le cas pour d'autres membres de la sous-famille ABCA : la détresse respiratoire néonatale pour ABCA3 [13], la maladie d'Alzheimer pour ABCA2 [14-15], la maladie de Stargardt pour ABCA4, et l'ichtyose Harlequin pour ABCA12.

### ***La sous-famille ABCB***

C'est la seule sous-famille chez les mammifères comportant à la fois des transporteurs entiers et des hémi-transporteurs. Actuellement, 11 membres constituent cette sous-famille dont 4 transporteurs entiers et 7 hémi-transporteurs. ABCB1 est le premier transporteur cloné ; il est aussi le premier transporteur caractérisé pour sa capacité à conférer le phénotype « MultiDrug Resistance » aux cellules cancéreuses.[16-18]. Nous reviendrons plus en détail dans cette introduction sur les propriétés de ce transporteur, sur lequel porte l'essentiel de mon travail de thèse.

Les protéines ABCB4 et ABCB11 sont toutes les deux localisées dans le foie et sont impliquées dans la sécrétion des composants de la bile, phosphatidylcholine et acides biliaires respectivement [8]. ABCB2 et B3 (TAP1/2) sont deux hémi-transporteurs qui dimérisent pour transporter des peptides de restriction antigénique à l'intérieur du réticulum endoplasmique, pour réaliser leur présentation antigénique aux cellules immunocompétentes en association aux molécules de classe 1 du complexe majeur d'histocompatibilité [19]. Le transporteur ABC le plus homologue aux TAP est ABCB9, et il a été localisé dans les lysosomes [20]. Les autres hémi-transporteurs, ABCB6, ABCB7, ABCB8 et ABCB10 ont été localisés dans les mitochondries où ils participent au le métabolisme du fer et transportent les peptides précurseurs de la protéine Fe/S [21].

Les maladies génétiques associées aux transporteurs de cette sous-famille sont: les cholestases intrahépatiques familiales de type 2 ou 3 dues respectivement à des mutations sur

ABCB11 [22] ou ABCB4 [23] et l'anémie sidéroblastique liée à l'X, due à des mutations d'ABCB7.

### *La sous-famille ABCC*

La sous-famille ABCC contient 12 transporteurs entiers avec des fonctions assez variées parmi lesquelles le transport d'ions, des activités de sécrétion de toxines ou encore, le rôle de récepteurs à la surface des cellules.

ABCC1 est le principal transporteur de conjugués du glutathion et du glucuronate ; c'est aussi l'un des transporteurs responsables d'efflux de médicaments cytotoxiques, cotransportés avec le glutathion. ABCC2 transporte les anions organiques du foie vers la bile, les mutations touchant ce transporteur sont responsables du syndrome de Dubin-Johnson [24]. Comme ABCC1 et C2, ABCC3 est aussi un transporteur de conjugués du glutathion alors qu'ABCC4 et ABCC5 sont des transporteurs de nucléotides cycliques et d'analogues de nucléotides [25]. Les trois membres de cette sous-famille, ABCC10, ABCC11 et ABCC12 sont également des pompes d'anions lipophiles. Un dernier membre de cette famille ABCC13 ne semble pas coder pour un transporteur fonctionnel chez l'Homme [4].

Le transporteur le plus étudié de cette sous-famille est ABCC7 ou CFTR. C'est un canal d'ions chlorures, régulé par les nucléotides, dont le gène muté est responsable chez l'Homme de la mucoviscidose [26]. ABCC8 (SUR1) et ABCC9 (SUR2) sont deux protéines membranaires ABC encore plus atypiques puisqu'elles ne présentent aucune fonction de transport connue à ce jour. ABCC8 a représenté un modèle de travail pour une partie de mes travaux de thèse présentés dans ce manuscrit. ABCC8/9 jouent le rôle de régulateurs de canaux potassiques impliqués dans la modulation de la sécrétion d'insuline. Des mutations d'ABCC8 causent une hypoglycémie due à l'excès de sécrétion d'insuline. Les protéines SUR sont en association hétérotétramérique avec les sous-unités Kir6 qui constituent le canal  $K^+$  proprement dit : les protéines ABC qui sont réparties autour du canal  $K_{ATP}$  dont la conductance est régulée par ATP et ADP (par liaison aux NBD) mais aussi par différents ligands pharmacologiques dont la reconnaissance multispécifique est une caractéristique importante de SUR.

### ***La sous-famille ABCD***

Cette sous-famille contient 4 hémi-transporteurs péroxisomaux [27] qui fonctionnent en homo et/ou hétérodimères pour réguler le transport des acides gras à très longues chaînes. Les mutations dans le gène ABCD1 ont été observées chez les personnes présentant une adrénoleucodystrophie lié à l’X [28].

### ***La sous-famille ABCE et ABCF***

Il s’agit des seules sous-familles composées de protéines dépourvues de domaines transmembranaires et n’ayant aucune fonction de transport membranaire. ABCE ne comporte qu’un seul membre : ABCE1, il a été décrit comme essentiel pour l’assemblage de la capsid du virus VIH [29], comme inhibiteur de la ribonucléase RNaseL [30] et assez récemment comme un intervenant dans l’initiation de la traduction [31]. La sous-famille ABCF comporte quant à elle trois membres, pour lesquels aucune fonction n’est connue à ce jour. Cependant, une surexpression d’ABCF2 suite à un traitement par chimiothérapie a été observée chez certains patients [32-33].

### ***La sous-famille ABCG :***

La sous-famille ABCG est composée de 5 hémi-transporteurs « inversés », c’est à dire ayant le NBD du côté N-terminal et le TMD du côté C-terminal (voir Figure-A-1). Les membres de cette sous-famille fonctionnent en homodimères (ABCG1, ABCG2 et ABCG4) ou en hétérodimères (ABCG5 et ABCG8). Depuis son clonage il y’a plus de dix ans par Doyle *et al.* (1998), ABCG2 est devenu l’un des transporteurs ABC les plus étudiés. Il est considéré aujourd’hui, avec ABCB1 et ABCC1, comme l’un des principaux transporteurs responsables de la résistance multiple aux médicaments au-delà de son rôle initialement attribué : BCRP pour « Breast Cancer Resistance Protein ». En outre, ABCG2 est responsable du phénotype de « side population » des cellules souches de différentes origines, en relation avec sa capacité à protéger ces cellules fragiles et précieuses contre des toxiques divers. La fonction d’ABCG4 demeure spéculative, mais elle semble compléter celle d’ABCG1 qui consiste au transfert de cholestérol de cellules périphériques (typiquement les macrophages) aux HDL matures. Quant à celle d’ABCG5 et d’ABCG8, elle est bien décrite, c’est le transport des stérols au niveau intestinal et biliaire, en particulier conduisant à l’exclusion de l’organisme des phytostérols apportés par l’alimentation, dont un excès est toxique ; des

mutations causant la perte de cette fonction sont impliquées dans la sitostérolémie due à l'accumulation des phytostérols.

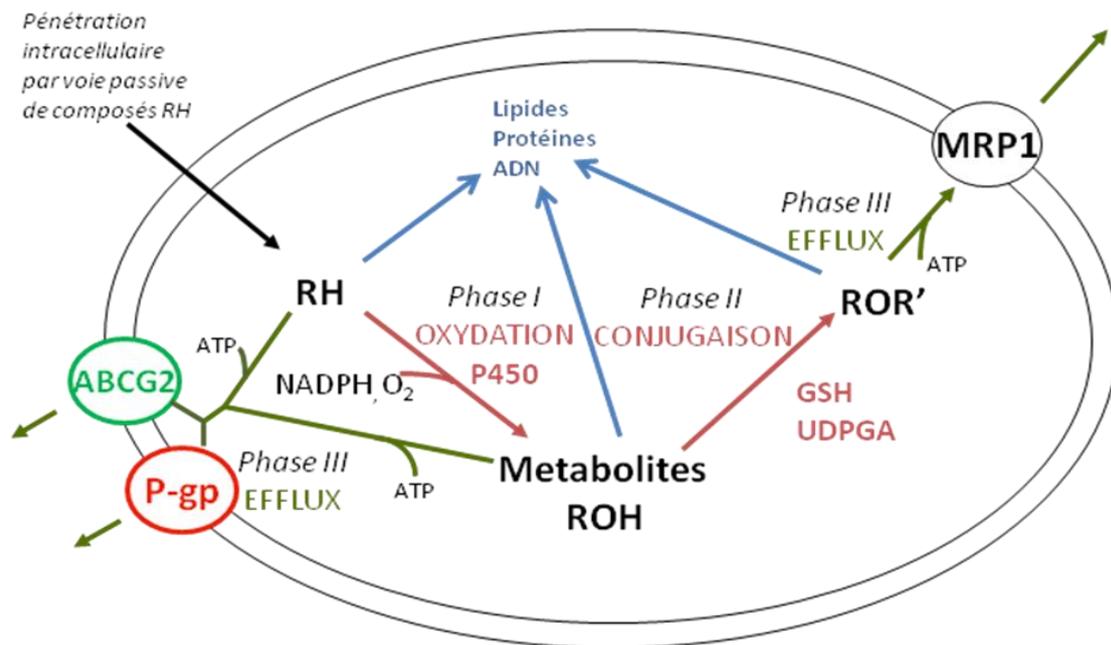
## **II. ABCB1 : La P-glycoprotéine ou P-gp**

### **II.1. Généralités : MDR et Détoxication**

Appelée aussi MDR1, la P-gp est le premier transporteur ABC découvert [16], et cette découverte est à l'origine de l'intérêt porté à l'étude des transporteurs ABC. Historiquement, la P-gp a été mise en évidence à partir de l'étude *in-vitro* d'un type particulier de résistance cellulaire à des médicaments cytotoxiques, le phénotype de résistance multiple aux médicaments (MDR). Il s'agit par définition de la résistance croisée à différents médicaments cytotoxiques, ne présentant pas d'analogie de structure ni de mode d'action ou de métabolisation, distinctes que celle qui a induit la résistance observée. Il a ainsi été constaté dans diverses lignées chimiorésistantes, obtenues par sélection progressive par augmentation de la concentration tolérée d'un médicament cytotoxique considérée, une diminution de la concentration intracellulaire des médicaments anticancéreux en comparaison avec les lignées parentales sensibles initiales [34]. Cette diminution peut avoir plusieurs origines : une diminution du flux entrant (influx), une augmentation du flux sortant (efflux), ou bien une combinaison de ces deux mécanismes. Les différentes études menées ont conduit à suggérer l'existence d'une pompe transmembranaire énergie-dépendante à l'origine de l'augmentation de l'efflux cellulaire et, par conséquent, à l'origine de la résistance pléiotropique: la première pompe décrite fut ainsi la P-gp en 1976 [16]. L'acquisition de la résistance à la colchicine dans des cellules tumorales était due à la P-gp modifiant la perméabilité membranaire à ce médicament et la surexpression de la P-gp est suffisante pour conférer le phénotype MDR. Ce phénotype est responsable d'environ 50% de l'ensemble des échecs thérapeutiques secondaires de la chimiothérapie [35]. La cellule tumorale s'adapte pour se protéger des cytotoxiques en augmentant l'expression de protéines de transport MDR, par des mécanismes soit de sélection soit d'induction. En effet, le niveau d'expression de cette protéine est d'autant plus élevé que les cellules MDR sont résistantes. La surexpression de la P-gp se traduit par une expulsion rapide, hors du cytoplasme de ces cellules MDR, des molécules cytotoxiques considérées dont la cible létale est intracellulaire. Plusieurs études ont

mis en évidence une augmentation significative de la concentration intracellulaire des médicaments dans les cellules MDR en présence du 2-déoxyglucose, inhibiteur de la production d'ATP [36-38]. En revanche, en présence de glucose, l'activité du transporteur est rétablie et la concentration intracellulaire de l'agent anticancéreux diminue dans ces cellules [39]. C'est ainsi que la P-gp a été identifiée comme le transporteur actif d'un grand nombre de composés ayant des structures chimiques très variées [18, 40].

Parallèlement, il a aussi été constaté que l'expression de la P-gp ne se limite pas uniquement aux cellules tumorales. La P-gp joue de manière générale un rôle important dans les mécanismes de protection de l'organisme contre les xénobiotiques potentiellement toxiques (poisons de l'environnement, médicaments) et, pour cela, la P-gp est également présente physiologiquement au niveau de nombreux organes et tissus sains. Elle participe au processus général de détoxication cellulaire et de protection de l'organisme en synergie avec d'autres enzymes (Figure-A-4).



GSH: Glutathion; UDPGA: Acide uridine diphosphoglucuronique, R': Sucre, Sulfate, Acetate, Méthyl

**Figure-A-4 : Schéma simplifié du métabolisme de détoxication des composés hydrophobes (Source : Cours de M. Delaforge).**

Ce schéma simplifié du processus de détoxication implique des enzymes différentes classées selon la phase de détoxication en trois catégories. Les enzymes de phase I, ou phase

de fonctionnalisation, sont majoritairement constituées des cytochromes P450. Ces enzymes se chargent de la mono-oxygénation (« oxydation ») des xénobiotiques ne pouvant être éliminés directement. Les enzymes de phase II, ou phase de conjugaison, comme la glutathion S-transférase, permettent de former des adduits covalents de groupements hydrophiles sur les xénobiotiques préalablement transformés par les cytochromes ou non. Enfin, les enzymes de phase III, dont font partie la P-gp, ABCG2 et la MRP1, tous les deux catalysent l'efflux des molécules transformées ou non par les enzymes des phases I et II [41]. Ce rôle de détoxification cellulaire proposé pour la P-gp est compatible avec ce qui est décrit précédemment concernant l'induction de l'expression de la P-gp lors de l'exposition prolongée de différentes lignées cellulaires à divers substrats de transport de la P-gp : elle doit être considérée comme une protéine inductible, de manière similaire aux P450, ce qui en fait une « protéine de stress » qui participe à la réponse cellulaire d'adaptation à une agression toxique [42].

## **II.2. Localisation tissulaire de la P-gp**

La P-gp a été mise en évidence dans de nombreux tissus sains comme le pôle apical des cellules épithéliales de la lumière intestinale où elle joue un rôle majeur dans la protection de l'organisme contre l'absorption intestinale des xénobiotiques apportés par l'alimentation. La P-gp est aussi présente au niveau des cellules du tubule proximal du rein, où elle favorise le processus d'élimination urinaire. Sa présence a aussi été mise en évidence dans les hépatocytes et dans les cellules endothéliales de la barrière hémato-encéphalique (BHE) [43-44]. Ces différents tissus participent au métabolisme des médicaments du point de vue de leur absorption, de leur et de leur élimination l'organisme.

Chez l'Homme, l'expression de la P-gp répond à une variabilité interindividuelle : par exemple, elle peut augmenter d'un facteur 8 au niveau entérocytaire et d'un facteur 2,4 au niveau hépatique chez des patients atteints de néoplasmes hépatiques [45]. Cependant, cette variabilité n'a jamais été confirmée sur d'autres tissus ou chez des sujets sains. Les facteurs présidant à cette variabilité (environnementaux, génétiques...) demeurent encore peu explorés, mais semblent parfois corrélés à des polymorphismes de type SNP (« single nucleotide polymorphism »).

Au niveau du système nerveux central (SNC), la P-gp a été identifiée sur la face luminale des cellules endothéliales de la BHE ainsi que sur les cellules endothéliales de la barrière hémato-méningée [46], mais c'est au niveau de la BHE que son rôle protecteur prend son importance, la surface vasculaire cérébrale de la BHE étant 5000 fois plus importante que la surface vasculaire cérébrale de la barrière hémato-méningée. Par son processus d'efflux, sa localisation et sa multispécificité de substrat, la P-gp protège le cerveau vis-à-vis de molécules potentiellement neurotoxiques comme cela a été démontré pour l'ivermectine ou la bromocriptine, ou pouvant provoquer des lésions irréversibles telles la 1-méthyl-4-phényl\_1,2,3,6-tétrahydropyridine (MPTP) par apoptose des neurones dopaminergiques (Tableau-A-4) [47].

<b>Organe</b>	<b>Localisation</b>	<b>Fonction</b>
<b>Intestin grêle, colon</b>	Pôle apical des cellules épithéliales	Sécrétion des xénobiotiques dans la lumière intestinale
<b>Rein</b>	Pôle apical des cellules du tube contourné proximal	Sécrétion des xénobiotiques dans l'urine
<b>Foie</b>	Membrane du canalicule biliaire des hépatocytes	Sécrétion des xénobiotiques dans la bile
<b>SNC</b>	Face luminale des cellules endothéliales de la BHE	Protection du SNC
<b>Placenta</b>	Trophoblastes	Protection du fœtus
<b>Cœur</b>	Membranes des cellules endothéliales myocardiaques	Protection du cœur
<b>Surrénales</b>	Surface des cellules de la medulla et du cortex	Sécrétion des stéroïdes endogènes
<b>Testicules</b>	Cellules endothéliales des capillaires sanguins = Barrière hémato-testiculaire	Protection des spermatozoïdes

**Tableau-A-4 : La localisation tissulaire de la P-gp et son rôle dans le transport des médicaments et des substances endogènes [47]**

### **II.3. La P-gp d'un point de vue fonctionnel**

Le développement de la résistance multiple aux médicaments constitue un obstacle majeur à l'efficacité du traitement par chimiothérapie de nombreux cancers humains, incluant les cancers de colon, du foie, les leucémies ainsi que les cancers pédiatriques. Ces tumeurs résistantes présentent la caractéristique d'une résistance croisée, mais bien définie, vis-à-vis d'un large spectre de médicaments cytotoxiques sans relations structurales. Il faut noter que le niveau de résistance des cellules d'une tumeur clinique est très inférieur (un facteur 2 est déjà

suffisant pour éliminer un médicament du fait de sa toxicité potentielle pour le malade) à celui qu'on peut observer *in-vitro* sur des cellules en culture avec une forte pression de sélection (jusqu'à plusieurs centaines de fois). L'étude *in vitro* des cellules MDR montre que l'efflux des médicaments est dépendant de l'ATP cellulaire. Le type majeur des mécanismes MDR est lié à la surexpression de la P-gp. C'est la raison pour laquelle il constitue le paradigme pour un fonctionnement en tant que pompe ATP-dépendante d'efflux de substrats hydrophobes [2].

### II.3.1. La P-gp : entre transport de substrats et hydrolyse d'ATP

Plusieurs arguments expérimentaux sont en faveur d'une hypothèse selon laquelle l'hydrolyse de l'ATP fournirait de l'énergie permettant à la P-gp de jouer son rôle de transporteur actif. La P-gp possède une fonction ATPasique qui lui permet d'hydrolyser l'ATP intracellulaire. Cette fonction est mesurable même en absence de substrat, et elle est en général augmentée en sa présence. L'énergie générée par l'hydrolyse de l'ATP permet à la P-gp d'assurer la fonction de transport des substrats à travers la membrane cytoplasmique dans le sens inverse de leur diffusion passive. La preuve que l'hydrolyse de l'ATP est nécessaire et suffisante au transport a été apportée pour la première fois dans les systèmes bactériens d'import du maltose et de l'histidine en utilisant des vésicules membranaires [48]. La P-gp lie aussi des analogues de l'ATP tels que le <sup>32</sup>P-8-azido-ATP, le GTP, ainsi que les analogues non hydrolysables de l'ATP (AMP-PNP), mais pas l'ADP ni le ribose-5-phosphate [49]. Il a été prouvé que ces analogues non hydrolysables sont les ligands compétitifs de l'ATP dans sa liaison à la P-gp mais ne peuvent la substituer en tant que « fournisseur d'énergie » [50]. L'absence d'une compétition entre l'ATP (ou analogues) et les divers substrats de la P-gp a permis de suggérer l'existence de deux types de sites de liaison distincts, l'un de nature nucléotidique, l'autre dédié aux substrats [51]. Il a aussi été prouvé que la P-gp, comme les autres membres de la superfamille des protéines ABC, possède deux sites de liaison des nucléotides. Ces deux sites sont fonctionnels et peuvent hydrolyser l'ATP indépendamment et à la même vitesse [52]. En 1994, Loo et Clarke avaient déjà démontré qu'en dissociant les deux moitiés N- et C- terminales du gène codant pour la P-gp et en les exprimant séparément, une activité catalytique peut être observée [53]. L'influence de la liaison et/ou de l'hydrolyse de l'ATP sur les changements conformationnels qui peuvent être observés au niveau de la P-gp ont fait l'objet de plusieurs études. Des données biochimiques ont ainsi démontré que la liaison de l'ATP aux domaines NBD produit des changements conformationnels au niveau

des TMD [54]. Il a aussi été observé que ces changements ne sont pas de la même amplitude et/ou de la même nature en présence d'un substrat à transporter. On suppose l'existence de communications allostériques entre les NBDs et les TMDs provoquant une perturbation fonctionnelle se manifestant par la translocation transmembranaire des substrats à transporter, et correspondant souvent à une accélération de l'hydrolyse d'ATP [49].

L'introduction de mutations ponctuelles au niveau des domaines de liaison de l'ATP, et plus précisément des motifs Walker A (Figure-A-2), inhibe l'acquisition par le gène du phénotype de résistance multiple aux médicaments. Ceci démontre l'importance des domaines NBD dans l'activité biologique de la P-gp [55]. Cette même étude a prouvé que la liaison de l'ATP n'est pas affectée par ces mutations : l'étape altérée serait celle de l'hydrolyse de l'ATP ou bien des changements conformationnels induits.

Deux types d'activités ATPasiques ont été observées chez la P-gp : une activité dite basale, mesurée en absence de substrats exogènes et provoquée par la prise en charge par la P-gp de substrats hydrophobes endogènes [56], et une activité dite stimulée, mesurée en présence de substrats exogènes. Plusieurs études ont été menées pour mesurer l'activité basale de la P-gp à la surface des cellules entières et des vésicules. Ces études ont permis à Higgins et Gottesman de proposer pour la P-gp, un rôle de translocation (ou « floppase », vers l'extérieur de la cellule) de composés endogènes et notamment de certains lipides membranaires. En plus de son rôle dans la protection de l'organisme contre les xénobiotiques, la P-gp intervient également dans le transport du cholestérol [57]. Il a ensuite été montré une corrélation entre l'activité ATPasique basale de la P-gp et la redistribution active intramembranaire du cholestérol par la P-gp [58]. En fait, la P-gp participe aussi à la régulation de la distribution intramembranaire de la sphingomyéline et des phospholipides à courtes chaînes.

L'existence des activités ATPasiques, basale ou stimulée, corrélées à la présence de deux types de substrats endogènes ou exogènes, est une caractéristique qui distingue la P-gp des autres transporteurs membranaires actifs, et qui confirme sa large multispécificité enzymatique.

### II.3.2. La P-gp : un modèle de pompe atypique?

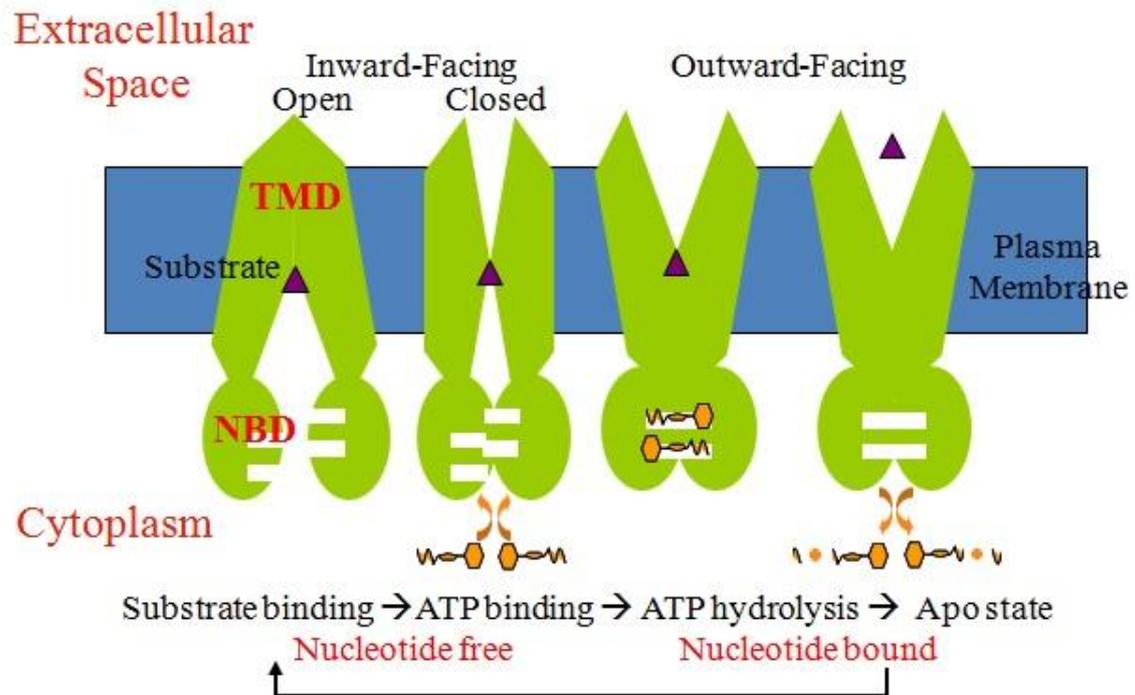
Plusieurs modèles ont été proposés pour décrire l'expulsion active de substrats réalisée par la P-gp. Les études menées sur la P-gp ont amené à conclure que le modèle de pompe

classique attribué aux transporteurs membranaires a fort probablement besoin d'être adapté dans le cas d'ABCB1. En effet, un transporteur membranaire « classique » tend à alterner entre une conformation avec le site de liaison du substrat accessible du côté cytosolique (*inward-facing*) et une conformation avec le site de liaison du substrat accessible du côté extracellulaire (*outward-facing*) [2]. Dans ce modèle, la liaison d'un substrat, en général unique, se fait à travers un site spécifique au sein du transporteur. Le substrat étant hydrophile, son interaction initiale avec son site de liaison se fait à partir de la phase aqueuse (le cytosol pour un exportateur), puis il est libéré au niveau de la phase aqueuse de l'autre côté de la membrane (le milieu extracellulaire). Le changement conformationnel induit par l'hydrolyse de l'ATP induit une libération du substrat de l'autre côté de la membrane. Les hélices transmembranaires sont supposées former une voie de transport hydrophile à travers la membrane, ce qui permet que les substrats ne soient en aucun cas en contact avec les lipides de la membrane. Ce modèle de pompe classique est largement soutenu pour les transporteurs de substrats hydrophiles, et il est également proposé pour les protéines ABC transportant des molécules polaires. Le succès de ce modèle ne permet néanmoins pas de le généraliser pour un transporteur ABC comme la P-gp. Même si ce modèle peut expliquer comment la P-gp exporte ses substrats (Figure-A-5), il est clair que plusieurs modifications doivent être faites :

i) Premièrement, à partir de la nature hydrophobe de la majorité des substrats de la P-gp, l'accès au site de liaison du substrat se fait à partir de la bicouche lipidique plutôt qu'à partir de la phase aqueuse. Rien n'empêche la libération du substrat de se faire également au niveau de l'environnement membranaire.

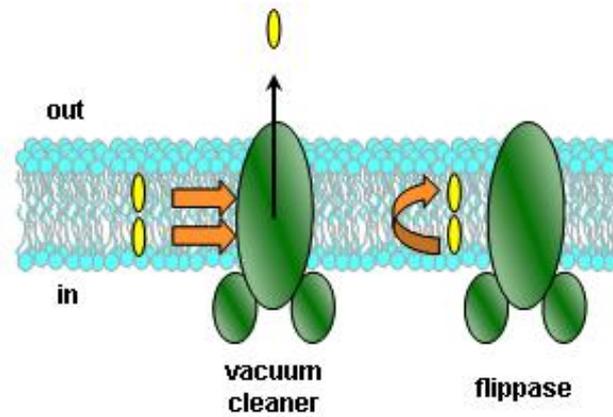
ii) Deuxièmement, la non-spécificité de la P-gp à un substrat donné nous éloigne d'un modèle avec un site de liaison bien spécifique des transporteurs membranaires classiques.

iii) Enfin, la nature lipophile des substrats doit être prise en compte pour comprendre certains aspects comme la voie de transport à travers la protéine, les cinétiques du transport, les gradients de concentration du substrat...



**Figure-A-5 :** Représentation schématique d'états conformationnels décrivant le cycle de translocation le plus simple imaginé pour la P-gp. En orange, les molécules d'ATP. En violet, le substrat.

En prenant en compte la nature « atypique » de la P-gp en tant que transporteur de substrats à majorité hydrophobe, deux modèles de pompe active semblent s'appliquer à la P-gp : le modèle « hydrophobic vaccum cleaner », proposé par Gottesman et Pastan et selon lequel le transport des composés se ferait directement de la membrane vers le milieu extracellulaire [40], et le modèle « floppase », proposée par Higgins et Gottesman, selon lequel le transport de composés se réaliserait du feuillet cytosolique à l'exoplasmique dans la membrane [59] (Figure-A-6). Ces deux modèles se confrontent depuis des années, et malgré de nombreux travaux, ils ne sont toujours pas départagés ; il est possible que le modèle effectif dépende de la nature du substrat.



**Figure-A-6 : Deux modèles alternatifs pour l'efflux actif de ses substrats transportés par la P-gp. Le substrat est représenté en jaune.**

### II.3.3. La P-gp : sa multispécificité de reconnaissance

Parmi les principales caractéristiques de l'étude de la P-gp, il y a sa grande variété de substrats transportés. En effet, la P-gp est capable de reconnaître des substrats de familles pharmacologiques, fonctions diverses et n'ayant aucune analogie structurale (Tableau-A-5) [35]. Ces substrats ont plutôt un caractère hydrophobe, de masse moléculaire variable allant de 250 Da (cimétidine) à 1250 Da (cyclosporine A, actinomycine D). Selon Ferté *et al.*, on distingue cinq grandes classes chimiques de molécules transportées par la P-gp [60]:

- Les bases faiblement lipophiles (méthotrexate)
- Les cations lipophiles (rhodamine-123, anthracyclines)
- Les composés neutres polycycliques (digoxine, stéroïdes)
- Les molécules amphiphiles (Triton X-100)
- Les peptides hydrophobes (cyclosporine A)

Les substrats de la P-gp sont souvent des substances médicamenteuses lipophiles. Ils sont à quelques exceptions près des substrats de l'isoenzyme 3A4 (CYP3A4) du cytochrome P450. [35, 61]. C'est pourquoi il est proposé que ces deux enzymes de détoxification aient des rôles complémentaires, voire synergiques.

Les substrats de la P-gp couvrent plusieurs classes thérapeutiques : nous trouvons notamment des agents cytotoxiques intercalants de l'ADN, des inhibiteurs de la topoisomérase II (doxorubicine, daunomycine, actinomycine D, étoposide), des poisons du fuseau mitotique (vinblastine, colchicine, taxol) ainsi que des immunosuppresseurs (cyclosporine A), des inhibiteurs calciques (vérapamil), des antiprotéases (indinavir). D'autres substrats sont présentés dans le Tableau-A-5.

<b>Classes thérapeutiques</b>	<b>Substrats</b>	<b>Inhibiteurs</b>	<b>Inducteurs</b>
<b>Cytostatiques</b>	Alcaloïdes de la Vinca Anthracyclines, Taxol Epipodophyllotoxines		Vinblastine, Doxorubicine
<b>Immunosuppresseurs</b>	Cyclosporine A Tacrolimus	Cyclosporine Valspodar	
<b>Antiarythmiques</b>	Amiodarone Propafénone, Digoxine	Amiodarone, Quinidine	
<b>Statines</b>	Simvastatine		
<b>Anticalciques</b>	Vérapamil, Mibéfradil	Vérapamil	
<b>Diurétiques</b>	Furosémide		
<b>Antihistaminiques H2</b>	Cimétidine, Ranitidine		
<b>Antibiotiques/ Macrolides</b>	Erythromycine, Clarithromycine	Erythromycine, Clarithromycine	
<b>Antifongiques imidazoles</b>	Kétaconazole	Kétaconazole	
<b>Antibiotiques/ antituberculeux</b>	Rifampicine		Rifampicine
<b>Antiépileptiques</b>	Phénytoïne		Phénytoïne
<b>Antiprotéases anti- VIH</b>	Saquinavir, Indinavir, Ritonavir	Ritonavir	
<b>Anthelminthiques</b>	Ivermectine		
<b>Antihistaminiques H1</b>	Terfénadine		
<b>Hormones stéroïdes</b>	Estrogènes, Aldostérone, Corticostéroïdes	Tamoxifène	Dexaméthasone
<b>Anti-émétiques</b>	Ondansétron		
<b>Antagonistes de dopamine</b>	Dompéridone		
<b>Antigoutteux</b>	Cochicine		
<b>Antidiarrhéiques</b>	Lopéramide		
<b>Analgésiques</b>	Morphine		
<b>Sédatifs, Antidépresseurs</b>	Midazolam	Midazolam	Millepertuis, Néfazodone
<b>Anti-PRL</b>	Bromocriptine		

**Tableau-A-5 : Exemples de substrats, inhibiteurs et inducteurs de la P-gp [35]**

\* La variabilité structurale des inducteurs est certes importante d'un point de vue biologique et pharmacologique, mais sans rapport avec la capacité de reconnaissance multispécifique de la P-gp elle-même.

\*\* La distinction entre « substrat » et « inhibiteur » fait référence à ce qui se passe au niveau de la cellule, où il est mesuré un flux net pour lequel intervient la vitesse du retour par diffusion passive (plus ou moins rapide selon leur hydrophobie) ; pour ce qui est du transporteur lui-même, d'un point de vue enzymologique, ce sont dans les deux cas des substrats reconnus et transloqués au travers de la membrane.

\*\*\* Il y a aussi la distinction importante du point de vue cellulaire entre drogues cytotoxiques et agents réversants, ou chimiosensibilisants (CS). Ces derniers sont des molécules transportées par la P-gp (le plus souvent « inhibiteurs » pour les cellules mais pas toujours) qui ne sont pas toxiques seuls pour les cellules mais qui inhibent l'efflux des drogues cytotoxiques, et qui ainsi leur reconfèrent leur efficacité. Les CS dits de première génération ont été des molécules déjà connues pour leurs propriétés pharmacologiques dans différents domaines et testées (plus ou moins par hasard) sur la P-gp ; lorsqu'elles ont été observées efficaces, elles n'ont cependant pas pu être utilisées en clinique car trop toxiques en relation avec leurs autres propriétés pharmacologiques qui sont patentées pour des plus faibles concentrations. Les CS de seconde génération sont des molécules obtenues par modifications chimiques de ceux de 1<sup>ère</sup> génération, mais ils ont aussi montré des effets toxiques importants (souvent du fait de leur affinité trop faible), bien que différents des propriétés des composés parents. Les CS de troisième génération sont des molécules « nouvelles » de synthèse totale, souvent obtenues par criblages extensifs de chimiothèques ; elles sont en général prometteuses à partir de leurs résultats *in-vitro* (bonne affinité), mais malheureusement elles se sont aussi révélées présenter des toxicités cliniques inattendues et inacceptables lors d'essais en phase I. Il faudrait donc passer à la quatrième génération, qui serait obtenue par design rationnel basé sur les données structurales du site de liaison de la P-gp, permettant (en théorie) d'élaborer des structures optimisées pour leur reconnaissance par la P-gp, avec affinité et sélectivité maximales.

Cette grande variété de substrats et l'absence de motifs chimiques communs et/ou de cibles pharmacologiques communes est une particularité de la P-gp (mais partagée avec CYP3A4). Un peu paradoxalement, cette particularité rend compliquée la découverte de nouveaux inhibiteurs spécifiques et efficaces n'engendrant pas d'effet toxique sur la cellule et pouvant donc être utilisés en clinique. En effet, on n'est pas comme dans le cas d'une enzyme « classique » dont on cherche à inhiber l'activité en étant guidé, par sa relation structure-

activité précise, vers des analogues du substrat bien spécifique. Dans le cas de la P-gp, de nombreuses molécules, connues pour l'inhiber *in-vitro*, présentent *in-vivo* des effets secondaire inacceptables en clinique à cause de l'existence d'autres cibles biologiques (prévisibles du fait de leurs propriétés pharmacologiques déjà connues ou bien imprévisibles dans le cas de molécules nouvellement isolées « au hasard » par criblage ou chimie combinatoire). L'identification des domaines protéiques et/ou des acides aminés impliqués dans la reconnaissance, la liaison, le transport à travers la bicouche lipidique membranaire et la libération d'un substrat donné, s'impose comme un pré-requis essentiel dans le but de comprendre la structure, la fonction et les mécanismes d'action de la P-gp [62]. Pour étudier l'interaction enzyme-ligand dans le cas de la P-gp, plusieurs approches ont été utilisées, des approches biochimiques se basant sur les propriétés fonctionnelles de l'enzyme, ou bien celles des ligands de la P-gp possédant une caractéristique permettant leur détection et une approche génétique analysant l'impact d'une mutation donnée sur l'interaction de la protéine avec son ligand. Ces approches ont été complétées par des études structure-activité de différents ligands. Ainsi, plusieurs études de mutants et de photomarquage ont suggéré que le domaine de liaison des substrats se localiserait au niveau des domaines transmembranaires [63]. Une idée soutenue par le fait qu'une activité ATPase stimulée n'est observée qu'en présence des deux domaines TMD, quand les deux moitiés de la P-gp sont co-exprimées [53]. Alors que la localisation transmembranaire du domaine de liaison du substrat est maintenant majoritairement admise, l'existence d'un ou plusieurs sites de liaison continue à faire débat au sein de la communauté scientifique. L'identification de deux types de relations entre les substrats de la P-gp, compétitive et non compétitive, montrent l'existence de plusieurs (sous-) sites de liaison sur l'enzyme, mais leur nombre, leurs relations mutuelles et leurs sélectivités respectives restent encore à être éclaircies. Les travaux menés par Loo et Clarke soutenant la possibilité à la P-gp de lier simultanément deux substrats différents sont en accord avec cette hypothèse [64]. Des travaux menés au sein du laboratoire ont aussi permis de soutenir cette hypothèse en proposant une cartographie éléments consensus de reconnaissance de certains ligand par la P-gp, ou « pharmacophore » sur l'enzyme [58].

Plusieurs études de mutagenèse dirigée ont aussi permis d'apporter de nombreuses informations. En effet, un certain nombre de mutations ont été définies comme affectant ou le profil de résistance vis-à-vis d'une série de substrats donnés (Tableau-A-6), donc très vraisemblablement leur liaison. Il est intéressant de remarquer dans les données de ce tableau que certaines mutations peuvent provoquer plusieurs phénomènes de résistance des cellules à divers substrats médicamenteux.

Mutation in P-glycoprotein <sup>b</sup>	Vinblastine <sup>c</sup>	Actinomycin D	Colchicine	Daunomycin	Etoposide
<b>N-terminal half</b>					
H61[G/A/S/C/P/T/D/V/N/I/E/L/Q] <sup>c</sup> (1)	1.0-2.1	ND	0.3-1	0.3-1.1	0.5-1.2
H61[M/F/Y/W] <sup>c</sup> (1)	0.3-0.9	ND	1.5-5.3	2.2-5.2	1.4-3.3
H61R <sup>c</sup> (1)	0.5	ND	2.7	0.5	3.2
G64R <sup>c</sup> (1)	0.5	ND	3.5	0.2	3.5
L65R <sup>c</sup> (1)	0.15	ND	2.9	0.2	1.8
I306R <sup>c</sup> (5)	ND	ND	ND	ND	ND
F335A <sup>c</sup> (6)	0.1	0.2	1.3	1.6	ND
V338A <sup>c</sup> (6)	0.3	1.0	2.8	2.0	ND
G341A <sup>c</sup> (6)	0.8	1.0	0.1	0.1	ND
N350I <sup>d</sup> (6)	ND	ND	≥ 0.5	≥ 2	ND
<b>C-terminal half</b>					
I837L <sup>d</sup> (9)	ND	1.6	0.8	0.9	ND
N839I <sup>d</sup> (9)	ND	1.5	0.2	0.5	ND
I862F <sup>d</sup> (10)	ND	ND	ND	≥ 2	ND
L865F <sup>d</sup> (10)	ND	ND	≥ 1.0	≥ 2	ND
A933T <sup>d</sup> (11)	ND	ND	ND	≥ 1.5	ND
F934A <sup>d</sup> (11)	1.0	0.75	0.5	0.4	ND
F938A <sup>d</sup> (11)	1.2	0.85	0.8	0.5	ND
S939F <sup>d</sup> (11)	2.7	11.8	28.7	13.5	ND
T941A <sup>d</sup> (11)	0.6	1.0	0.3	0.3	ND
Q942A <sup>d</sup> (11)	1.3	1.2	0.4	0.3	ND
Y946A <sup>d</sup> (11)	0.5	0.3	0.8	1.1	ND
S948A <sup>d</sup> (11)	1.0	0.5	0.8	0.7	ND
Y949A <sup>d</sup> (11)	0.9	0.4	0.4	0.2	ND
F953A <sup>d</sup> (11)	0.5	0.1	0.1	0.2	ND
L975A <sup>c</sup> (12)	ND	ND	ND	†	ND
V981A <sup>c</sup> (12)	ND	ND	ND	†	ND
F983A <sup>c</sup> (12)	ND	ND	ND	†	ND
F978L <sup>c</sup> (12)	0.9	1.0	0.3	0.3	ND

**Tableau-A-6 : Analyses des changements de résistance associés à des mutations qui modifient la spécificité de la P-glycoprotéine humaine et murine. Adapté de [65].** La nomenclature à une lettre a été utilisée pour désigner les acides aminés concernés par les mutations. Les numéros entre parenthèses indiquent ***l'hélice transmembranaire*** de l'acide aminé. Le profil de résistance des cellules est : =1, quand la mutation ne modifie pas résistance des cellules; < 1, la mutation augmente la sensibilité des cellules en comparaison au type sauvage ; > 1, la mutation augmente la résistance des cellules en comparaison au type sauvage, ND : Non Déterminé.

Il est possible de remarquer que les mutations affectant la spécificité de la P-gp sont localisées au niveau des hélices transmembranaires. Des mutations localisées au niveau des domaines NBD et des domaines intracellulaires et altérant la fonction de transport de la P-gp ont cependant été révélées par plusieurs études [66-67], mais ces mutations n'altèrent pas le profil de reconnaissance, contrairement à certaines mutations affectant les domaines TMD. Ces constatations suggèrent donc un rôle important des régions transmembranaires dans la détermination des sites de liaison des substrats mais ne permettent pas de comprendre leur mode de fonctionnement (autonome ou inter-dépendant). En effet, et en tenant compte de sa grande multispécificité, il n'est pas absurde d'imaginer l'intervention, de manière directe ou indirecte, de différents domaines membranaires de la P-gp dans la description globale des sites de liaison de ses substrats.

#### II.3.4. La P-gp : le phénotype MDR

La notion de MDR appelée aussi résistance pléiotropique, a été définie suite à la découverte d'une résistance croisée entre l'actinomycine D, agent anticancéreux, et d'autres composés chimiques [34]. La MDR a donc été définie comme la faculté de certaines cellules à présenter une résistance croisée à des agents cytotoxiques sans analogie de structure [68-69]. La caractéristique de ce type de résistance est que les cellules concernées développent une résistance envers des molécules cytotoxiques auxquelles elles n'ont jamais été exposées, en même temps que la résistance acquise contre les molécules anticancéreuses utilisées lors du traitement initial, ce qui complique la stratégie thérapeutique.

Les facteurs à l'origine de la résistance multiple aux médicaments correspondent à des altérations génétiques au niveau cellulaire : altération du niveau d'expression de diverses protéines spécifiques ou synthèse de protéines mutées. Il existe plusieurs types de résistance multiple aux médicaments : celles dites de type « classique » par efflux actif et qui implique des protéines comme la P-gp ou MRP1, et celles dites « atypiques » et qui impliquent des protéines telles que la glutathion-S-transférase, responsable de la métabolisation de différentes drogues, ou la topoisomérase II, cible commune à différentes drogues. Le phénotype MDR « classique » se traduit par une accumulation réduite des composés cytotoxiques due à l'augmentation de leur efflux vers le milieu extracellulaire. Le mécanisme MDR le plus étudié est celui dû à la surexpression d'une phosphoglycoprotéine présente au niveau de la membrane plasmique des cellules MDR. Cette protéine a été surnommée P-glycoprotéine par Juliano et Ling en 1976, le « P » désignant alors le terme perméabilité [16]. Dans ce même article, Juliano et Ling ont mis en évidence l'existence d'une corrélation entre le taux de résistance cellulaire et la quantité de P-gp exprimée à la surface des cellules.

Il existe des moyens pharmacologiques pour moduler le phénotype MDR. En effet, un certain nombre de composés non cytotoxiques connus sous le nom d'agents réversants ou chimiosensibilisants ou modulateurs MDR, permettent de sensibiliser les cellules résistantes à l'action des médicaments cytotoxiques. Ces chimiorésistants incluent des peptides cycliques, des analogues de médicaments, des stéroïdes. Des efforts sont menés dans plusieurs laboratoires pour comprendre les bases fonctionnelles et structurales de ce type d'inhibition. On suppose un mécanisme d'action selon lequel l'agent réversant entrerait en compétition avec le composé cytotoxique, ce qui provoquerait une réduction de l'expulsion de ce dernier vers le milieu extracellulaire, mais il a depuis été montré que l'inhibition observée pouvait aussi être compétitive. Des chimiosensibilisants comme le vérapamil et la trifluoropérazine

ont été utilisés sans réelle efficacité, et souvent avec des effets toxiques inacceptables, chez des patients dont les tumeurs étaient cliniquement multirésistantes. Ces échecs peuvent être liés, d'une part, à la toxicité propre de ces modulateurs, qui ne peuvent être employés *in vivo* qu'à des concentrations sub-optimales, et, d'autres part, au fait que ces tumeurs avaient peut-être d'autres mécanismes de chimiorésistance associés [70].

## II.4. La P-gp d'un point de vue structural

### II.4.1. Structure du gène ABCB1

Chez l'homme, le gène *ABCB1* est situé sur le bras long du chromosome 7. Il est composé de plus de 200kb et possède 29 exons et 28 introns. Le gène *ABCB1* et le gène *CYP3A4* sont très proches sur le même chromosome (7q21 et 7q22, respectivement) (Figure-A-7). Une corégulation de ces deux gènes pourrait exister [71].

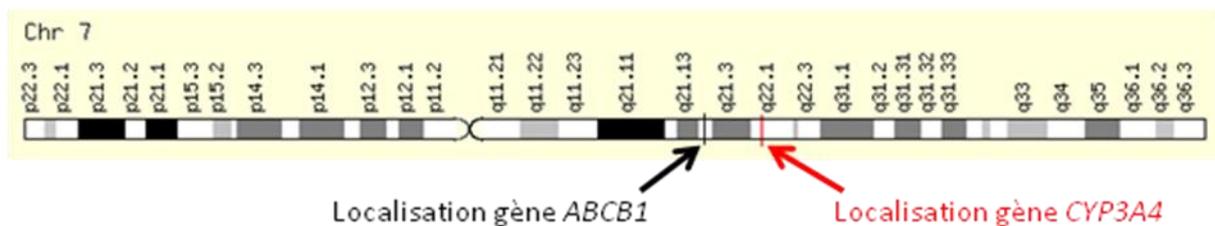


Figure-A-7 : Localisation sur le chromosome 7 des gènes *ABCB1* et *CYP3A4* (Source : [www.genecards.org](http://www.genecards.org))

### II.4.2. Structure de la protéine P-gp

#### II.4.2.1. La structure primaire

La P-gp a été identifiée en utilisant des anticorps monoclonaux spécifiques [72] dans plusieurs lignées cellulaires MDR, dans plusieurs tissus normaux et tumoraux, et chez plusieurs espèces. Ainsi, le séquençage complet des gènes codant pour les différentes P-gp est disponibles chez plusieurs mammifères (Homme, souris, hamster chinois, rat, chien, porc)

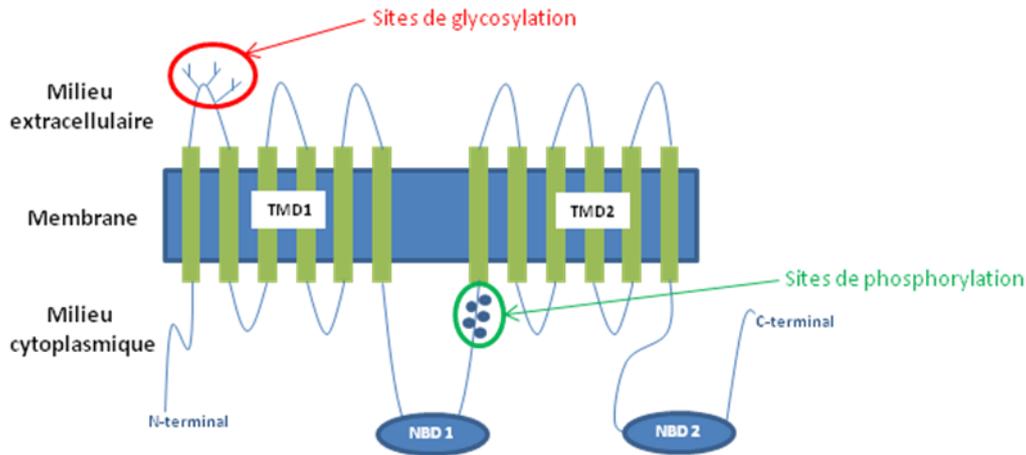
depuis le milieu des années 80. Ces gènes produisent plusieurs isoformes qui ont été différenciées en trois classes en fonction de l'espèce d'origine et en fonction de la protéine issue du gène considéré (Tableau-A-7) [73].

Espèces	Classe I	Classe II	Classe III
<b>Homme</b>	MDR1 = B1		MDR2/3 = B4
<b>Souris</b>	mdr3	mdr1	mdr2
	mdr1a	mdr1b	mdr2
<b>Hamster chinois</b>	pgp1	pgp2	pgp3

**Tableau-A-7 : Nomenclature des gènes codant pour la P-gp.** *La nomenclature utilisée pour les différents isoformes de la P-gp. Les gènes des classes I et II codent pour des protéines conférant le phénotype de résistance MDR. Ceux de la classe III codent pour des protéines possédant la fonction de floppase de la phosphatidylcholine.*

Chez l'Homme, la P-gp dans son état mature présente une masse moléculaire de 170 kDa, et est composée de 1280 acides aminés [16]. Dans les lignées cellulaires à partir desquelles la P-gp a été isolée, on trouve 100 000 à 200 000 molécules de P-gp par cellule, soit un pourcentage non négligeable des protéines membranaires. Cette protéine relativement stable a une demi-vie décrite entre 18 et 72 heures [74]. Les masses moléculaires élevées de la P-gp mature résultent de modifications post-traductionnelles incluant des glycosylations et des phosphorylations [75] : la masse moléculaire de la P-glycoprotéine en l'absence de glycolysation est de 140 kDa [7]. Les sites de glycosylation se situent du côté N-terminal au niveau de la première boucle extracellulaire (Figure-A-8). La structure primaire de la protéine prédit 2, 3 ou 4 sites possibles (les Asparagine 91-94 ou 99 sont des sites confirmés) pour la glycosylation [62]. De nombreuses études menées sur la P-gp ont montré que la glycosylation ne jouait pas un rôle important dans la fonction de transport des substrats. Par contre, des mutants de la P-glycoprotéine perdant ses sites de glycosylation ont révélé un rôle du groupement N-glycosylé dans le bon repliement et/ou l'adressage correct de la P-gp dans la membrane plasmique. La phosphorylation de la P-gp a été observée dans plusieurs lignées cellulaires résistantes chez les humains et les rongeurs [62]. Cette universalité a été le sujet de plusieurs études suggérant un rôle des mécanismes de phosphorylation et déphosphorylation dans la régulation de l'activité d'efflux des médicaments. Les résidus susceptibles d'être concernés par la phosphorylation sont situés au niveau du segment séparant les moitiés N- et C- terminales de la P-glycoprotéine (Figure-A-8). Les protéines kinases cellulaires qui

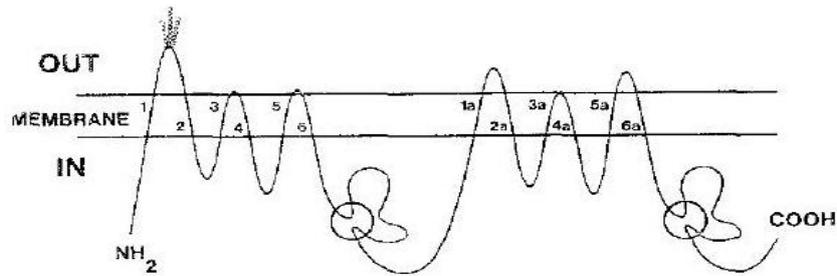
interviennent, ainsi que les répercussions fonctionnelles respectives des différents sites de phosphorylation, sont encore largement débattues.



**Figure-A-8 : Représentation schématique de la P-glycoprotéine et de ses sites de glycosylation et de phosphorylation.** Adapté de [73]

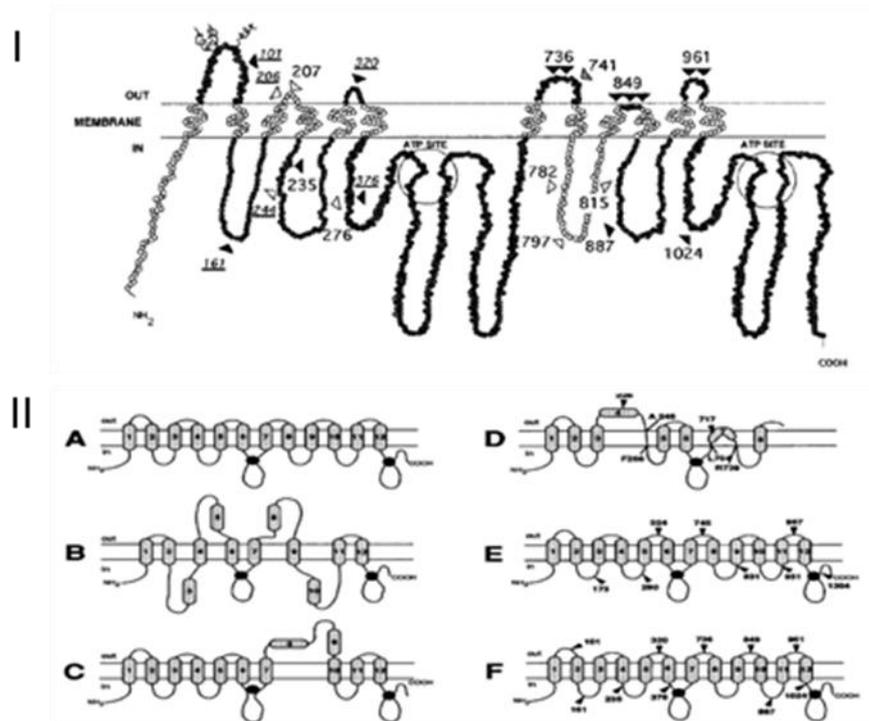
#### II.4.2.2. Structure secondaire et topologie

La séquence du gène codant pour la P-gp a permis de déduire sa composition de deux moitiés homologues (43% d'identité et 78% d'homologie). Une fois la séquence primaire identifiée, une prédiction structurale pouvait être établie. Les deux moitiés symétriques comportent chacune un domaine TMD hydrophobe et un domaine NBD hydrophile. Un domaine TMD est composé de six segments transmembranaires (TM) possédant chacun une vingtaine d'acides aminés hydrophobes reliés entre eux par des boucles extra- et intra-cellulaires. Un domaine NBD est un domaine cytoplasmique liant les nucléotides et comportant les motifs spécifiques décrits précédemment (Figure-A-2). Il compte environ 250 acides aminés. Les deux moitiés homologues de la P-gp sont séparées par un domaine appelé « région linker » comportant environ 60 acides aminés [75] (Figure-A-9) [7].



**Figure-A-9: Première représentation schématique de la P-gp et de l'orientation de ses domaines transmembranaires par [7].** Les domaines sont numérotés de 1 à 6 pour la première moitié et de 1a à 6a pour la seconde. Les sites de liaison de l'ATP sont représentés par des cercles, ceux de la glycolysation sont représentés au niveau du domaine extracellulaire entre les domaines transmembranaires 1 et 2.

Ces résultats ont permis d'orienter les études menées pour établir la topologie membranaire de la P-gp. Les premiers modèles établis s'étaient basés sur les profils d'hydrophatie [7] et d'homologie de séquence que possède la P-gp avec les transporteurs bactériens [76]. Ce n'est qu'en 1995, en utilisant la technique de marquage par épitope peptidique que Kast *et al.* ont pu déterminer la topologie de la moitié N-terminale de la P-gp [77]. Les résultats de cette étude ont indiqué que les segments séparant les TM1 et TM2 ainsi que ceux séparant TM5 et TM6 sont extracellulaires alors que les segments séparant TM2 et TM3 et celui en aval de TM6 sont intracellulaires. Ces observations sont en cohérence avec les modèles prédits dans les années 80 (Figure-A-9). Cette étude a aussi confirmé la localisation extracellulaire des sites de glycolysation et la localisation intracellulaire du premier site de liaison de l'ATP. L'équipe Kast *et al.* a pu l'année suivante poursuivre l'étude et identifier avec la même technique la localisation des autres segments constituant la P-gp. Ainsi, le segment séparant TM4 et TM5, celui séparant TM10 et TM11 et celui en aval de TM12 sont intracellulaires alors que ceux situés entre TM7 et TM8, TM9 et TM10, TM11 et TM12 sont extracellulaires [78]. Ces résultats ont confirmé le modèle de structure secondaire de Gottesman et Pastan (1988), construit par homologie, distinguant la localisation et l'orientation des 12 segments transmembranaires, des sites de glycolysation au niveau de la première boucle extracellulaire et des sites nucléotidiques au niveau intracellulaire (Figure-A-10).



**Figure-A-10 : Les modèles proposés pour décrire la topologie membranaire de la P-gp. I- La topologie aujourd'hui admise. II- L'ensemble des topologies proposées depuis 1986 jusqu'à celle proposée en 1996 par Kast et al. (F) [78]**

En plus de la description de la topologie membranaire de la P-gp, des études ont été menées dans le but d'évaluer quantitativement la structure secondaire de MDR1. Parmi ces études, celle de Vigano *et al.* en 2002, a utilisé la spectroscopie infrarouge à transformée de Fourier (FTIR) pour déterminer la composition en hélices  $\alpha$ , feuillets  $\beta$ , et coudes de la structure secondaire de la P-gp sauvage et de P-gp mutées au niveau des motifs Walker B des domaines NBD1 ou NBD2 (Tableau-A-8).

	$\alpha$ -Helix	$\beta$ -Sheet	$\beta$ -Turn	Random
	%	%	%	%
WT	$46 \pm 3$	$18 \pm 3$	$8 \pm 3$	$28 \pm 3$
E552Q	$47 \pm 3$	$19 \pm 3$	$7 \pm 3$	$27 \pm 3$
E1197Q	$47 \pm 3$	$20 \pm 3$	$7 \pm 3$	$26 \pm 3$

**Tableau-A-8 : Composition de la structure secondaire de la P-gp sauvage et mutée au niveau des acides aminés 552 et 1197 des motifs Walker B de NBD1 et de NBD2 [79]**

On peut remarquer que cette suggestion de composition en structure secondaire ne reflète pas de différences significatives entre formes sauvages et mutées des domaines Walker B. L'hydrolyse de l'ATP n'influerait donc pas sur la structure secondaire de la P-gp. Une autre étude menée par Dong *et al.* [80] en utilisant la technique du dichroïsme circulaire a donné des résultats quantitatifs comparables: 43% d'hélice  $\alpha$ , 16% de feuillet  $\beta$ , 15% de coudes. Au final, il est possible de retenir une topologie « tandem » avec deux moitiés N- et C-terminales homologues (Figure-A-1A).

#### II.4.2.3. Structure tertiaire

Depuis sa découverte, la P-gp a suscité un grand intérêt et son étude d'un point de vue structural comme d'un point de vue fonctionnel était et demeure l'objet de multiples travaux. Cependant, la complexité de son fonctionnement en tant qu'enzyme multispécifique, sa nature même de protéine possédant à la fois une surface hydrophobe en contact avec la membrane et une ou deux surface(s) hydrophile(s) en contact avec le milieu aqueux ainsi que la faible quantité de protéines disponibles à la surface des cellules ont limité les données expérimentales concernant sa structure tridimensionnelle. Ces difficultés concernent toutes les protéines membranaires. Ainsi, le nombre de structures tridimensionnelles de protéines membranaires actuellement disponibles dans la PDB (Protein Data Bank) est de 279 structures ([http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), date : 31 Mars 2011). La parution de la première structure tridimensionnelle d'une protéine membranaire ne date que de 1985 [structure du centre réactionnel photosynthétique de *Rhodospseudomonas viridis* (Deisenhofer *et al.*, 1985 non paru dans *PubMed*)]. S'agissant des transporteurs ABC, les efforts très soutenues pour obtenir des structures tridimensionnelles n'ont porté leur fruit que dans la dernière décennie. Ainsi, il a fallu attendre 2002 [81] pour pouvoir observer la première structure tridimensionnelle à une assez bonne résolution d'une protéine ABC, la protéine BtuCD d'*Escherichia. coli* (*E. coli*, 3.2 Å, voir Tableau-A-9).

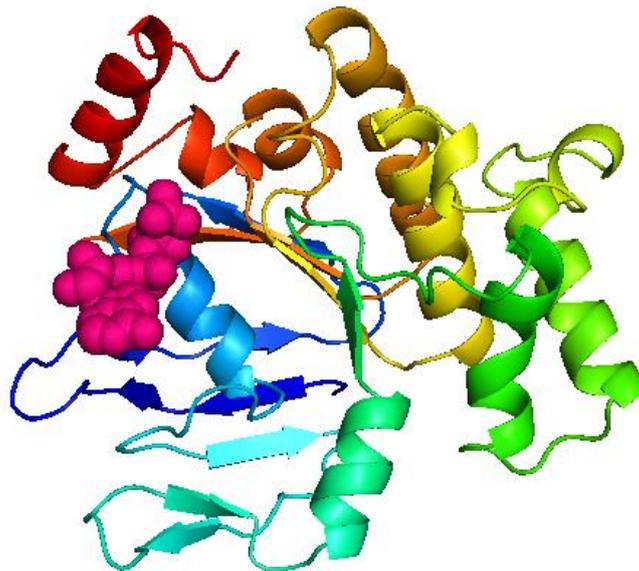
Protéine ABC	Organisme	Code PDB	Références
<b>BtuCD vitamin B<sub>12</sub> transporter</b> 3.2 Å	<i>Escherichia coli</i>	1L7V	Locher <i>et al</i> 2002
<b>BtuCD-F Complex ; BtuCD B<sub>12</sub> transporter + Btu F binding protein;</b> 2.6 Å	<i>Escherichia coli</i>	2QI9	Hvorup <i>et al</i> 2007
<b>Sav1866 multidrug transporter</b> 3.0 Å	<i>Staphylococcus aureus</i>	2HYD	Dawson et Locher 2006
<b>Molybdate transporter ModB<sub>2</sub>C<sub>2</sub> complexed with ModA; 3.1 Å</b> ModA with bound MoO <sub>4</sub> : 2ONR, 1.60 Å ModA with bound WO <sub>4</sub> : 2ONS, 1.55 Å	<i>Archaeoglobus fulgidus</i>	2ONK	Hollenstein <i>et al</i> 2007
<b>ModBC molybdate ABC transporter</b> 3.0 Å	<i>Methanosarcina acetivorans</i>	3D31	Gerber <i>et al</i> 2008
<b>HI1470/1 putative metal-chelate-type ABC transporter</b> 2.4 Å	<i>Haemophilus influenzae</i>	2NQ2	Pinkett <i>et al</i> 2007
<b>MsbA lipid « flippase » with AMPPNP</b> 3.7 Å ADP+Vanadate-bound conformation: 3B5Z, 4.2 Å Open apo-conformation ( <i>E.coli</i> ): 3B5W, 5.3 Å Closed apo-conformation ( <i>Vibrio cholerae</i> expressed in <i>E.coli</i> ): 3B5X, 5.5 Å	<i>Salmonella typhimurium</i> (expressed in <i>E.coli</i> )	3B60	Ward <i>et al</i> 2007
<b>P-Glycoprotein</b> 3.8 Å With bound QZ59-RRR : 3G60, 4.4 Å With bound QZ59-SSS:3G61, 5.5 Å	<i>Mus musculus</i> (Mouse) (expressed in <i>Pichia pastoris</i> )	3G5U	Aller <i>et al</i> 2009
<b>MalFGK<sub>2</sub> MBP maltose uptake transporter complex</b> 2.8 Å	<i>Escherichia coli</i>	2R6G	Oldham <i>et al</i> 2007
<b>MalFGK<sub>2</sub> uptake transporter</b> 4.5 Å	<i>Escherichia coli</i>	3FH6	Khare <i>et al</i> 2009
<b>MetNI methionine uptake transporter complex</b> 3.7 Å MetN-C2 domain : 3DHX, 2.1 Å	<i>Escherichia coli</i>	3DHW	Kadaba <i>et al</i> 2008
<b>FbpC ferric iron-uptake transporter nucleotide-binding domain</b> 1.9 Å	<i>Neisseria gonorrhoeae</i>	3FVQ	Newstead <i>et al</i> 2009

Tableau-A-9 : Les structures des transporteurs ABC disponibles dans la PDB<sup>1</sup> à la date du 31 Mars 2011

Dans ce tableau, on constate qu'à ce jour, la majorité des protéines ABC pour lesquelles une structure tridimensionnelle est disponible, sont d'origine bactérienne, souvent des « demi-transporteurs » plus faciles à obtenir en grandes quantités. Ce tableau ne contient que les structures des protéines ABC entièrement résolues (parties NBD et TMD). Des

<sup>1</sup> [http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html)

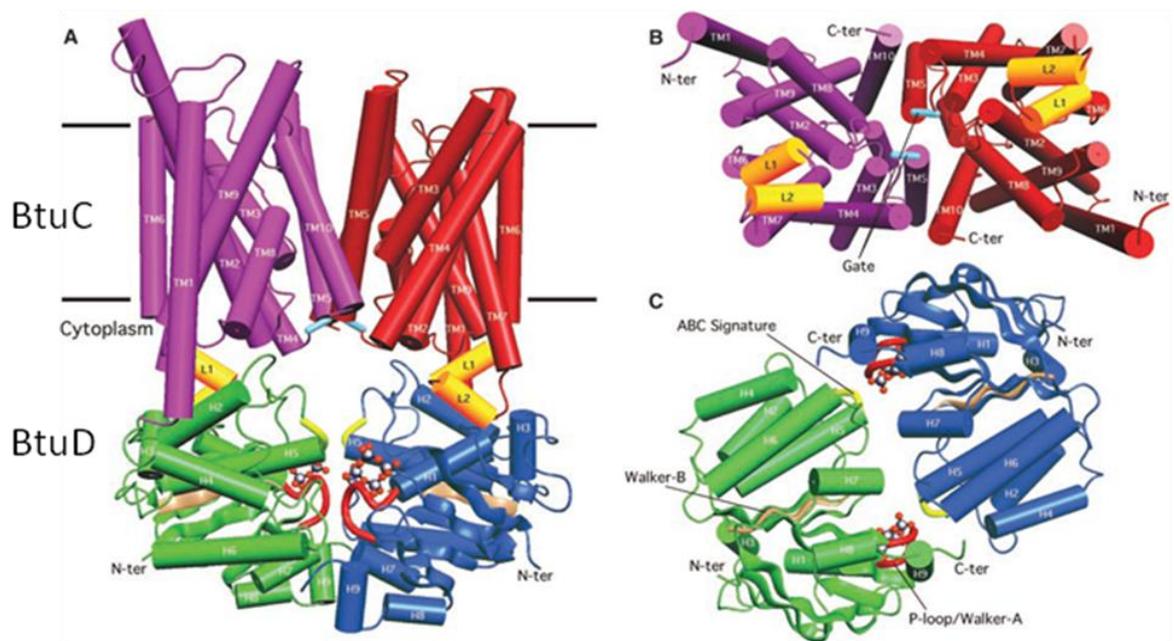
structures représentatives des transporteurs ABC ont paru quelques années avant la publication des structures contenues dans le Tableau-A-9 mais ne représentaient que les domaines solubles, les domaines nucléotidiques étant plus faciles à exprimer et cristalliser. La première structure de haute résolution d'un domaine ABC a paru en 1998. Il s'agissait de la structure représentant le domaine catalytique liant l'ATP de l'histidine perméase (Figure-A-11) [82]. Cette structure a été la première représentation tridimensionnelle du domaine nucléotidique conservé chez les transporteurs ABC. Hung *et al.* ont mis en évidence une corrélation entre les propriétés des mutants CFTR et la structure de HisP prouvant ainsi la possibilité de prendre HisP comme modèle représentatif des domaines NBD des transporteurs ABC en général. Cette première réussite a été suivie les années suivantes par d'autres structures de domaines nucléotidiques très proches de la structure de HisP et avec une assez haute résolution (exemples : MalK, Rad50, MJ0796, MJ1267 et TAP1, respectivement, les codes PDB suivants : 1G29, 1F2U, 1F3O, 1GAJ, 1JJ7).



**Figure-A-11 : Structure du domaine de liaison de l'ATP de l'histidine perméase (2.20 Å) (Code PDB : 1BOU).** Coloration en arc en ciel des hélices  $\alpha$  et des feuillets  $\beta$  avec le côté N-terminal en bleu et le côté C-terminal en rouge. L'ATP est représentée en sphères de couleur magenta.

En 2002, une structure complète résolue aux rayons X était enfin disponible [81]. La protéine BtuCD est un transporteur chargé de l'import de la vitamine B<sub>12</sub>. Elle est arrangée en quatre sous-unités, deux sous-unités membranaires BtuC comportant une vingtaine d'hélices

transmembranaires et deux sous-unités cytoplasmiques BtuD présentant des sites de liaison de l'ATP et formées de chacune de neuf hélices  $\alpha$  et de neuf feuilletts  $\beta$  (Figure-A-12). L'arrangement dimérique des deux sous-unités BtuC offre le chemin possible de translocation vers le cytoplasme alors que l'organisation spatiale des sous-unités BtuD ressemble à celui de HisP et Rad50.

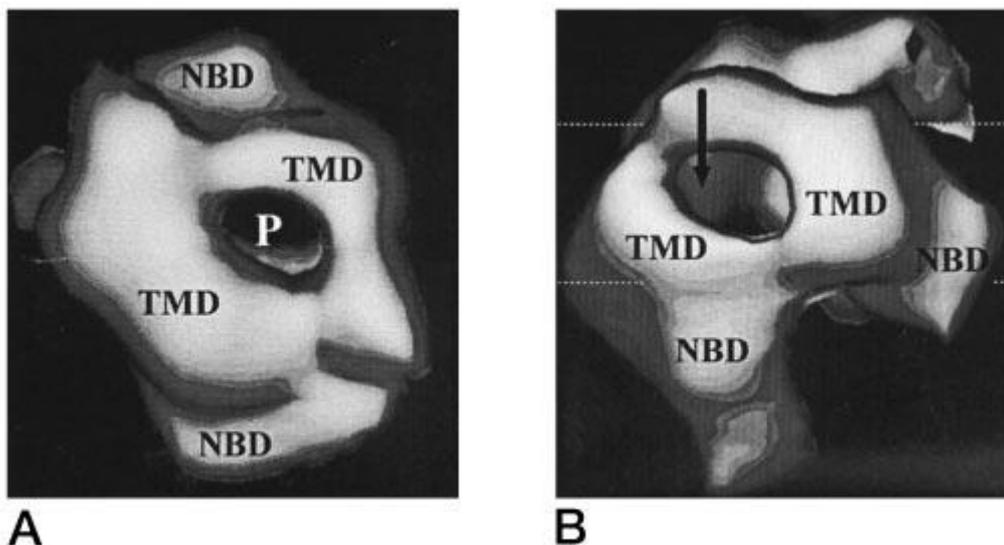


**Figure-A-12 : Structure tridimensionnelle du transporteur de la vitamine B<sub>12</sub> [81].** Les deux sous-unités BtuC sont représentées en rouge et en violet, celles de BtuD en vert et bleu. **A-** Vue latérale, **B-** Vue du côté cytoplasmique des sous-unités BtuC, **C-** Vue de dessous des sous-unités BtuD.

Plusieurs tentatives infructueuses de détermination de structures entières de protéines ABC par diffraction aux rayons X ont précédé la parution de la structure de BtuCD. D'ailleurs, cette structure était considérée jusqu'en 2006 comme la seconde structure à paraître pour un transporteur ABC. En effet, une structure cristallisée du transporteur membranaire ABC homodimérique, MsbA, avait paru en 2001 suite à des travaux menés par Chang et Roth [83]. Mais, cette structure a été retirée de la PDB en Décembre 2006, tout comme cinq autres structures du même groupe du Scripps, parues entre 2001 et 2006 (trois structures décrivaient le transporteur MsbA, les deux autres décrivaient un autre transporteur bactérien, EmrE). Ces rétractations de structures ainsi que celles de leurs articles de dépôt, avaient eu un grand écho au sein de la communauté scientifique internationale étant donné les

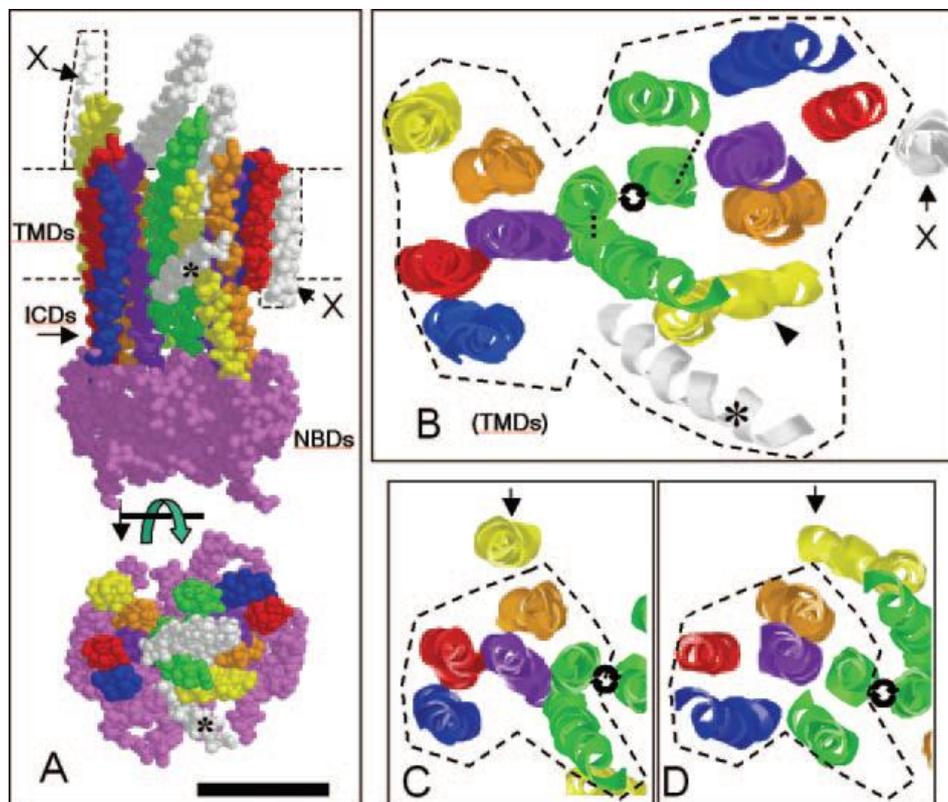
cinq années séparant la parution et la rétractation des structures, avec pour conséquence non négligeable la parution pendant cette période de nombreux travaux basés sur ces structures modèles et aux conclusions erronées. Certaines hypothèses basées sur ces structures sont toutefois restées plausibles comme le modèle de translocation décrit par Higgins et Linton [84] et décrivant le cycle des changements conformationnels à l'origine du transport des solutés à travers la membrane.

L'obtention d'une structure tridimensionnelle d'une assez haute résolution représentant un importateur ABC (BtuCD) représentait une grande avancée dans l'étude structurale des transporteurs ABC. Mais l'absence d'une homologie structurale entre BtuCD (avec ses 20 segments TM et son substrat de grande taille et hydrophile) et les protéines exportatrices ABC, comprenant les transporteurs d'intérêt clinique comme la P-gp, n'avait pas permis d'avancées structurales sur les mécanismes de résistance. La première structure d'une protéine ABC décrite dans la littérature a néanmoins été celle de la P-gp en 1997 (Figure-A-13) [85]. Cette structure, de faible résolution ( $\sim 25 \text{ \AA}$ ), représentait la P-gp dans l'état liant un nucléotide. Elle a été obtenue par une reconstruction tridimensionnelle d'images obtenues par microscopie électronique de microcristaux bidimensionnels.



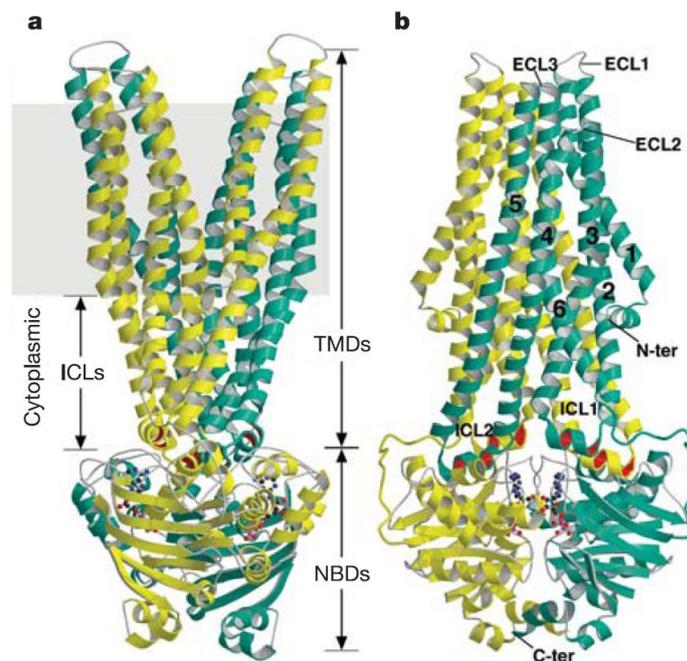
**Figure-A-13 : Structure à basse résolution ( $25 \text{ \AA}$ ) de la P-gp [85].** *Projections de reconstruction 3D de la P-gp. Les TMD désignent les domaines transmembranaires et les NBD, les domaines nucléotidiques. « P » désigne le pore central dans la vue de dessus (A), ce pore est désigné par une flèche au niveau de la vue latérale (B). Les pointillés blancs désignent les limites de la membrane.*

Malgré sa faible résolution, cette première avancée dans l'étude structurale de la P-gp de manière particulière et des exportateurs ABC de manière générale a permis de visualiser pour la première fois un pore central (désigné par « P » sur la Figure-A-13A) qui permettrait la communication entre la protéine et le milieu extracellulaire aqueux. Un pore au niveau latéral a aussi été imaginé (avec une flèche sur la Figure-A-13B), un pore qui permettrait une éventuelle prise en charge des substrats au niveau de la bicouche lipidique [86]. L'équipe Rosenberg a proposé plus tard de nouvelles structures tridimensionnelles de la P-gp avec des meilleures résolutions, d' $\sim 10$  Å [87] puis d' $\sim 8$  Å (Figure-A-14) [88] et obtenues grâce à la même technique.



**Figure-A-14 : Interprétation de la structure tridimensionnelle de la P-gp (8 Å) [88].** *A- Vue latérale et vue de dessus de la P-gp. En violet, les domaines NBD, les hélices transmembranaires sont colorées par paire pour distinguer les deux moitiés de la protéine. B- Vue de dessus avec une représentation en spirales des hélices transmembranaires. C et D- Vues identiques à B- évoquant une pseudo-symétrie entre les hélices entourées de tirets en pointillés.*

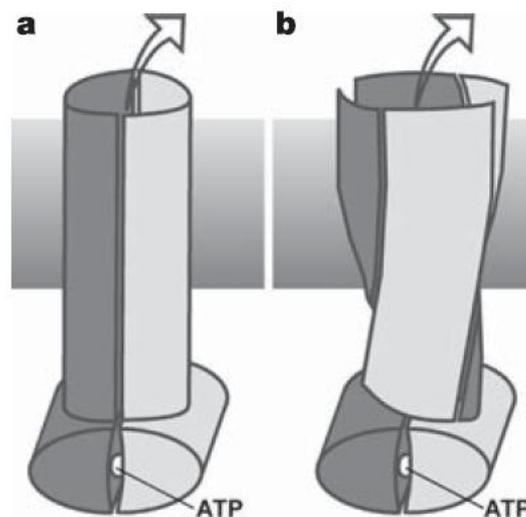
Ces travaux uniques à ce jour de la protéine P-gp humaine liant l'ATP ont permis d'orienter de nombreux travaux de détermination des structures tridimensionnelles de transporteurs ABC. La première structure tridimensionnelle avec une assez haute résolution (3.0 Å) d'un exportateur ABC était celle de Sav1866 de *Staphylococcus aureus* (Figure-A-15) [89]. Sav1866 est un héli-transporteur dont la forme fonctionnelle est homodimérique. La parution de cette structure de protéine homodimérique a permis de justifier l'invalidation des structures de MsbA parues entre 2001 et 2006 et a fourni des preuves concrètes que dans la conformation *nucleotide-bound*, les deux tandems TMD-NBD sont interactivement liés avec l'interface principale des boucles intracellulaires (ICL) ancrée au niveau du NBD de l'unité opposée.



**Figure-A-15 : Structure de Sav1866 [89].** *a.* Le backbone de la protéine homodimérique dans une représentation spirale avec les deux sous-unités colorées en jaune et en turquoise. Le carré gris représente la probable localisation de la membrane. *b.* Une vue de *a.* après une rotation de 90° autour d'un axe vertical. ICL : les boucles intracellulaires. ECL : les boucles extracellulaires.

Les deux sous-unités exhibent un considérable twist avec une vraie interaction entre les domaines transmembranaires et les domaines nucléotidiques. Dawson et Locher expliquent les changements conformationnels générés par la liaison et l'hydrolyse de l'ATP et

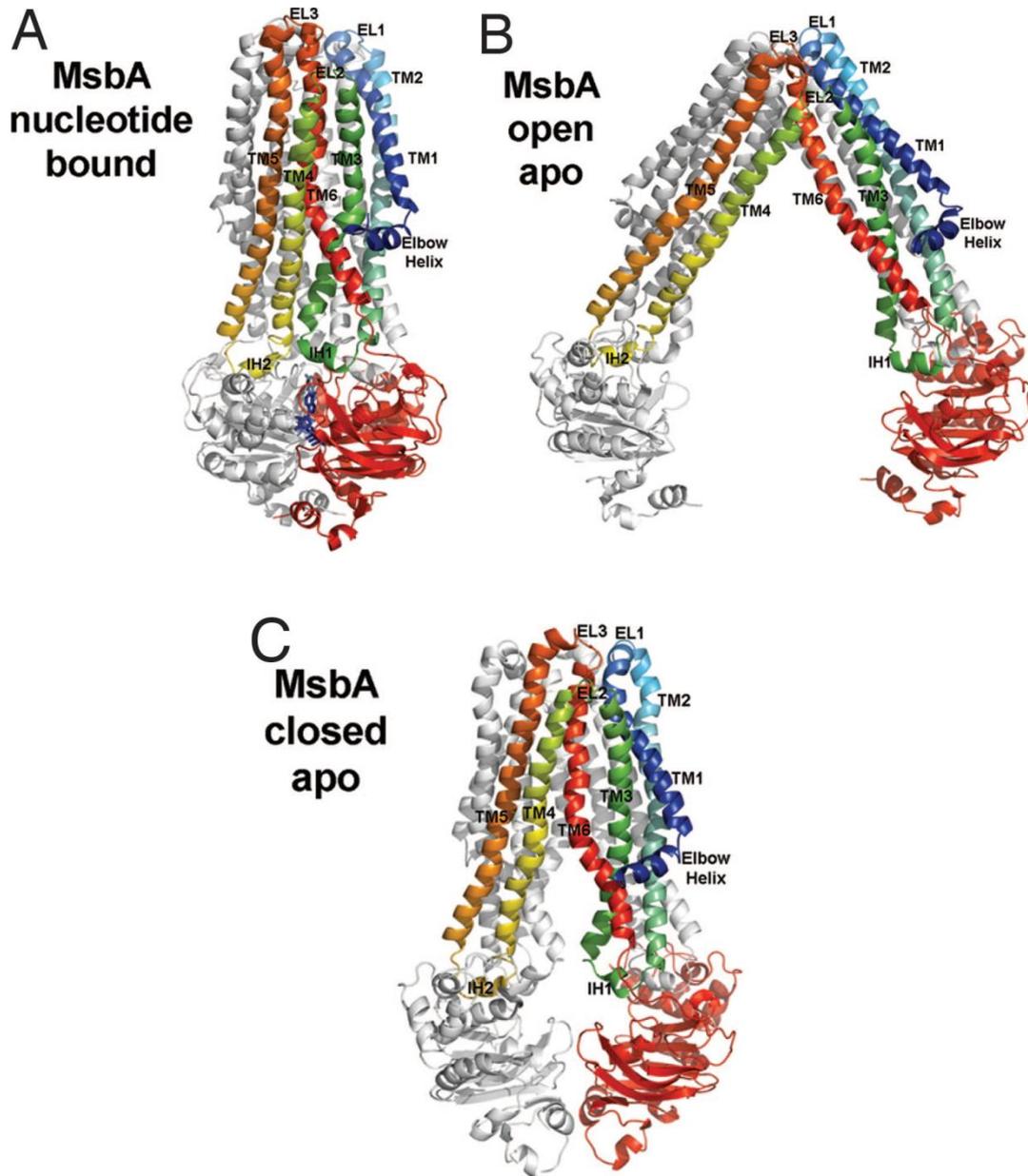
transmis des NBD aux TMD par l'existence d'une interface de transmission. Les domaines TMD contribueraient à cette interface principalement à travers les boucles intracellulaires ICL1 et ICL2. Une hypothèse en accord avec ce qui a été observée à travers des études mutationnelles et de comparaison de séquence qui indiquaient que ICL2 de TAP1:TAP2, tout comme ICL4 de CFTR fourniraient des contacts similaires [89]. Une large cavité observée à l'interface des deux domaines transmembranaires rappelle celle observée dans la conformation *outward-facing* de MDR1 humain [88]. Dawson et Locher ont mis en évidence la prédominance des acides aminés chargés et polaires au niveau de cette cavité tout en suggérant que celle-ci reflèterai une voie d'extrusion pour les substrats puisque l'affinité aux composés hydrophobes sera faible voir absente. A partir de ces résultats, un second modèle de translocation a été proposé différent du modèle largement utilisé et imaginé jusque là (Tableau-A-9) (Figure-A-16) [89].



**Figure-A-16 : Représentation schématique d'un exportateur ABC. [89].** *a.* Le premier modèle de translocation imaginé avec les deux sous-unités arrangées en face à face. *b.* Représentation schématique du modèle Sav1866 dans une conformation *Outward-facing*. Les carrés gris représentent la localisation de la membrane.

Alors que la structure de l'importateur BtuCD était plutôt en accord avec le modèle décrit dans la Figure-A-16-a, la structure de Sav1866 a suggéré pour la première fois l'association sous forme « d'aile » des hélices transmembranaires tout en étant bien ancrées

dans la membrane (Figure-A-15-a, Figure-A-16-b). L'architecture de Sav1866 a permis d'orienter les travaux de Ward *et al.* [90] dans une « resoumission » de nouvelles structures du transporteur bactérien MsbA (Tableau-A-9 et Figure-A-17)



**Figure-A-17 : La structure de MsbA dans trois conformations différentes [90].** *A- Nucleotide bound. B- Open apo. C- Closed apo. Pour chaque conformation, seul un monomère est coloré selon un gradient arc en ciel (N-terminus en bleu et C-terminus en rouge). L'autre monomère est en blanc. Les molécules AMPPNP sont représentées en bâtonnets bleus dans la structure nucleotide-bound.*

Ces conformations correspondent à quatre structures déposées dans la PDB et déterminées par diffraction aux rayons X (Tableau-A-9) : deux structures en présence d'un nucléotide (AMPNP et ADP-Vanadate) et deux structures en absence de nucléotide. Grâce à une comparaison des diverses conformations, Ward *et al.* ont révélé la flexibilité de la charnière formée par les deux boucles extracellulaires EL2 et 3 (Figure-A-17). Cette charnière permettrait aux domaines NBD de se dissocier grâce à la mobilité du « complexe » TM4/TM5/IH2 (Figure-A-17). Pour expliquer le changement conformationnel observé, Ward *et al.* suggèrent aussi qu'une liaison des nucléotides au niveau des domaines NBD causerait le réarrangement des hélices transmembranaires et changerait l'accessibilité du transporteur d'une conformation ouverte vers le milieu cytoplasmique (*inward-facing*) vers une conformation ouverte vers le milieu extracellulaire (*outward-facing*). Ces résultats sont les premiers à représenter structurellement le cycle de translocation supposé pour les transporteurs ABC (Figure-A-5). Il faut, néanmoins, remarquer que seule la structure dans la conformation *outward-facing* liée au nucléotide AMPPNP représente une structure d'une assez haute résolution (3.8 Å) et a guidé la reconstruction des trois autres structures parues. Les deux structures *apo* ne représentent que des traces des C $\alpha$ .

Les structures *apo* de MsbA sont les premières structures décrivant un exportateur ABC dans les conformations *open* et *closed inward-facing*. La structure de l'importateur ABC HI1470/1 [91] dans une conformation *inward-facing* était parue quelques mois avant ces structures MsbA mais les architectures structurales des importateurs et des exportateurs ABC semblent assez différentes [92]. MsbA et Sav1866 ne présentent pas de homologie structurale avec des importateurs ABC plus disponibles dans la PDB (Tableau-A-9). En effet, les arrangements des domaines TMD semblent être très différents entre les importateurs et les exportateurs ABC. Cependant, les architectures structurales de Sav1866 et de MsbA (2HYD et 3B60) en complexe avec un nucléotide semblent décrire une conformation *outward-facing* commune. La P-gp humaine présente un nombre d'hélices transmembranaires égal à celui de ces transporteurs bactériens dans leur état homodimérique. La P-gp possède aussi une identité de séquence raisonnablement élevée avec Sav1866 et MsbA (respectivement à 28 et 31%), qui permet de lui attribuer un repliement structural identique. Cette identité de séquence est encore plus élevée s'agissant des domaines les plus conservés c'est-à-dire les domaines NBD (50% d'identité de séquence avec le domaine NBD de Sav1866) [92]. Cette homologie structurale a aussi été mise en évidence grâce à des approches de *cross-linking* chimique [92] ainsi qu'à partir de cartes de densité électronique [93]. L'ensemble de ces critères a soutenu des travaux de modélisation par homologie du MDR1 humain [94]. O'Mara et Tieleman ont

proposé trois modèles de la P-gp humaine dans trois conformations différentes : une conformation *nucleotide-bound* basée sur le modèle de Sav1866 (Figure-A-15) et deux conformations *apo* nommées *semi-open* et *open* basées sur une reconstruction de domaines NBD avec comme modèles des structures de la sous-unité ATPase du transporteur de maltose d'*E. coli* (MalK) ne liant pas l'ATP [95].

#### II.4.2.4. *Structure quaternaire*

Dans la littérature, il existe plusieurs arguments expérimentaux en faveur d'une organisation dimérique de la P-gp [96-97]. Bien que la mise en évidence de la fonctionnalité de la P-gp se soit faite dans ses états homodimériques, on ignore si ce type d'association est indispensable pour que la P-gp soit fonctionnelle dans les conditions physiologiques. Les données actuelles de structure tridimensionnelle de la P-gp ont permis de révéler certains états conformationnels sans privilégier un état par rapport à un autre. Cependant, il a été montré qu'une expression séparée des moitiés N- et C- terminales d'un monomère de la P-gp ne fait pas perdre à la protéine son activité d'hydrolyse de l'ATP basale alors que leur co-expression simultanée est indispensable à la restauration de l'activation de la P-gp par ses substrats de transport [53].

### III. Objectifs de la thèse

Depuis la découverte du premier transporteur ABC en 1976 (ABCB1), de nombreuses études ont porté sur le fonctionnement moléculaire de ces transporteurs ABC, du fait de leur rôle déterminant dans la résistance aux traitements chimio-thérapeutiques. Ils ont aussi focalisé l'attention de la communauté scientifique parce qu'ils sont présents dans tous les organismes, et qu'ils remplissent probablement d'autres fonctions que le transport d'une grande variété de substrats.

Leur fonction de transporteur leur confère d'importants rôles dans l'absorption, la distribution, et l'excrétion des composés pharmacologiques. Pour cette fonction, ils se trouvent ainsi impliqués dans certaines maladies héréditaires et physiopathologies graves. Pour comprendre aujourd'hui le fonctionnement intime d'une protéine, la résolution de sa structure est primordiale. Dans le cas de la P-gp, l'approche structurale est d'autant plus cruciale que ce transporteur est le siège de changements conformationnels de grande ampleur qu'il faut caractériser. Cependant, 35 ans après sa découverte et bien que de multiples études lui aient été consacrées, de nombreux points restent à préciser.

Mon travail de thèse s'est articulé autour des 4 priorités suivantes, pour apporter de nouveaux éléments à une vision moléculaire du fonctionnement de la P-glycoprotéine :

- 1) Etudier le lien pouvant exister entre la phylogénie des transporteurs ABC et leur mécanisme de reconnaissance de substrats.
- 2) Réaliser des modèles tridimensionnels de la P-gp dans trois conformations catalytiques différentes pouvant expliquer le déroulement du cycle de transport de cette protéine d'intérêt.
- 3) Localiser les résidus, impliqués dans le changement de résistance chez la P-gp, vis-à-vis de la bicouche lipidique.
- 4) Localiser des sous-sites de liaison de substrats connus de la P-gp par le biais de travaux de docking réalisés sur les modèles tridimensionnels construits par homologie.

## IV. Bibliographie

1. Higgins, C.F., *ABC transporters: from microorganisms to man*. Annu Rev Cell Biol, 1992. **8**: p. 67-113.
2. Sharom, F.J., *The P-glycoprotein efflux pump: how does it transport drugs?* J Membr Biol, 1997. **160**(3): p. 161-75.
3. Dean, M., A. Rzhetsky, and R. Allikmets, *The human ATP-binding cassette (ABC) transporter superfamily*. Genome Res, 2001. **11**(7): p. 1156-66.
4. Dean, M. and T. Annilo, *Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates*. Annu Rev Genomics Hum Genet, 2005. **6**: p. 123-42.
5. Walker, E.J. and J.W. Dow, *Location and properties of two isoenzymes of cardiac adenylate kinase*. Biochem J, 1982. **203**(2): p. 361-9.
6. Szakacs, G., et al., *Targeting multidrug resistance in cancer*. Nat Rev Drug Discov, 2006. **5**(3): p. 219-34.
7. Chen, C.J., et al., *Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells*. Cell, 1986. **47**(3): p. 381-9.
8. Stefkova, J., R. Poledne, and J.A. Hubacek, *ATP-binding cassette (ABC) transporters in human metabolism and diseases*. Physiol Res, 2004. **53**(3): p. 235-43.
9. Schuierer, M.M. and T. Langmann, *Molecular diagnosis of ATP-binding cassette transporter-related diseases*. Expert Rev Mol Diagn, 2005. **5**(5): p. 755-67.
10. Dean, M., *ABC transporters, drug resistance, and cancer stem cells*. J Mammary Gland Biol Neoplasia, 2009. **14**(1): p. 3-9.
11. Brooks-Wilson, A., et al., *Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency*. Nat Genet, 1999. **22**(4): p. 336-45.
12. Rust, S., et al., *Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1*. Nat Genet, 1999. **22**(4): p. 352-5.
13. Shulenin, S., et al., *ABCA3 gene mutations in newborns with fatal surfactant deficiency*. N Engl J Med, 2004. **350**(13): p. 1296-303.
14. Chen, Z.J., et al., *Association of ABCA2 expression with determinants of Alzheimer's disease*. FASEB J, 2004. **18**(10): p. 1129-31.
15. Mace, S., et al., *ABCA2 is a strong genetic risk factor for early-onset Alzheimer's disease*. Neurobiol Dis, 2005. **18**(1): p. 119-25.
16. Juliano, R.L. and V. Ling, *A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants*. Biochim Biophys Acta, 1976. **455**(1): p. 152-62.
17. Carlsen, S.A., J.E. Till, and V. Ling, *Modulation of membrane drug permeability in Chinese hamster ovary cells*. Biochim Biophys Acta, 1976. **455**(3): p. 900-12.
18. Riordan, J.R., et al., *Amplification of P-glycoprotein genes in multidrug-resistant mammalian cell lines*. Nature, 1985. **316**(6031): p. 817-9.
19. Abele, R. and R. Tampe, *The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing*. Physiology (Bethesda), 2004. **19**: p. 216-24.
20. Zhang, F., et al., *Characterization of ABCB9, an ATP binding cassette protein associated with lysosomes*. J Biol Chem, 2000. **275**(30): p. 23287-94.
21. Kispal, G., et al., *The ABC transporter Atm1p is required for mitochondrial iron homeostasis*. FEBS Lett, 1997. **418**(3): p. 346-50.
22. Strautnieks, S.S., et al., *A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis*. Nat Genet, 1998. **20**(3): p. 233-8.
23. Deleuze, J.F., et al., *Defect of multidrug-resistance 3 gene expression in a subtype of progressive familial intrahepatic cholestasis*. Hepatology, 1996. **23**(4): p. 904-8.

24. Wada, M., et al., *Mutations in the canalicular multispecific organic anion transporter (cMOAT) gene, a novel ABC transporter, in patients with hyperbilirubinemia II/Dubin-Johnson syndrome*. Hum Mol Genet, 1998. **7**(2): p. 203-7.
25. Ritter, C.A., et al., *Cellular export of drugs and signaling molecules by the ATP-binding cassette transporters MRP4 (ABCC4) and MRP5 (ABCC5)*. Drug Metab Rev, 2005. **37**(1): p. 253-78.
26. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA*. Science, 1989. **245**(4922): p. 1066-73.
27. Shani, N. and D. Valle, *Peroxisomal ABC transporters*. Methods Enzymol, 1998. **292**: p. 753-76.
28. Mosser, J., et al., *Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters*. Nature, 1993. **361**(6414): p. 726-30.
29. Zimmerman, C., et al., *Identification of a host protein essential for assembly of immature HIV-1 capsids*. Nature, 2002. **415**(6867): p. 88-92.
30. Bisbal, C., et al., *Cloning and characterization of a RNase L inhibitor. A new component of the interferon-regulated 2-5A pathway*. J Biol Chem, 1995. **270**(22): p. 13308-17.
31. Chen, Z.Q., et al., *The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors*. J Biol Chem, 2006. **281**(11): p. 7452-7.
32. Tsuda, H., et al., *Identification of overexpression and amplification of ABCF2 in clear cell ovarian adenocarcinomas by cDNA microarray analyses*. Clin Cancer Res, 2005. **11**(19 Pt 1): p. 6880-8.
33. L'Esperance, S., et al., *Gene expression profiling of paired ovarian tumors obtained prior to and following adjuvant chemotherapy: molecular signatures of chemoresistant tumors*. Int J Oncol, 2006. **29**(1): p. 5-24.
34. Biedler, J.L. and H. Riehm, *Cellular resistance to actinomycin D in Chinese hamster cells in vitro: cross-resistance, radioautographic, and cytogenetic studies*. Cancer Res, 1970. **30**(4): p. 1174-84.
35. Gutmann, H., et al., *Transport of amentoflavone across the blood-brain barrier in vitro*. Planta Med, 2002. **68**(9): p. 804-7.
36. Dano, K., *Active outward transport of daunomycin in resistant Ehrlich ascites tumor cells*. Biochim Biophys Acta, 1973. **323**(3): p. 466-83.
37. Inaba, M., et al., *Active efflux of daunorubicin and adriamycin in sensitive and resistant sublines of P388 leukemia*. Cancer Res, 1979. **39**(6 Pt 1): p. 2200-3.
38. See, Y.P., et al., *Increased drug permeability in Chinese hamster ovary cells in the presence of cyanide*. Biochim Biophys Acta, 1974. **373**(2): p. 242-52.
39. Inaba, M., et al., *Restored in vitro sensitivity of adriamycin- and vincristine-resistant P388 leukemia with reserpine*. Biochem Pharmacol, 1981. **30**(15): p. 2191-4.
40. Gottesman, M.M. and I. Pastan, *Biochemistry of multidrug resistance mediated by the multidrug transporter*. Annu Rev Biochem, 1993. **62**: p. 385-427.
41. Ishikawa, T., *The ATP-dependent glutathione S-conjugate export pump*. Trends Biochem Sci, 1992. **17**(11): p. 463-8.
42. Leveille-Webster, C.R. and I.M. Arias, *The biology of the P-glycoproteins*. J Membr Biol, 1995. **143**(2): p. 89-102.
43. Thiebault, J.B., et al., *The role of neurosurgery in the non-specific treatment of prostatic cancer pain*. Prog Clin Biol Res, 1987. **243B**: p. 493-6.
44. Cordon-Cardo, C., et al., *Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites*. Proc Natl Acad Sci U S A, 1989. **86**(2): p. 695-8.
45. Schuetz, E.G., K.N. Furuya, and J.D. Schuetz, *Interindividual variation in expression of P-glycoprotein in normal human liver and secondary hepatic neoplasms*. J Pharmacol Exp Ther, 1995. **275**(2): p. 1011-8.

46. Rao, V.V., et al., *Choroid plexus epithelial expression of MDR1 P glycoprotein and multidrug resistance-associated protein contribute to the blood-cerebrospinal-fluid drug-permeability barrier*. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3900-5.
47. Vautier, S., et al., *ABCB1 (P-glycoprotein) and blood-brain barrier: role in neurological diseases and their treatments*. Journal de pharmacie clinique, 2006. **25**(4): p. 11.
48. Mourez, M., et al., *Rôle, fonctionnement et structure des transporteurs à ATP binding cassette (ABC)*. M/S médecine sciences, 2000. **16**(3): p. 9.
49. Sarkadi, B., et al., *Expression of the human multidrug resistance cDNA in insect cells generates a high activity drug-stimulated membrane ATPase*. J Biol Chem, 1992. **267**(7): p. 4854-8.
50. Horio, M., M.M. Gottesman, and I. Pastan, *ATP-dependent transport of vinblastine in vesicles from human multidrug-resistant cells*. Proc Natl Acad Sci U S A, 1988. **85**(10): p. 3580-4.
51. Cornwell, M.M., et al., *ATP-binding properties of P glycoprotein from multidrug-resistant KB cells*. FASEB J, 1987. **1**(1): p. 51-4.
52. Loo, T.W. and D.M. Clarke, *Covalent modification of human P-glycoprotein mutants containing a single cysteine in either nucleotide-binding fold abolishes drug-stimulated ATPase activity*. J Biol Chem, 1995. **270**(39): p. 22957-61.
53. Loo, T.W. and D.M. Clarke, *Reconstitution of drug-stimulated ATPase activity following co-expression of each half of human P-glycoprotein as separate polypeptides*. J Biol Chem, 1994. **269**(10): p. 7750-5.
54. Callaghan, R., R.C. Ford, and I.D. Kerr, *The translocation mechanism of P-glycoprotein*. FEBS Lett, 2006. **580**(4): p. 1056-63.
55. Azzaria, M., E. Schurr, and P. Gros, *Discrete mutations introduced in the predicted nucleotide-binding sites of the mdr1 gene abolish its ability to confer multidrug resistance*. Mol Cell Biol, 1989. **9**(12): p. 5289-97.
56. Borgnia, M.J., G.D. Eytan, and Y.G. Assaraf, *Competition of hydrophobic peptides, cytotoxic drugs, and chemosensitizers on a common P-glycoprotein pharmacophore as revealed by its ATPase activity*. J Biol Chem, 1996. **271**(6): p. 3163-71.
57. Johnstone, R.W., A.A. Ruefli, and M.J. Smyth, *Multiple physiological functions for multidrug transporter P-glycoprotein?* Trends Biochem Sci, 2000. **25**(1): p. 1-6.
58. Garrigues, A., et al., *Characterization of two pharmacophores on the multidrug transporter P-glycoprotein*. Mol Pharmacol, 2002. **62**(6): p. 1288-98.
59. Higgins, C.F. and M.M. Gottesman, *Is the multidrug transporter a flippase?* Trends Biochem Sci, 1992. **17**(1): p. 18-21.
60. Ferte, J., *Analysis of the tangled relationships between P-glycoprotein-mediated multidrug resistance and the lipid phase of the cell membrane*. Eur J Biochem, 2000. **267**(2): p. 277-94.
61. Benet, L.Z., C.L. Cummins, and C.Y. Wu, *Transporter-enzyme interactions: implications for predicting drug-drug interactions from in vitro data*. Curr Drug Metab, 2003. **4**(5): p. 393-8.
62. Germann, U.A., *P-glycoprotein--a mediator of multidrug resistance in tumour cells*. Eur J Cancer, 1996. **32A**(6): p. 927-44.
63. Loo, T.W. and D.M. Clarke, *Cross-linking of human multidrug resistance P-glycoprotein by the substrate, tris-(2-maleimidoethyl)amine, is altered by ATP hydrolysis. Evidence for rotation of a transmembrane helix*. J Biol Chem, 2001. **276**(34): p. 31800-5.
64. Loo, T.W., M.C. Bartlett, and D.M. Clarke, *Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein*. J Biol Chem, 2003. **278**(41): p. 39706-10.
65. Shilling, R.A., et al., *New light on multidrug binding by an ATP-binding-cassette transporter*. Trends Pharmacol Sci, 2006. **27**(4): p. 195-203.
66. Hanna, M., et al., *Mutagenesis of transmembrane domain 11 of P-glycoprotein by alanine scanning*. Biochemistry, 1996. **35**(11): p. 3625-35.
67. Kwan, T. and P. Gros, *Mutational analysis of the P-glycoprotein first intracellular loop and flanking transmembrane domains*. Biochemistry, 1998. **37**(10): p. 3337-50.

68. Bailly, J.D., et al., *Lack of correlation between expression and function of P-glycoprotein in acute myeloid leukemia cell lines*. *Leukemia*, 1995. **9**(5): p. 799-807.
69. Goldstein, L.J., et al., *Expression of a multidrug resistance gene in human cancers*. *J Natl Cancer Inst*, 1989. **81**(2): p. 116-24.
70. Marie, J.P., *Le phénomène de résistance multiple aux anti-cancéreux: les gènes MDR et la P-gp*. *M/S médecine sciences*, 2000. **6**(5): p. 6.
71. Marzolini, C., et al., *Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance*. *Clin Pharmacol Ther*, 2004. **75**(1): p. 13-33.
72. Kartner, N., et al., *Detection of P-glycoprotein in multidrug-resistant cell lines by monoclonal antibodies*. *Nature*, 1985. **316**(6031): p. 820-3.
73. Bosch, I. and J. Croop, *P-glycoprotein multidrug resistance and cancer*. *Biochim Biophys Acta*, 1996. **1288**(2): p. F37-54.
74. Gottesman, M.M., et al., *Genetic analysis of the multidrug transporter*. *Annu Rev Genet*, 1995. **29**: p. 607-49.
75. Endicott, J.A. and V. Ling, *The biochemistry of P-glycoprotein-mediated multidrug resistance*. *Annu Rev Biochem*, 1989. **58**: p. 137-71.
76. Gottesman, M.M. and I. Pastan, *Resistance to multiple chemotherapeutic agents in human cancer cells*. *Trends Pharmacol Sci*, 1988. **9**(2): p. 54-8.
77. Kast, C., et al., *Membrane topology of P-glycoprotein as determined by epitope insertion: transmembrane organization of the N-terminal domain of mdr3*. *Biochemistry*, 1995. **34**(13): p. 4402-11.
78. Kast, C., et al., *Transmembrane organization of mouse P-glycoprotein determined by epitope insertion and immunofluorescence*. *J Biol Chem*, 1996. **271**(16): p. 9240-8.
79. Vigano, C., et al., *Structural and functional asymmetry of the nucleotide-binding domains of P-glycoprotein investigated by attenuated total reflection Fourier transform infrared spectroscopy*. *J Biol Chem*, 2002. **277**(7): p. 5008-16.
80. Dong, M., et al., *Secondary structure of P-glycoprotein investigated by circular dichroism and amino acid sequence analysis*. *Biochim Biophys Acta*, 1998. **1371**(2): p. 317-34.
81. Locher, K.P., A.T. Lee, and D.C. Rees, *The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism*. *Science*, 2002. **296**(5570): p. 1091-8.
82. Hung, L.W., et al., *Crystal structure of the ATP-binding subunit of an ABC transporter*. *Nature*, 1998. **396**(6712): p. 703-7.
83. Chang, G. and C.B. Roth, *Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters*. *Science*, 2001. **293**(5536): p. 1793-800.
84. Higgins, C.F. and K.J. Linton, *Structural biology. The xyz of ABC transporters*. *Science*, 2001. **293**(5536): p. 1782-4.
85. Rosenberg, M.F., et al., *Structure of the multidrug resistance P-glycoprotein to 2.5 nm resolution determined by electron microscopy and image analysis*. *J Biol Chem*, 1997. **272**(16): p. 10685-94.
86. Sharom, F.J., *The P-glycoprotein multidrug transporter: interactions with membrane lipids, and their modulation of activity*. *Biochem Soc Trans*, 1997. **25**(3): p. 1088-96.
87. Rosenberg, M.F., et al., *Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle*. *EMBO J*, 2001. **20**(20): p. 5615-25.
88. Rosenberg, M.F., et al., *Three-dimensional structure of P-glycoprotein: the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state*. *J Biol Chem*, 2005. **280**(4): p. 2857-62.
89. Dawson, R.J. and K.P. Locher, *Structure of a bacterial multidrug ABC transporter*. *Nature*, 2006. **443**(7108): p. 180-5.
90. Ward, A., et al., *Flexibility in the ABC transporter MsbA: Alternating access with a twist*. *Proc Natl Acad Sci U S A*, 2007. **104**(48): p. 19005-10.
91. Pinkett, H.W., et al., *An inward-facing conformation of a putative metal-chelate-type ABC transporter*. *Science*, 2007. **315**(5810): p. 373-7.

92. Zolnerciks, J.K., C. Wooding, and K.J. Linton, *Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein*. *FASEB J*, 2007. **21**(14): p. 3937-48.
93. McDevitt, C.A., et al., *Structural insights into P-glycoprotein (ABCB1) by small angle X-ray scattering and electron crystallography*. *FEBS Lett*, 2008. **582**(19): p. 2950-6.
94. O'Mara, M.L. and D.P. Tieleman, *P-glycoprotein models of the apo and ATP-bound states based on homology with Sav1866 and Malk*. *FEBS Lett*, 2007. **581**(22): p. 4217-22.
95. Oloo, E.O., E.Y. Fung, and D.P. Tieleman, *The dynamics of the MgATP-driven closure of Malk, the energy-transducing subunit of the maltose ABC transporter*. *J Biol Chem*, 2006. **281**(38): p. 28397-407.
96. Naito, M. and T. Tsuruo, *Functionally active homodimer of P-glycoprotein in multidrug-resistant tumor cells*. *Biochem Biophys Res Commun*, 1992. **185**(1): p. 284-90.
97. Jette, L., M. Potier, and R. Beliveau, *P-glycoprotein is a dimer in the kidney and brain capillary membranes: effect of cyclosporin A and SDZ-PSC 833*. *Biochemistry*, 1997. **36**(45): p. 13929-37.

# Etude phylogénétique

# Chapitre B : Etude Phylogénétique

---

## I. Introduction

Dans le chapitre précédent, nous avons évoqué la remarquable conservation de séquences et d'organisation architecturale observée chez les protéines ABC. Au début de mes travaux de thèse, j'ai pu remarquer que peu d'études phylogénétiques avaient été menées pour étudier le lien entre l'évolution et cette conservation observée de séquences. C'est ainsi que j'ai décidé de réaliser ma propre étude phylogénétique sur les protéines ABC humaines dans le but d'analyser, même partiellement, l'histoire évolutive de ces protéines chez l'Homme, en mettant en parallèle phylogénie et signatures distinctives. Cette étude phylogénétique servira également à estimer le degré d'éloignement évolutif entre les ABC bactériennes et les humaine afin de valider l'utilisation des structures 3D de transporteurs ABC bactériens dans la reconstruction de modèles 3D de transporteurs ABC humains.

Intéressé en priorité aux protéines membranaires pour la reconstruction de modèles structuraux des protéines d'efflux, j'ai opté pour la sélection des 5 sous-familles ABC possédant des domaines transmembranaires, excluant ainsi les sous-familles E et F de l'étude. Un tableau contenant les 43 séquences protéiques sélectionnées est présenté dans la section Annexes, Tableau-Ax-1. Ces séquences sont issues des bases de données UniprotKB<sup>2</sup>, une base de données collectant des informations fonctionnelles sur les protéines avec une annotation fiable et précise [1], et TCDB<sup>3</sup>, une base de données regroupant les transporteurs membranaires dans une classification reconnue au sein la communauté scientifique [2].

## II. Analyse des séquences protéiques

Dans une étude phylogénétique, la qualité des résultats obtenus dépend étroitement de la qualité de l'alignement de séquences. L'étape d'alignement est donc cruciale et son

---

<sup>2</sup> <http://www.uniprot.org>

<sup>3</sup> <http://www.tcdb.org/>

optimisation nécessaire du fait de la relativement faible identité de séquences dans certaines régions des transporteurs ABC humains.

Dans le cas des transporteurs ABC, et devant la grande divergence de la taille des séquences entre les transporteurs entiers eux-mêmes et les hémi-transporteurs, une étude phylogénétique se base logiquement sur l'alignement des régions les plus conservées. Ces régions sont, comme décrit dans le précédent chapitre, les domaines nucléotidiques NBD et leurs motifs distinctifs (Figure-A-2). Pour identifier ces domaines au niveau des diverses séquences, j'ai utilisé la banque de données *Prosite*<sup>4</sup>[3]. Il s'agit d'une banque originale considérée comme un dictionnaire qui recense des motifs (ou signature) protéiques ayant une signification biologique. Elle est établie en regroupant, quand cela est possible, les protéines contenues dans UniprotKB par famille comme par exemple les kinases ou les protéases. Dans cette banque de données, le domaine décrivant les trois motifs distinctifs (Walker A, Walker B et Signature « LSGGQ ») de la P-gp est répertorié sous l'identifiant : PS50893. Ce domaine, d'environ 240 acides aminés, est nommé ABC\_transporter\_2 (Figure-B-1).

Dans le but de localiser ce domaine d'intérêt dans les séquences ABC, j'ai procédé à un criblage de ces séquences grâce au programme *Scanprosite*<sup>5</sup>. Le résultat de cette opération est la détection de 70 « hits » représentant les 27 transporteurs entiers qui possèdent deux domaines ABC\_transporter\_2, et les 16 hémi-transporteurs qui ne possèdent qu'un seul domaine ABC\_transporter\_2. Ces « hits », correspondant à des séquences d'une taille moyenne de 234 acides aminés, représentent les domaines NBD présents en deux copies chez les transporteurs entiers et en une seule copie chez les hémi-transporteurs (Figure-A-1). En prenant en compte l'organisation architecturale des transporteurs ABC, j'ai renommé les hits décrivant les transporteurs entiers par une terminaison « \_N » quand il s'agissait d'un domaine localisé du côté N-terminal, et une terminaison « \_C » quand il s'agissait d'un domaine localisé du côté C-terminal (Exemple : ABCA1\_N et ABCA1\_C pour les deux hits correspondant aux deux domaines du transporteur ABCA1). Dans le résultat du crible, figure un second domaine commun à certains transporteurs ABC (ABCB, ABCC et ABCD). Ce domaine, d'environ 276 acides aminés, est répertorié sous l'identifiant : PS50929. Il est nommé ABC\_TM1F (Figure-B-1). Il correspond, quand il est présent, à la quasi-intégralité du domaine transmembranaire.

---

<sup>4</sup> <http://expasy.org/prosite/>

<sup>5</sup> <http://expasy.org/tools/scanprosite/>



**Figure-B-1 : Domaines Prosite identifiés dans la séquence protéique du transporteur ABCD1**

### **III. Alignement des séquences protéiques:**

Tous les alignements de séquences cités dans ce chapitre ont été réalisés avec le programme *Dialign*<sup>6</sup> [4]. Il s'agit d'un programme d'alignement multiple qui effectue une comparaison de segments de séquences, mettant ainsi en évidence des similarités locales entre séquences éloignées. Ces qualités font de ce programme un outil d'alignement adéquat pour l'analyse des séquences protéiques des transporteurs ABC.

Les 70 séquences, issues du criblage réalisé à l'étape précédente, ont été alignées via *Dialign*, puis, dans le but d'améliorer la qualité de l'alignement obtenu, ce dernier a été soumis au programme *Gblocks*<sup>7</sup> version 0.91b [5]. *Gblocks* sert à éliminer les positions les moins bien alignées et les régions les plus divergentes. Le résultat obtenu est présenté Figure-B-2.

Cet alignement de séquences représente les 115 positions les mieux alignées, révélées par *Gblocks*. Ces positions décrivent les 5 blocs les plus conservés de l'alignement de séquences résultant de *Dialign*. Cette étape de sélection des blocs conservés a pour but d'affiner les analyses phylogénétiques.

Il est possible de remarquer que, sur l'alignement de blocs présenté Figure-B-2, les trois blocs représentant les motifs distinctifs des transporteurs ABC sont les mieux alignés. Ce même type d'alignement est celui que j'ai utilisé pour aborder l'étape des analyses phylogénétiques.

<sup>6</sup> <http://bibiserv.techfak.uni-bielefeld.de/dialign/>

<sup>7</sup> [http://molevol.cmima.csic.es/castresana/Gblocks\\_server.html](http://molevol.cmima.csic.es/castresana/Gblocks_server.html)



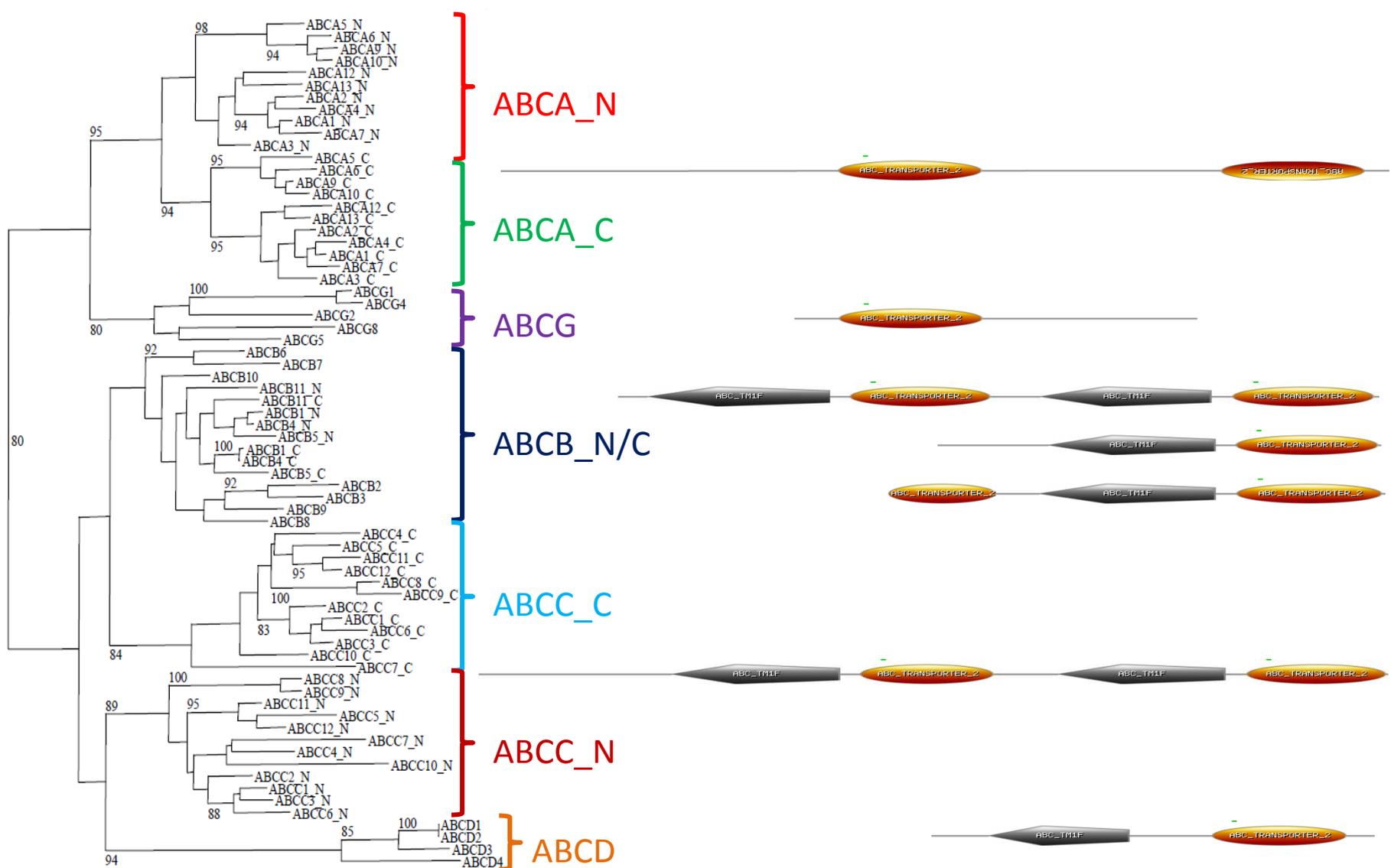
## IV. Analyses phylogénétiques

### IV.1. Méthodologie

Toutes les analyses phylogénétiques ont été réalisées grâce au programme PhyML [7]. Ce programme, largement utilisé par la communauté scientifique, sert à construire des arbres phylogénétiques à partir de séquences nucléotidiques ou protéiques. Il se base sur une méthode d'estimation dite de maximum de vraisemblance "maximum likelihood". Cette méthode évalue la probabilité qu'un modèle évolutif choisi ait donné lieu aux séquences observées. Le maximum de vraisemblance offre plusieurs avantages :

- On suppose que c'est la méthode d'estimation la moins affectée par des erreurs d'échantillonnage.
- Même pour les courtes séquences, comme c'est le cas pour l'alignement issu de *Gblocks*, cette méthode est reconnue comme performante par rapport à d'autres. Ainsi, elle permet une utilisation optimale du modèle d'évolution en tenant compte de l'information contenue dans tous les sites.

Pour tous les arbres phylogénétiques qui seront présentés dans cette partie, j'ai réalisé un test de "*bootstrap*" de 100 répliques. Ce test, proposé dans la fin des années 70 par Bradley Efron [8] a pour but de fournir des indications sur la précision des estimations réalisées. Nous avons ainsi des informations sur la robustesse de l'arbre fourni par PhyML et son pourcentage d'exactitude. En règle générale, si la valeur fournie par le *bootstrap* à une branche donnée est supérieure à 90%, la topologie de la branche est considérée comme fiable. Un premier arbre phylogénétique issu de l'alignement *Gblocks* (Figure-B-2) est présenté Figure-B-3.



**Figure-B-3 : Arbre phylogénétique représentant le lien évolutif entre les moitiés N- et C- terminales des transporteurs ABC entiers et les héli-transporteurs.** Cet arbre de maximum de vraisemblance a été construit sur la base de l'alignement de 70 séquences protéiques (Figure-B-2). Sept « branches » sont possibles: ABCA\_N (Rouge), ABCA\_C (Vert), ABCG (Violet), ABCB\_N/C (Bleu), ABCC\_N (Rouge foncé) ABCC\_C (Bleu ciel) et ABCD (Orange). Seules les valeurs de bootstrap supérieures ou égales à 80% sont montrées. (Les moitiés des 35 transporteurs entiers sont distingués par des terminaisons « \_N » et « \_C » pour respectivement les zones N- et C- terminales, exemple : ABCC8\_N et ABCC8\_C). Les membres des sous-familles ABCE et ABCF, étant non membranaires, n'ont pas fait partie de cette analyse phylogénétique.

## IV.2. Phylogénie des transporteurs ABC humains

L'arbre présenté Figure-B-3 nous suggère fortement qu'il existe un ancêtre commun à tous les transporteurs ABC humains, indépendamment de leur fonction et de leur organisation architecturale. Deux grandes branches sont observées: l'une représentant les transporteurs des sous-familles ABCA et ABCG, l'autre représentant les transporteurs des sous-familles ABCB, ABCC et ABCD. Rappelons que les transporteurs non membranaires des sous-familles ABCE et ABCF ont été exclus de l'étude. Dans le but d'avoir une vision plus large de la robustesse des liens phylogénétiques entre les différentes sous-familles, j'indique sur cet arbre les valeurs de *bootstrap* supérieures à 80%.

Au niveau de la première branche ABCA-ABCG, les séquences représentant les moitiés N- et C- terminales des transporteurs de la sous-famille ABCA sont regroupées dans des clusters indépendants, mais liés par le même nœud. Les membres de la sous-famille ABCA et ABCG se distinguent par la présence du seul domaine ABC\_Transporter\_2. Ce domaine existe en deux copies dans les transporteurs entiers ABCA, et en une seule copie dans les hémi-transporteurs ABCG. Ce premier résultat peut suggérer un phénomène de duplication interne intervenu au cours de l'histoire évolutive des membres de la sous-famille ABCA, après leur séparation de la sous-famille G. Au sein de cette sous-famille, les séquences sont regroupées d'une manière quasi identique en N- comme en C- terminal. Ainsi, on distingue deux sous-clusters représentant d'un côté les transporteurs ABCA-5, 6, 9 et 10, de l'autre les transporteurs ABCA-12, 13, 2, 4, 1, 7 et 3. En confrontant ce résultat à la taille de séquences des différents transporteurs ABCA, on peut constater qu'hormis ABCA3, tous les transporteurs du second sous-cluster ont une taille de séquences de plus de 2000 acides aminés. Les transporteurs du premier sous-cluster sont plus courts. Au niveau de la sous-famille ABCG, on retrouve deux sous-clusters, avec les membres fonctionnant en homodimère d'un côté (ABCG-1, 2 et 4), ceux fonctionnant en hétérodimère de l'autre (ABCG-5 et 8). Il est possible, dans ce cas, de réaliser un parallèle entre fonctionnement et phylogénie.

Au niveau de la seconde branche ABCB-ABCC-ABCD, on retrouve regroupés, les transporteurs possédant à la fois le domaine ABC\_TM1F et le domaine ABC\_Transporter\_2. Pour la sous-famille ABCB, composée de 4 transporteurs entiers (ABCB-1, 4, 5, 11) et de 7 hémi-transporteurs (ABCB-2, 3, 6, 7, 8, 9, 10), on remarque que les deux sous-groupes sont liés à des branches indépendantes. Les isoformes 1, 4 et 11 possèdent une architecture avec deux domaines ABC\_TM1F et deux domaines ABC\_Transporter\_2, alors que l'isoforme 5

présente une architecture unique, avec deux domaines ABC\_Transporter\_2 mais un seul domaine ABC\_TM1F présent du côté C-terminal. Il faut néanmoins remarquer qu'ABCB5 est le seul transporteur ABC entier avec une séquence de 812 acides aminés seulement, ce qui laisse penser à une perte d'un domaine ABC\_TM1F au cours de l'évolution.

De leur côté, les hémi-transporteurs ABCB forment des sous-clusters indépendants. Ainsi, les membres TAP1 (ABCB2) et TAP2 (ABCB3) sont liés à une seule branche, avec une valeur de *bootstrap* de 92% reflétant sa bonne robustesse. Pour rappel, TAP1 et TAP2 forment un hétérodimère fonctionnel. Sur une branche voisine aux TAP, on trouve ABCB9, un hémi-transporteur qui leur est homologue.

Sous cette même grande branche, on remarque que les membres de la sous-famille ABCC avec ses séquences représentant la moitié N-terminale, sont plus proches des séquences de la sous-famille ABCD, et que celles représentant la moitié C-terminale sont plus proches des séquences de la sous-famille ABCB. Les valeurs de *bootstrap* assez élevées (84 et 89%) dans les branches respectives de ces deux ensembles de séquences indiquent la robustesse de cette partie de l'arbre. Le domaine transmembranaire optionnel (TMD0) qui caractérise la plupart des membres de cette sous-famille serait probablement à l'origine de cette particularité. Par contre, on distingue, comme pour les autres sous-familles, des groupes de séquences assemblés sous le même nœud en N- et C- terminal. C'est le cas par exemple des deux ABCC homologues SUR1 (ABCC8), SUR2 (ABCC9), liés par des branches bien robustes (des valeurs de *bootstrap* = 100%). Ceci suggère une duplication récente pour les deux récepteurs aux sulfonylurées. Cette hypothèse est valable également pour ABCC11 et 12 (valeurs de *bootstrap* de 95%). Ces deux transporteurs sont proches d'un point de vue phylogénétique d'ABCC5. Les transporteurs ABCC-1, 2, 3 et 6 forment de leur côté des sous-clusters bien définis en N- comme en C- terminal.

Les membres de la sous-famille ABCD forment un dernier cluster, avec de bonnes valeurs de *bootstrap* associées à toutes les branches. Les 4 hémi-transporteurs possèdent une architecture semblable, avec un domaine ABC\_TM1F et un domaine ABC\_Transporter\_2. Un autre phénomène de duplication pourrait être suggéré entre les deux homologues ABCD1 et ABCD2.

Cette analyse des relations évolutives entre les transporteurs ABC humains nous permet d'évoquer les hypothèses suivantes :

- Un premier grand phénomène de duplication a probablement eu lieu, donnant naissance à deux grands clusters : l'un contenant les sous-familles ABCA-ABCG,

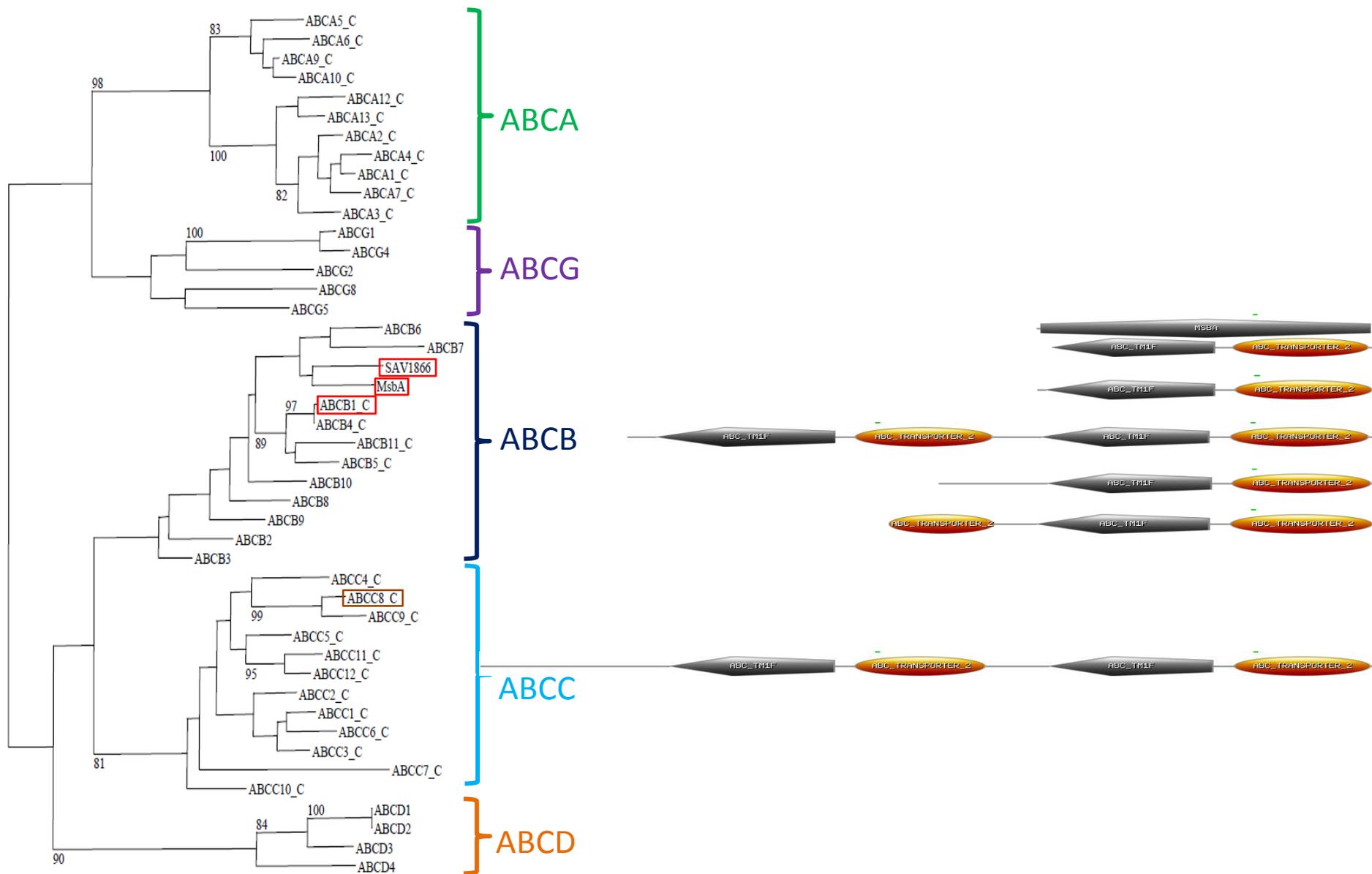
l'autre, les sous familles ABCB-ABCC-ABCD, chacun évoluant séparément ensuite.

- Des phénomènes de duplication interne (c'est-à-dire sans séparation des gènes, donnant des protéines "en tandem") peuvent aussi être suggérés pour expliquer l'histoire évolutive des sous-familles comportant des transporteurs entiers : ABCA-ABCB-ABCC, survenant après le premier phénomène duplication qui a donné naissance aux différentes sous-familles.

### **IV.3. La phylogénie au service de la modélisation**

Nous avons évoqué, dans le chapitre précédent, qu'en l'absence d'informations sur la structure tridimensionnelle de la P-gp, les diverses études structurales menées sur ce transporteur ABC s'étaient jusqu'alors basées sur les modèles ABC bactériens homologues disponibles dans la PDB: MsbA et Sav1866.

Avant de réaliser une étude structurale de la P-glycoprotéine, et au vu des résultats obtenus à l'issue de la première analyse phylogénétique, j'ai décidé d'intégrer Sav1866 et MsbA dans une nouvelle reconstruction phylogénétique des transporteurs ABC. Le but, étant d'étudier si un éventuel lien évolutif existe entre la P-gp et ces deux transporteurs ABC, et si l'intégration de ces transporteurs bactériens dans une étude phylogénétique avec des transporteurs humains modifierait le résultat obtenu pour la précédente étude. Le phénomène de duplication interne observé chez les transporteurs entiers permet de réduire le nombre de séquences incluses dans cette nouvelle analyse en gardant un seul pool des séquences qui leur sont représentatives. Ainsi, j'ai maintenu uniquement les séquences représentatives de la moitié C-Terminale. Pour cette analyse et grâce à *Prosite*, j'ai pu identifier sur les séquences MsbA et Sav1866 les deux domaines ABC\_Transporter\_2 et ABC\_TM1F (Tableau-Ax-1). La suite de la procédure est identique à celle de la première analyse avec un alignement *Dialign* puis un alignement des blocs conservés avec *Gblocks*. L'arbre phylogénétique obtenu est présenté Figure-B-4.



**Figure-B-4 : Arbre phylogénétique représentant la relation entre les transporteurs ABC humains et deux transporteurs bactériens.** Cet arbre de maximum de vraisemblance a été construit sur la base de 45 séquences protéiques (43 humains et 2 bactériens). Cinq « branches » principales sont mis en évidence: ABCA (Vert), ABCG (Violet), ABCB (Bleu), ABCC (Bleu ciel) et ABCD (Orange). P-gp (ABCB1), MsbA et Sav1866 sont repérés dans l'arbre par des cadres rouges, SUR1 (ABCC8) par un cadre marron. Seules les valeurs de bootstrap supérieures ou égales à 80% sont montrées. (Les transporteurs entiers sont représentés par leur moitié C-terminale « \_C », exemple : ABCC8\_C)

Sur l'arbre Figure-B-4, il est clair que la topologie générale de l'arbre présenté Figure-B-3 a été conservée. On distingue, ainsi, les deux grandes branches ABCA-ABCG et ABCB-ABCC-ABCD. Pour les séquences bactériennes intégrées dans l'analyse, elles sont liées à une même branche, au niveau du sous-cluster représentant la sous-famille ABCB. Les branches qui lui sont voisines sont celles liant les transporteurs ABCB-6 et 7, mais aussi celles représentant le groupement ABCB-1, 4, 11 et 5. Alors que les deux transporteurs bactériens contiennent les mêmes domaines que les autres membres de la branche ABCB-ABCC et ABCD, leur localisation au niveau du sous-cluster ABCB traduit un lien phylogénétique plus important avec les transporteurs de la sous-famille ABCB, dont la P-gp. Ce premier résultat va dans le sens d'une homologie architecturale entre ces transporteurs bactériens et notre transporteur d'intérêt.

Dans le cadre d'une collaboration entre notre équipe et une équipe partenaire du CEA de Grenoble (Dr. Michel Vivaudou), j'ai invité à réaliser un modèle tridimensionnel d'un transporteur de la sous-famille ABCC. Il s'agit du transporteur ABCC8 (SUR1). Le but est d'étudier un lien fonctionnel possible entre cette protéine ABC et la P-gp, et de mettre en valeur des résultats obtenus via des études d'enzymologie. En effet, au niveau de l'arbre de la Figure-B-4, SUR1, pour lequel aucune fonction de transport n'a été observée à ce jour, fait partie du sous-cluster ABCC qui est le sous-cluster le plus proche des transporteurs ABCB. Au niveau architectural, on peut noter, qu'hormis le domaine transmembranaire supplémentaire TMD0, l'architecture générale des moitiés N- et C-terminales de ce transporteur est identique à celle des ABCB entiers. Ce lien phylogénétique est de tout premier intérêt pour l'étude structurale menée sur ABCB1 et ABCC8, qui sera présentée dans le prochain chapitre.

La séquence de la P-gp de souris, pour laquelle une structure tridimensionnelle est disponible dans la PDB depuis 2009, n'a pas été intégrée dans l'étude phylogénétique. La P-gp de souris étant déjà reconnue comme orthologue de la P-gp humaine [9].

## V. Bibliographie

1. Bairoch, A., et al., *The Universal Protein Resource (UniProt)*. Nucleic Acids Res, 2005. **33**(Database issue): p. D154-9.
2. Saier, M.H., Jr., C.V. Tran, and R.D. Barabote, *TCDB: the Transporter Classification Database for membrane transport protein analyses and information*. Nucleic Acids Res, 2006. **34**(Database issue): p. D181-6.
3. Hulo, N., et al., *The PROSITE database*. Nucleic Acids Res, 2006. **34**(Database issue): p. D227-30.
4. Morgenstern, B., *DIALIGN: multiple DNA and protein sequence alignment at BiBiServ*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W33-6.
5. Castresana, J., *Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis*. Mol Biol Evol, 2000. **17**(4): p. 540-52.
6. Gouy, M., S. Guindon, and O. Gascuel, *SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building*. Mol Biol Evol, 2010. **27**(2): p. 221-4.
7. Guindon, S. and O. Gascuel, *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood*. Syst Biol, 2003. **52**(5): p. 696-704.
8. Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional conference series in applied mathematics. 1982: Society for Industrial and Applied Mathematics. 38.
9. Dean, M., A. Rzhetsky, and R. Allikmets, *The human ATP-binding cassette (ABC) transporter superfamily*. Genome Res, 2001. **11**(7): p. 1156-66.

# Modélisation Moléculaire

# Chapitre C : Modélisation Moléculaire

---

## I. Introduction

Dans le cadre de la collaboration citée dans le chapitre précédent, des résultats biochimiques, obtenus par Alexia Garrigues, Solenne Martin et Stéphane Orłowski (Saclay) et par Elisabeth Garcia, H el ene Jacquet et Michel Vivaudou (Grenoble), avaient r ev el e la reconnaissance par la P-gp de ligands sp ecifiques de SUR1. Ces deux prot eines ABC n'ont pas de fonction commune, et n'avaient pas de substrats communs connus avant ces travaux. La P-gp et SUR1 partagent, n eanmoins, le caract ere d'une reconnaissance multisp ecifique de substrats. Le fonctionnement complexe de SUR1 dans le cadre de la r egulation de la conductance des canaux  $K_{ATP}$  (Une conductance d ependante  a la fois de la liaison « classique » de nucl eotides aux domaines NBD mais aussi de la liaison de ligands pharmacologiques) a limit e les donn ees fonctionnelles et structurales disponibles sur cette prot eine. La P-gp, quant  a elle, est la prot eine ABC jusqu'ici la plus  etudi ee par la communaut e scientifique.

J'ai consacr e cette partie de mon travail de th ese  a l' etude structurale de la P-gp humaine dans le but de comprendre sa multisp ecificit e de reconnaissance de substrats dans trois  etats catalytiques diff erents « *Open-Inward* », « *Closed-Inward* » et « *Outward-Facing* ». Pour interpr eter les r esultats biochimiques cit es ci-dessous pour comparer les deux prot eines ABC, j'ai  egalement d evelopp e un mod ele tridimensionnel repr esentant SUR1 dans l' etat catalytique « *Outward-Facing* ».

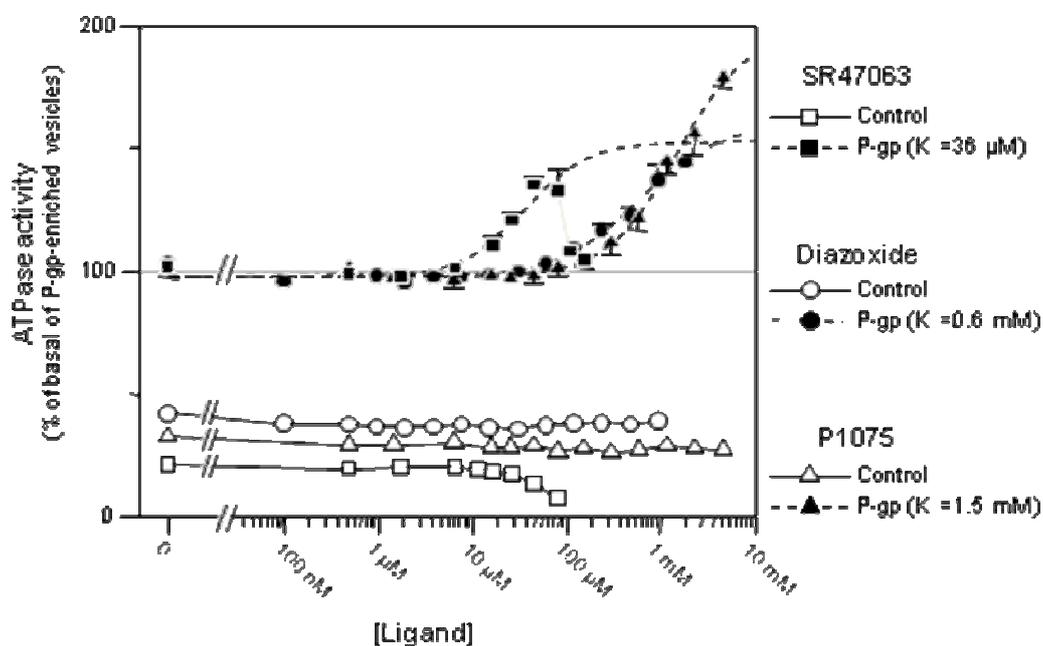
Les r esultats de ce travail ont fait l'objet d'une publication parue en F evrier 2011 dans *The Journal of Biological Chemistry* [1] et qui constituera le c oeur de ce chapitre.

## II. Etude structurale de transporteurs ABC humains

### II.1. Contexte scientifique

Avant d'aborder l'étude structurale de transporteurs ABC, il est impératif de présenter le contexte scientifique dans lequel ce travail a eu lieu. Il s'agit des résultats biochimiques et enzymologiques intéressants obtenus précédemment par nos collaborateurs. Ces résultats ont confirmé le large spectre de reconnaissance de substrats par la P-gp. Ainsi, les mesures des activités ATPasiques dans des vésicules membranaires contenant la P-gp ont révélé la possibilité de leur stimulation par des ouvreurs [Diazoxide, SR47063 (dérivé du cromakalim) et P1075 (dérivé du pinacidil)] ou des bloqueurs (glibenclamide, meglitinide et tolbutamide) des canaux  $K_{ATP}$  (Exemple : Figure-C-1).

A



**B**

	<b>Activité ATPasique stimulée (% de l'activité basale)</b>	<b>K<sub>1/2</sub> (μM)</b>
<b>Verapamil (VRP)</b>	<b>173</b>	<b>1.3</b>
<b>Progestérone (PRG)</b>	<b>166</b>	<b>17</b>
<b>Vinblastine (VBL)</b>	<b>inhibition</b>	<b>31</b>
<b>Glibenclamide (GLB)</b>	<b>160</b>	<b>11</b>
<b>SR47063</b>	<b>155</b>	<b>36</b>

**Figure-C-1 : Effets de la concentration des ligands de SUR sur l'activité ATPasique stimulée de la P-gp.** **A**, Effets des ligands ouvreurs des canaux  $K_{ATP}$  (ligands de SUR1) sur l'activité ATPasique de la P-gp. **B**, Données d'interaction déduites des mesures des effets des ligands de la P-gp (en rouge) et de SUR1 (en bleu) sur l'activité basale de la P-gp. Son indiqués les simulations d'activité ATPasique (exprimées en % de l'activité basale) et les concentrations  $K_{1/2}$  observées pour la moitié de l'effet maximal.

La Figure-C-1 est représentative des résultats révélant l'effet des ouvreurs et des bloqueurs des canaux  $K_{ATP}$  sur l'activité ATPasique de la P-gp. Nous constatons ainsi que des ligands de SUR1 (Figure-C-1B) peuvent stimuler l'activité ATPasique basale de la P-gp à des niveaux comparables à ceux de substrats connus de la P-gp (VBL, PRG, VRP). Un autre volet de cette étude enzymologique a mis en évidence des interactions possibles entre des ligands de SUR1 et des substrats de la P-gp (Voir article [1]).

Les résultats obtenus suggéraient donc des similarités fonctionnelles entre SUR1 et P-gp dans le cadre de leur reconnaissance de substrats et ceci malgré la divergence de leurs fonctions biologiques connues à ce jour. Devant l'absence de structures tridimensionnelles d'une assez haute résolution de ces deux protéines ABC, une interprétation plus détaillée de ces résultats exigeait la construction de modèles tridimensionnels des deux protéines ABC. Une information structurale complémentaire était importante pour interpréter les données d'interaction et mettre en valeur les diverses hypothèses issues des résultats de l'étude enzymologique. Lorsque j'ai entamé ce travail en 2008, aucune structure de P-gp de mammifères n'était disponible dans la PDB. La parution récente d'une structure de P-gp de

souris [2] a permis d'améliorer la construction des modèles de P-gp humaine mais n'a pas modifié notre stratégie globale de modélisation qui s'appuie aussi sur les structures homologues disponibles. Il faut préciser, qu'à ce jour, il n'existe pas de structure de SUR1 dans la PDB ni d'un autre membre de la famille ABCC.

Grâce aux analyses phylogénétiques, premier volet de mon travail doctoral, nous avons opté pour une reconstruction par homologie de modèles tridimensionnelles de protéines ABC humaines (P-gp et SUR1) serait possible en se basant sur les structures cristallographiques des protéines ABC bactériennes (Sav1866 et MsbA, les seules structures 3D disponibles en début de thèse).

## **II.2. Analyse de la topologie membranaire des transporteurs ABC**

### **II.2.1. ABCB1 et ABCC8 : des topologies membranaires similaires**

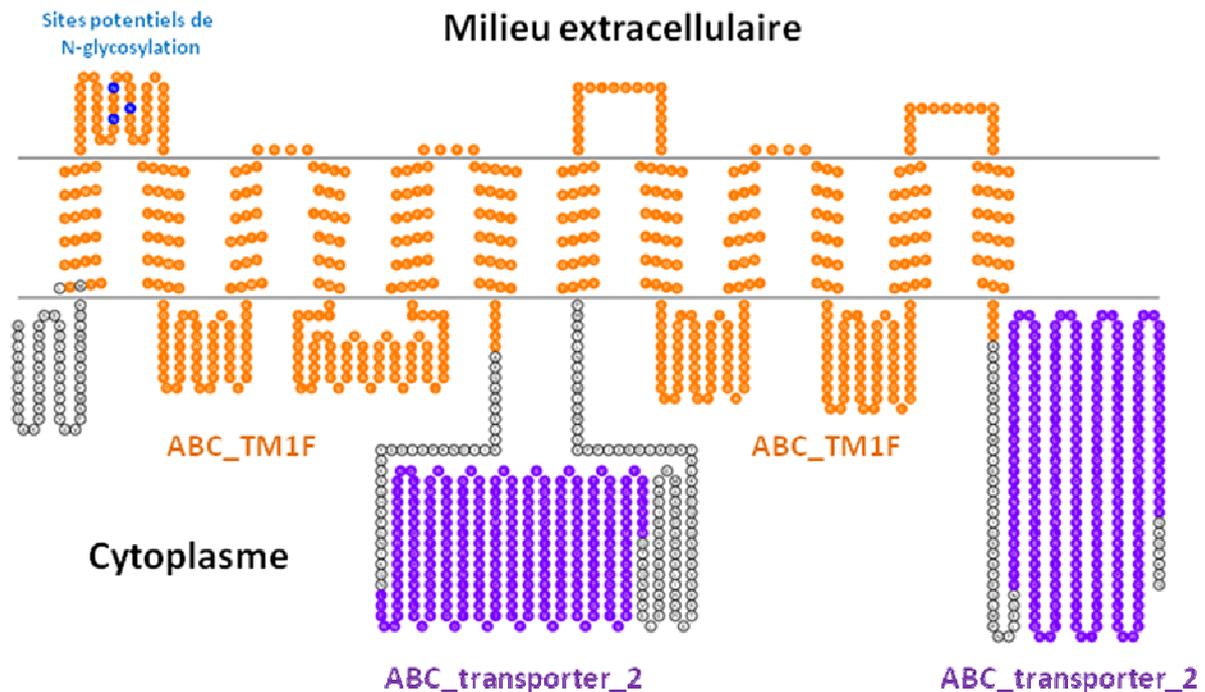
Nous avons évoqué dans le chapitre Introduction la multitude des études réalisées dans le but de décrire la topologie membranaire de la P-gp. Devant les progrès et les améliorations qu'ont connu les programmes de prédiction des structures secondaires de manière générale et ceux des hélices transmembranaires de manière spécifique, nous avons décidé de soumettre les séquences protéiques de la P-gp et de SUR1 à une analyse topologique par un programme de prédiction spécialisé (Voir Tableau-Ax-1 pour les numéros d'accession). Il s'agit du programme HMMTOP (v2.0)<sup>8</sup> [3]. La spécificité de ce programme est qu'il opère en se basant sur l'hypothèse que la topologie et la localisation des segments transmembranaires est déterminée par la différence dans les distributions des acides aminés dans diverses parties structurales d'une protéine plus que par la composition spécifique en acides aminés de ces parties. Cinq zones structurales ont été définies dans les protéines membranaires: les hélices membranaire (H), les terminaisons des hélices en intra et en extracellulaire (i et o), les boucles intra ou extracellulaire (I et O). Deux terminaisons connectées forment une courte boucle associée à la membrane, tandis que l'architecture "terminaison-boucle-terminaison" forme une longue boucle dans le cytosol ou dans le milieu extra-cytosolique. La topologie est déterminée en partitionnant la séquence protéique de manière à produire les fréquences

---

<sup>8</sup> <http://www.enzim.hu/hmmtop/>

relatives aux acides aminés de ces zones structurales le long de la séquence doivent être maximales. Les fichiers de sortie sont présentés (Figure-Ax-1).

Pour une meilleure interprétation et une meilleure lisibilité de ces fichiers, j'ai visualisé le résultat en utilisant le programme TOPO2<sup>9</sup>. La topologie structurale prédite de la P-gp est ainsi schématisée (Figure-C-2).



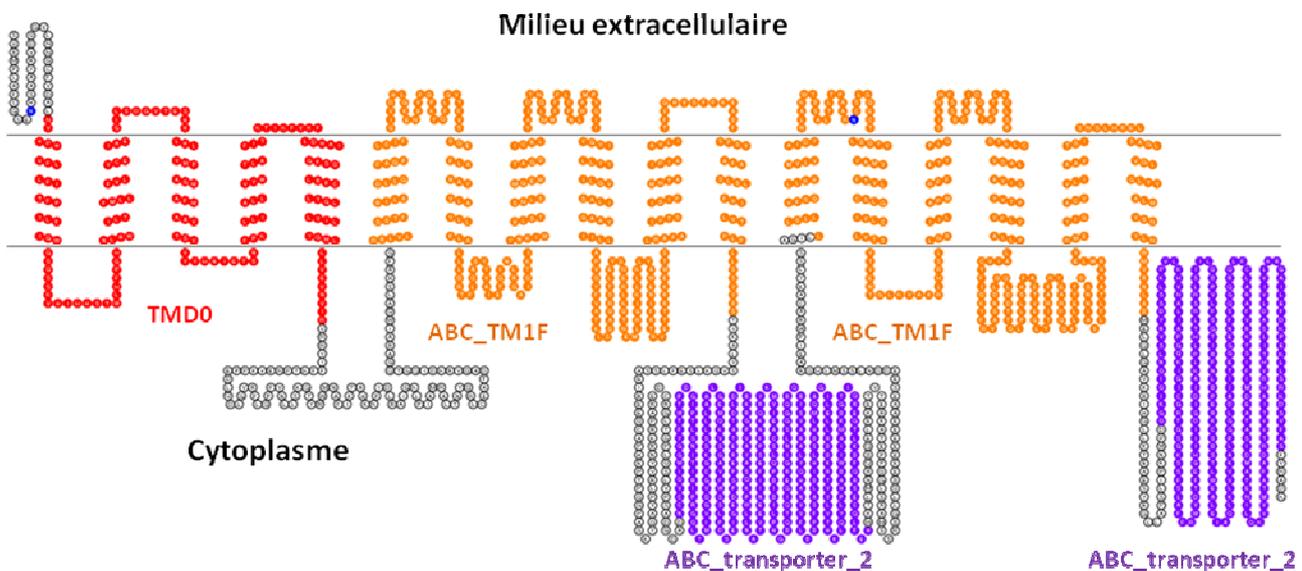
**Figure-C-2 : Topologie structurale de la P-gp déterminée par HMMTOP et visualisée par TOPO2.** On distingue 12 hélices transmembranaires et deux domaines NBD. Les 2 motifs ABC\_TM1F localisés entre les acides aminés 52-357 et 712-1000 sont représentés en orange et les 2 motifs ABC\_transporter\_2 localisés entre les acides aminés 392-628 et 1035-1273 sont représentés en violet. Les 3 sites potentiels des N-glycosylations, 91-94-99, sont représentés en bleu.

Grâce à la prédiction par HMMTOP de la structure topologique de la P-gp, on retrouve bien les 12 hélices caractéristiques des domaines TMD des protéines ABC. Ces hélices sont respectivement localisées entre les acides aminés : 49-73, 117-141, 190-209, 214-233, 299-323, 328-352, 710-733, 754-778, 833-852, 857-874, 935-957 et 974-996 (voir Figure-Ax-1). HMMTOP a également détecté les sites potentiels de glycolysation de la P-gp situés au niveau des résidus Arg 91, 94 et 99. Si on se base sur la structure cristallographique

<sup>9</sup> <http://www.sacs.ucsf.edu/TOPO2/>

de la P-gp de souris, proche à 87% d'identité de séquence de la P-gp humaine, on constate une excellente adéquation entre les segments TM prédits et la topologie observée. L'efficacité de cette méthode de prédiction des segments transmembranaires est valorisée, non seulement par la qualité de ses résultats mais également par le fait qu'en comparaison avec une méthode « classique », basée sur l'étude des profils d'hydrophobicité, comme celle fournie par le programme TMHMM v2<sup>10</sup> [4], seules 10 hélices transmembranaires sont détectées (Figure-Ax-2).

Dans le cas de SUR1, le résultat de la prédiction HMMTOP est présenté Figure-C-3 :



**Figure-C-3 : Topologie structurale de SUR1 déterminée par HMMTOP et visualisée par TOPO2.** On distingue 17 hélices transmembranaires, dont 5 faisant partie du domaine TMD0 (37-185) ainsi que deux domaines NBD. Les 2 motifs ABC\_TM1F localisés entre les acides aminés 300-602 et 1014-1306 sont représentés en orange et les 2 motifs ABC\_transporter\_2 localisés entre les acides aminés 679-929 et 1344-1578 sont représentés en violet. Les deux sites potentiels des N-glycosylations, 10-1049, sont présentés en bleu.

On constate qu'une topologie membranaire similaire à celle de la P-gp est retrouvée, avec en plus la présence du domaine caractéristique supplémentaire TMD0 en amont du premier domaine transmembranaire « classique ». TMD0 est composé de 5 hélices transmembranaires, ce qui amène à 17 le nombre total des hélices transmembranaires contenues dans SUR1. Ces hélices sont respectivement localisées entre les acides aminés : 37-

<sup>10</sup> <http://www.cbs.dtu.dk/services/TMHMM/>

54, 75-94, 107-124, 135-152, 161-185 pour le domaine TMD0 et entre les acides aminés 300-324, 351-368, 399-423, 450-473, 536-560, 575-594, 1010-1034, 1061-1085, 1104-1121, 1148-1172, 1253-1270 et 1279-1298 pour le deux autres domaines TMD restants. Deux sites potentiels de glycosylation sont suggérés pour SUR1 mais sont localisés dans deux régions extracellulaires différentes de la protéine au niveau des résidus Asn 10 et 1049. Comme réalisé avec la P-gp, les résultats issus de HMMTOP ont été comparés avec ceux issus de TMHMM. Ce dernier ne détecte que 16 hélices transmembranaires (Figure-Ax-2).

## II.2.2. Quelle topologie pour la région linker ?

L'une des particularités communes entre les structures des transporteurs ABC (Sav1866, MsbA et P-gp de souris) disponibles dans la PDB est l'absence de la région linker (une soixantaine d'acides aminés non résolue dans le cristal). Devant l'absence de cette information structurale chez les templates<sup>11</sup> et avant de procéder à la construction de modèles tridimensionnels de transporteurs ABC, j'ai souhaité évaluer les possibilités de modélisation de cette région dans la P-gp humaine en l'étudiant grâce à un programme de prédiction de structures secondaires. J'ai ainsi soumis la séquence de la P-gp humaine à un programme de prédiction de structures secondaires, PSIPRED (v3.0)<sup>12</sup> [5]. Ce programme est reconnu pour la qualité et l'exactitude de ses résultats ; il se base sur l'optimisation de résultats générés par PSI-BLAST [6]. Le fichier de sortie est présenté dans le chapitre Annexes (Figure-Ax-3).

L'avantage de cet outil est la génération d'un indice de confiance à chaque position prédite des structures secondaires. Cet indice varie de 0 à 9 selon la fiabilité de la prédiction. Ainsi, dans le fichier de sortie, on peut remarquer qu'une région entre les acides aminés 630 et 694 se distingue par un indice de confiance inférieur à 4. Cette région se distingue par une prédiction de structure majoritairement désordonnée (Coil) avec une majorité de degrés de confiance < 4. On voit en effet que d'autres structures prédites « C » dans le reste de la séquence le sont avec un bon degré de confiance en moyenne, correspondant aux boucles de « jonction » extra-membranaires.

Dans le cas du segment « linker », il n'y a manifestement pas de structures ordonnées (les rares régions « H » ont des degrés de confiance de 0 à 2), et un aussi long segment

---

<sup>11</sup> Un template : Une structure protéique utilisée comme « guide » pour reconstruire un modèle 3D par homologie

<sup>12</sup> <http://bioinf.cs.ucl.ac.uk/psipred/>

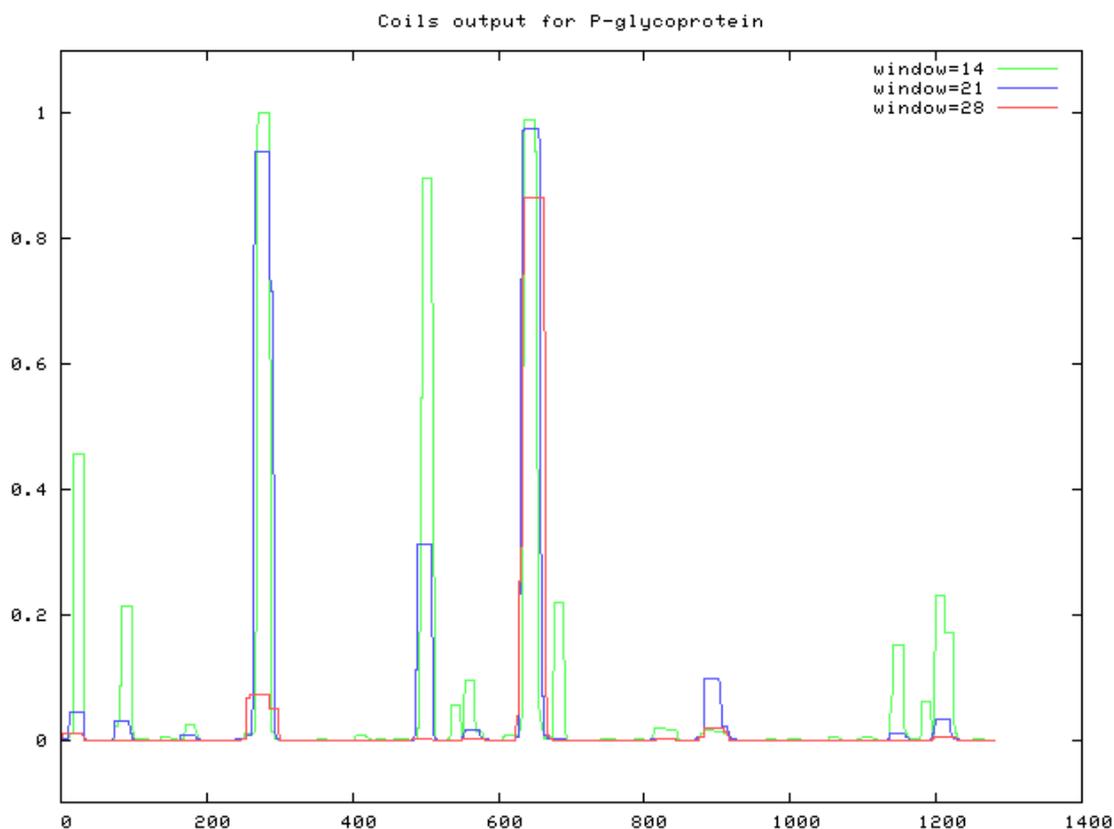
désordonné (Coil) est impossible à modéliser sans template. Ces résidus correspondent à la région linker et ces valeurs signifient qu'une reconstruction de cette région de la P-gp ne serait certainement pas fiable sans l'utilisation d'un template.

### II.2.3. Analyse des enroulements superhélicoïdaux

Nous avons évoqué dans le chapitre Introduction, le modèle de translocation suggéré par Dawson et Locher [7] (Figure-A-16) grâce à la structure tridimensionnelle de Sav1866 dans sa conformation « *Outward-facing* ». Ce modèle avait décrit pour la première fois un enlacement (enroulement) des hélices transmembranaires. Ce même enroulement a, par la suite, été observé chez MsbA par Ward *et al.* [8], dans la même conformation. Ce modèle laisse imaginer la présence d'une topologie dite superhélicoïdale qui implique l'enroulement de deux ou de plusieurs hélices. Ce type de conformation n'est pas une structure secondaire mais un type particulier de structure tertiaire. On prédit ce type de superstructure lorsqu'on trouve des motifs répétés de 7 acides aminés de type *hxxhcx*, *h* pour résidu hydrophobe et *c* pour résidu chargé. La présence de ce type de conformation chez les structures bactériennes suggère l'intérêt de son étude chez la P-gp et SUR1. Pour ce fait, j'ai utilisé un autre programme de prédiction spécifique COILS (v1.0)<sup>13</sup> [9]. Ce programme sert à détecter les probables domaines composés de superhélices ou «coiled-coil». COILS est un programme qui compare une séquence particulière à une base de données de conformations «coiled-coil» connues, dans le but de générer un score de similarité. En comparant ce score à une distribution des scores dans les protéines globulaires et les protéines «coiled-coil», COILS calcule la probabilité que la séquence adopte cette conformation spécifique. Les résultats des analyses COILS réalisées sur la P-gp et SUR1 sont présentés respectivement Figure-C4 et Figure-C-5.

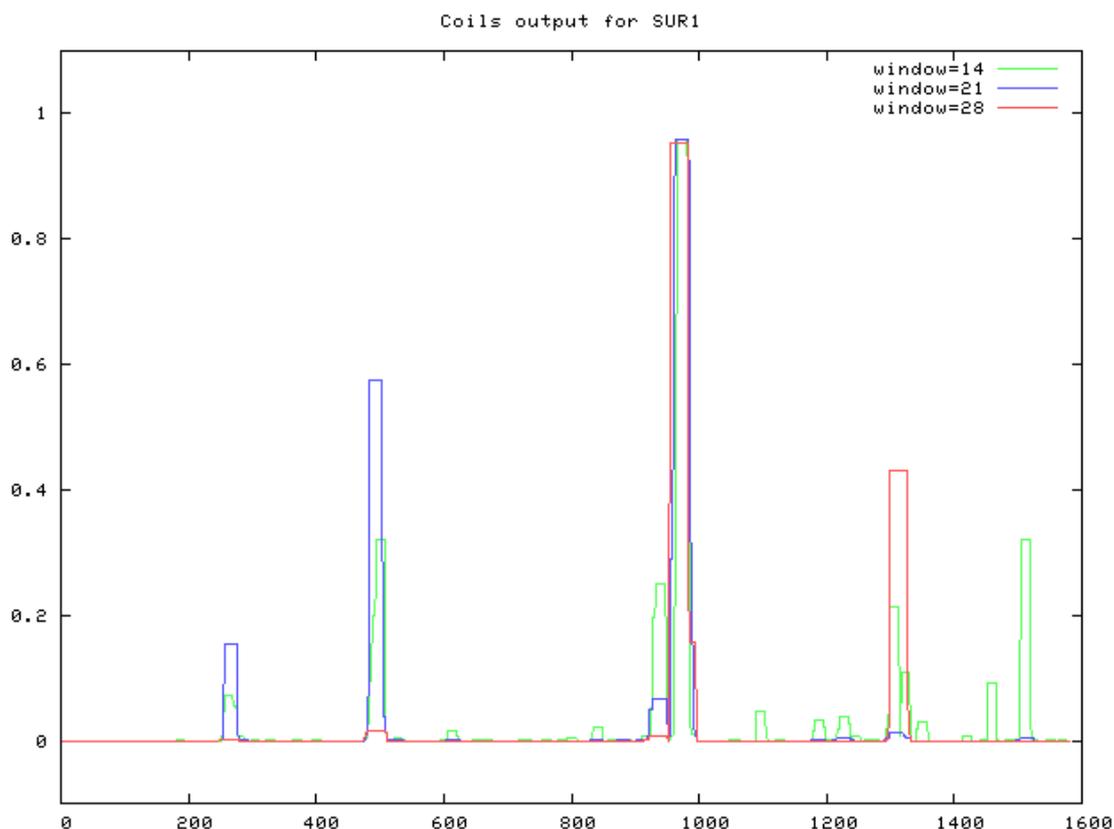
---

<sup>13</sup> [http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)



**Figure-C-4 : Localisation des régions « coiled-coil » chez la P-gp.** *Le « coiled-coil » est une région amphipathique représentant une heptade de 7 résidus hydrophobes et hydrophiles. Les fenêtres de glissement pour la détection des motifs sont ainsi des multiples de 7, c'est-à-dire 14 (vert), 21 (bleu) et 28 (rouge).*

COILS (v1.0) prédit des enroulements superhélicoïdaux dans diverses régions de la P-gp mais deux régions se distinguent avec des valeurs de probabilités assez élevées (> à 0.8 en utilisant deux types de fenêtres de glissement). Ces régions sont situées entre les acides aminés : 269-287 et 636-652. Ces deux régions sont localisées du côté N-terminal et plus précisément au niveau intracellulaire. On peut constater, que la seconde région prédite, fait partie de la région-linker. Elle est prédite avec une grande probabilité quelle que soit la taille de la fenêtre de glissement. Sachant que les domaines « coiled-coil » sont considérés, à cause de leur interaction spécifique, comme des « tags »/signaux/marqueurs de dimérisation, on peut dire que les résultats de cette analyse suggèrent que des motifs potentiels de dimérisation sont présents au niveau de la fin de la région N-terminale de la P-gp.



**Figure-C-5 : Localisation des régions « coiled-coil » chez SUR1.** Le code couleur appliqué aux fenêtres de glissement est le même que celui de la Figure-C-4

Dans le cas de SUR1, plusieurs enroulements superhélicoïdaux sont localisables tout au long de la séquence protéique mais une seule région se distingue réellement, celle située entre les acides aminés 965 et 984. Pour cette région, la probabilité de présence d'une conformation « coiled-coil » est supérieure à 90% quelque soit la fenêtre de glissement. Il est intéressant de remarquer que les résidus de 965 à 984 font partie de la région-linker de SUR1.

Pour la P-gp comme pour SUR1, on peut supposer la présence d'un enroulement superhélicoïdal et donc celle de motifs de dimérisation potentiels au niveau de la région-linker. Cette hypothèse est en harmonie avec celle soumise dans le chapitre précédent, suggérant un phénomène de duplication interne pour les transporteurs entiers. Nous remarquons aussi l'absence des « coiled-coil » au niveau des régions transmembranaires de P-gp et de SUR1 alors que les mouvements de twist observé au cours du cycle catalytique laisseraient supposer l'existence de ces motifs. Par contre, la région linker pourrait jouer un rôle dans le changement conformationnel donnant lieu à l'enlacement des hélices transmembranaires dans la conformation *Outward-Facing*.

### II.3. Modélisation comparative de transporteurs ABC humains

Nous allons nous intéresser, dans ce paragraphe, aux travaux réalisés dans le cadre de l'étude structurale de transporteurs ABC, et plus spécialement la P-gp. Les résultats de ces travaux sont complémentaires à ceux présentés dans l'article inclus dans ce manuscrit.

#### II.3.1. Sélection des templates et alignements de séquences

Alors que les résultats des analyses phylogénétiques indiquent une proximité d'évolution entre la P-gp et les exportateurs ABC bactériens, nous avons procédé à une méthode de confirmation plus classique du choix de nos templates en réalisant un PSI-Blast de la séquence requête (la P-gp humaine ou SUR1) contre la banque de données PDB. PSI-Blast permet d'identifier les séquences protéiques les plus identiques à la séquence de la P-gp et pour lesquelles des structures tridimensionnelles sont déjà disponibles dans la PDB seront révélées.

Hormis la structure de la P-gp de souris parue en 2009 [2] et dont la séquence protéique est à 87% identique à celle de la P-gp humaine, seules les séquences des transporteurs ABC bactériens, MsbA (des trois espèces : *Escherichia coli*, *Vibrio cholerae* et *Salmonella typhimurium*) et Sav1866 (*Staphylococcus aureus*) sont détectées par PSI-Blast avec respectivement des scores d'identité vis-à-vis la séquence de la P-gp humaine de : 31, 28, 31 et 28%. Ces scores, étant proches de 30%, il est communément admis qu'ils sont suffisants pour générer des modèles structuraux fiables par le biais de la modélisation comparative.

Pour ABCC8, les scores d'identité vis-à-vis des séquences bactériennes sont plus faibles mais avoisinent néanmoins les 20%. Le fonctionnement de SUR1 dans le cadre d'un assemblage hétérotétramérique avec des sous-unités Kir, semble difficilement compatible avec le changement conformationnel décrit par la P-gp. Le mouvement d'alternance entre une ouverture vers le milieu intracellulaire puis vers le milieu extracellulaire (Figure-A-5) est imaginé pour des protéines ABC assurant une fonction de transport, fonction ce qui n'a toujours pas été mise en évidence jusqu'à présent pour SUR1.

Nous avons analysé en détail ces scores d'identité en observant leur répartition le long des séquences, et plus précisément au niveau des domaines NBD et TMD (Tableau-Ax-2). La séquence protéique de la P-gp de souris a été intégrée dans cette analyse (Tableau-C-1).

TMD		P-gp_Hum		SUR1		Sav1866	Ec-MsbA	Vc-MsbA	St-MsbA	P-gp_Mouse	
		TMD1	TMD2	TMD1	TMD2					TMD1	TMD2
P-gp_Hum	TMD1	100									
	TMD2	28	100								
SUR1	TMD1	12	9	100							
	TMD2	8	10	10	100						
Sav1866		<b>13</b>	<b>16</b>	<b>11</b>	<b>12</b>	100					
Ec-MsbA		<b>17</b>	<b>16</b>	5	12	16	100				
Vc-MsbA		<b>16</b>	<b>14</b>	14	11	19	65	100			
St-MsbA		<b>18</b>	<b>16</b>	<b>5</b>	<b>11</b>	16	97	65	100		
Pgp_Mouse	TMD1	<b>76</b>	26	<b>11</b>	8	13	15	16	15	100	
	TMD2	27	<b>77</b>	11	<b>11</b>	11	14	16	15	24	100

NBD		P-gp_Hum		SUR1		Sav1866	Ec- MsbA	Vc- MsbA	St- MsbA	P-gp_Mouse	
		NBD1	NBD2	NBD1	NBD2					NBD1	NBD2
P-gp_Hum	NBD1	100									
	NBD2	61	100								
SUR1	NBD1	26	24	100							
	NBD2	29	30	23	100						
Sav1866		50	47	27	31	100					
Ec-MsbA		53	48	31	31	53	100				
Vc-MsbA		45	45	30	31	51	72	100			
St-MsbA		54	48	30	32	55	94	73	100		
Pgp_Mouse	NBD1	92	62	26	29	50	51	46	52	100	
	NBD2	60	91	25	30	47	47	45	48	61	100

**Tableau-C-1 : Les scores d'alignement des domaines TMD et NBD.** Les scores ont été obtenus grâce au programme d'alignement multiple ClustalW [10]. Sont présentés en rouge, les scores d'alignement des domaines représentatifs des templates et ceux représentatifs des modèles à construire. P-gp\_Hum : la P-gp humaine, Pgp\_Mouse : la P-gp de souris, Sa : Staphylococcus aureus, Ec : Escherichia coli, Vc : Vibrio cholerae, St : Salmonella typhimurium

A la lecture du tableau-C-1, nous pouvons constater que les scores d'alignement sont plutôt indépendants des espèces. Ainsi, les domaines TMD et NBD de la P-gp et de SUR1 humaines partagent plus d'identité avec les séquences des exportateurs ABC bactériens qu'entre eux. Ces scores représentent un autre argument supplémentaire en faveur de notre choix de templates. Ils sont aussi en accord avec les résultats issus des analyses phylogénétiques ainsi qu'avec les scores d'alignements globaux des séquences entières (Tableau-Ax-3).

Nous constatons que les scores d'identité sont nettement supérieurs au niveau des domaines NBD en comparaison à ceux des domaines TMD. Ce résultat n'est pas surprenant,

les NBD contiennent les motifs conservés entre toutes les protéines ABC. Ils sont impliqués dans la liaison d'un même substrat qui est l'ATP alors que les domaines TMD sont impliqués dans le transport de substrats variables selon la protéine ABC.

Il faut signaler que trois structures de P-gp de souris ont été déposées dans la PDB alors que le travail de modélisation avait été entamé. Ces structures étaient les premières disponibles pour un transporteur ABC de mammifère. La P-gp de souris n'a été cristallisée que dans une seule conformation du cycle de translocation, « *Inward-Facing* » ouverte vers le milieu cytoplasmique. Une de ces structures représentait la P-gp dans un état dit *apo* (sans substrat ou nucléotide liés), les deux autres avec deux cyclo-peptides liés [Le cyclic-tris-(R)-valineselenazole (QZ59-RRR) et le cyclic-tris-(S)-valineselenazole (QZ59-SSS)].

### II.3.2. Génération et sélection des modèles

Les structures des protéines ABC bactériennes ainsi que celle de la P-gp de souris ont donc été les templates sélectionnés dans notre travail de modélisation comparative. Ces structures ne représentent pas les mêmes conformations, ce qui nous a permis d'entreprendre un travail de construction de modèles structuraux décrivant la P-gp humaine dans trois conformations catalytiques différentes. Pour SUR1, la faiblesse des scores d'identité associée à la disponibilité de deux structures templates dans une même conformation « *Outward-Facing* » m'ont contraint à construire un modèle tridimensionnel unique décrivant SUR1 dans cette conformation (Tableau-C-2).

Templates (Code PDB)	Résolution (Å)	Modèles (conformation)
Ec-MsbA (3B5W)	5.3	Pgp_Hum (open inward-facing)
Vc-MsbA (3B5X)	5.5	Pgp_Hum (closed inward-facing)
Sav1866 (2HYD)	3.0	Pgp_Hum (outward-facing)
St-MsbA (3B60)	3.7	Pgp_Hum (outward-facing)
Pgp-Mouse (3G5U)	3.8	Pgp_Hum (open inward-facing)
Sav1866 (2HYD)	3.0	SUR1 (outward-facing)
St-MsbA (3B60)	3.7	SUR1 (outward-facing)

**Tableau-C-2 : Les modèles P-gp et SUR1 construits et leurs templates correspondants.**

*P-gp\_Hum* : la P-gp humaine, *Pgp\_Mouse* : la P-gp de souris, *Sa* : *Staphylococcus aureus*, *Ec* : *Escherichia coli*, *Vc* : *Vibrio cholerae*, *St* : *Salmonella typhimurium*.

Les données du tableau-C-2 indiquent que les structures dans la conformation *Outward-Facing*, à priori compétentes pour l'efflux du substrat transporté, sont les plus nombreuses et les mieux résolues.

Par contre, il n'existe qu'une seule structure template pour la forme intermédiaire dite *Closed Inward-Facing* et qui a été résolue à basse résolution, ce qui limite la qualité des modèles qui peuvent être dérivés. Pour la conformation *OpenInward-Facing*, deux structures sont disponibles, dont une à résolution correcte (3.8 Å), ce qui permet de construire des modèles corrects, grâce à l'alignement multiple avec deux templates.

Il faut noter qu'étant donné la méthode d'obtention des structures Ec-MsbA et Vc-Msba (avec St-MsbA comme point de départ et en positionnant manuellement les monomères dans une carte expérimentale de densité électronique) [8], elles ne sont représentées dans la PDB que par le biais de traces de C $\alpha$ . La chaîne principale et les chaînes latérales ont été reconstruites respectivement grâce à SABBAC<sup>14</sup> [11] et Scit<sup>15</sup> [12], deux programmes fournis par le serveur RPBS (Ressource Parisienne en Bioinformatique Structurale).

Pour chaque procédure de modélisation, nous avons utilisé le programme Modeller dans sa version 9v6 [13]. Ce programme de modélisation comparative est le plus utilisé dans la communauté scientifique et il est implémenté par de nombreux serveurs en ligne. Gratuit et à usage académique, Modeller est un programme automatisé : l'utilisateur doit essentiellement fournir un alignement de séquence (réalisé par ses propres soins). Cette étape reste la plus délicate de l'ensemble de la procédure étant donné la faible conservation de séquences observée au niveau des domaines transmembranaires. L'alignement est amélioré manuellement de manière à respecter la topologie membranaire prédite. Un exemple d'un alignement type soumis à Modeller est présenté Figure Ax-4.

Grâce à la procédure « automodel » de ce programme, nous avons généré 10 modèles de chaque conformation étudiée. Les modèles générés ont été affinés grâce à la procédure « refinement slow » de Modeller. Il noter qu'étant donné les absences, d'un domaine TMD0, de la région-linker ainsi que de la première boucle extracellulaire, au niveau des structures templates (bactériennes ou de souris), ces régions n'ont pu être modélisées.

Le meilleur modèle généré a, par la suite, été sélectionné grâce à une combinaison de programmes de scoring fournis par Modeller et qui sont : GA341 et DOPE (Discrete Optimized Protein Energy). Tous les modèles générés, pour la P-gp ou pour SUR1, ont été

---

<sup>14</sup> <http://bioserv.rpbs.jussieu.fr/cgi-bin/SABBAC>: Reconstruction de la chaîne principale

<sup>15</sup> <http://bioserv.rpbs.jussieu.fr/cgi-bin/SCit>: Analyse et édition des chaînes latérales

validés grâce à des scores GA341 > 0,9. Un exemple d'un script type soumis à Modeller décrivant ces étapes de génération et de sélection de modèles est présenté Figure-Ax-5.

### II.3.3. Evaluation des modèles

Afin d'interpréter et d'exploiter les modèles 3D de protéines, il est essentiel d'estimer leur précision aussi bien globale que locale. Les erreurs dans les modèles proviennent principalement de deux sources: (i) l'échec de la recherche conformationnelle pour trouver la conformation optimale ainsi que (ii) l'échec de la fonction *scoring* à identifier la conformation optimale. C'est ainsi que les qualités stéréochimiques des modèles générés ont été vérifiées par le biais de divers programmes d'analyse et de validation des structures protéiques. Le but de cette stratégie est d'avoir un consensus permettant de juger la fiabilité d'un modèle donné.

**ProSA** [14]: Ce programme calcule un score décrivant la qualité générale d'un modèle donné puis le compare avec celui décrivant des protéines natives de la PDB (Rayons-X ou RMN). Le modèle est de bonne qualité s'il est classé au même rang que des protéines cristallisées de taille similaire. ProSA donne des Z-scores. Plus le Z-score est bas, plus le modèle est bon. Le Z-score dépend de la taille de la séquence protéique. Exemple : pour une protéine de plus de 1000 acides aminés, un bon Z-score est compris entre -9 et -15.

**Procheck** [15]: Procheck évalue la «normalité» de la géométrie des résidus d'un modèle protéique donné par rapport à des structures d'une assez haute résolution. Parmi les paramètres décrits par ce programme figure le diagramme de Ramachandran, qui décrit les angles de torsion *phi-psi* de tous les résidus d'une structure protéique. Il est reconnu qu'un bon modèle protéique est celui dont les combinaisons des angles de torsion *phi-psi* de ses résidus sont contenues dans les régions dites favorables du diagramme. D'autres paramètres de Procheck sont les paramètres « Goodness-factors » ou « G-factors » qui décrivent la qualité des angles dièdres, des liaisons covalentes et de la géométrie générale de la protéine. Pour un modèle fiable, les valeurs de ces paramètres sont aux alentours de -0.5

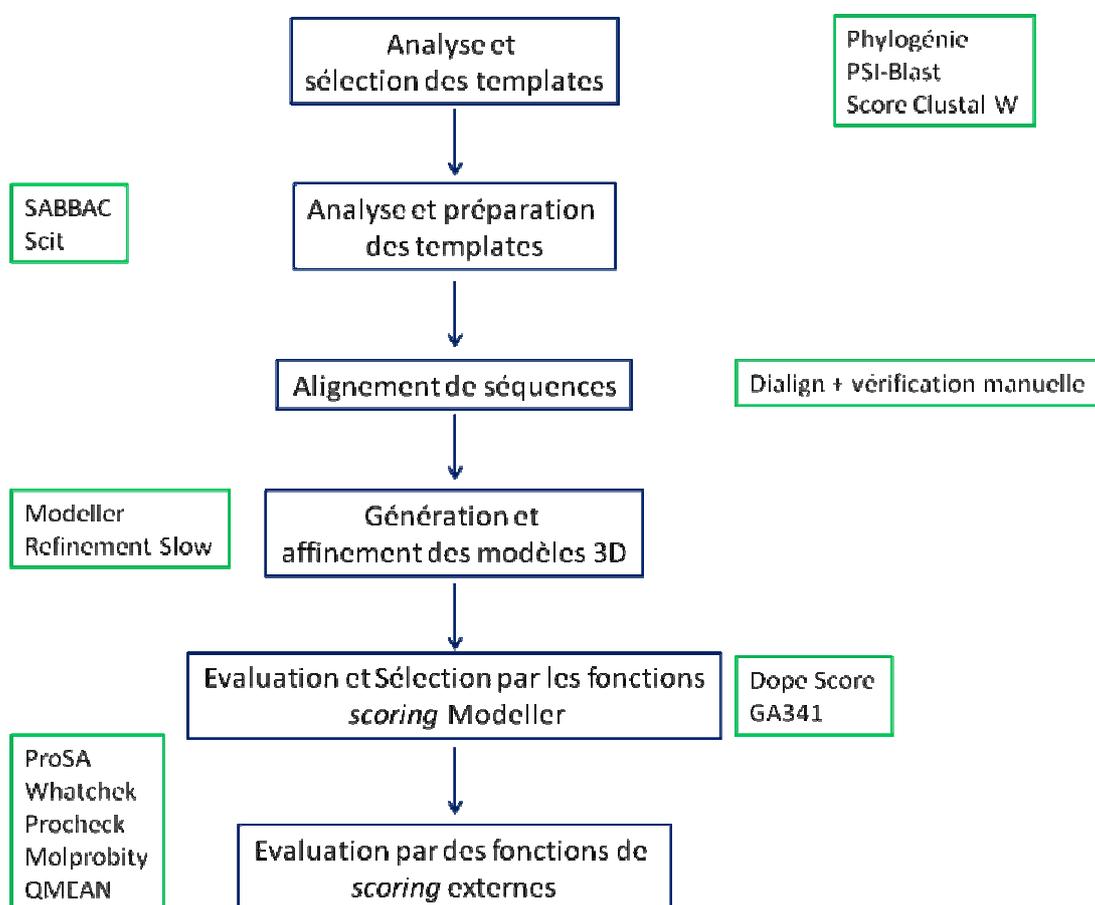
**Whatcheck** [16]: Ce programme évalue aussi les qualités stéréochimiques de structures modélisées et ceci grâce à des paramètres de scoring ; certains décrivant la qualité de la conformation générale du modèle (Z-score structural, bonne qualité de structure quand les valeurs sont positives), d'autres traduisant le degré de conformité de ce modèle vis-à-vis

des contraintes communes d'affinement (RMS Z-scores, bonne qualité de structure quand les valeurs sont proches de 1).

**Molprobability**<sup>16</sup> [17]: Il s'agit d'un serveur fournissant d'autres outils d'analyses des contacts entre les résidus permettant d'évaluer le pourcentage de ceux ayant des mauvais angles ou des mauvaises liaisons.

**QMEAN**<sup>17</sup> [18]: C'est un score qui estime de manière statistique la qualité des modèles protéiques. La valeur du score est comprise entre 0 et 1, et le modèle est fiable quand cette valeur est proche de 1.

Un diagramme graphique récapitulatif est présenté Figure-C-6 :



**Figure-C-6 : Diagramme récapitulatif du processus de modélisation réalisé**

Les résultats de ces divers tests d'évaluation sont résumés dans le Tableau-C-3

<sup>16</sup> <http://molprobability.biochem.duke.edu/>

<sup>17</sup> <http://swissmodel.expasy.org/qmean>

**Tableau-C-3: Evaluation des modèles structuraux.** *P-gp Open Inward-Facing A: le modèle P-gp reconstruit en se basant sur la structure d'Ec-MsbA structure comme template, P-gp Open Inward-Facing B: le modèle P-gp reconstruit en se basant sur la structure de la P-gp de souris comme template. Pour les autres structures modélisées, les templates sont résumés dans le tableau-C-2*

Méthodes	Modèles structuraux		P-gp Open Inward-Facing A	P-gp Open Inward-Facing B	Pgp Closed Inward-Facing	Pgp Outward-Facing	SURI Outward-Facing
ProSA	Z-Score		-9,62	-12,09	-4,21	-11,41	-9,59
Procheck	Ramachandran (%)		97,8	99,4	94,7	99,5	98,7
	G-Factors	Dihedral	0	-0,15	0	-0,13	0
		Covalent	-0,62	-0,25	-0,27	-0,73	-0,27
		Overall	-0,03	-0,18	-0,1	-0,04	-0,1
Whatcheck	Structure Z-scores	1st generation packing quality	0,644	0,656	1,109	-0,293	1,23
		2nd generation packing quality	-2,645	-2,148	-2,187	-3,451	-2,341
		Ramachandran Plot appearance	-2,823	-3,827	-2,115	-3,39	-2,182
		chi-1/chi-2 rotamer normality	-1,164	-2,866	-1,999	-1,567	-2,846
		Backbone conformation	0,984	1,18	2,682	-0,719	1,948
	RMS Z-scores	Bond lengths	0,969	0,955	0,962	0,972	0,978
		Bond angles	1,339	1,269	1,291	1,421	1,279
		Omega angle restraints	0,13	0,138	0,126	0,159	0,126
		Side chain planarity	0,575	0,35	0,465	0,407	0,326
		Improper dihedral distribution	1,121	1,224	1,109	1,272	1,147
		Inside/outside distribution	1,145	1,149	1,171	1,209	1,222
MOLPROBITY	Residues with bad bonds (%)		0,00	0,00	0,00	0,00	0,00
	Residues with bad angles (%)		1,52	0,68	3,25	0,68	1,07
QMEAN	QMEANscore		0,512	0,607	0,256	0,742	0,525

A partir de ce tableau, il est possible de constater la bonne qualité des modèles générés par Modeller, même ceux basés sur les structures d'Ec-MsbA et Vc-MsbA (de basse résolution et contenant uniquement les traces de Ca). Par exemple et en se basant sur le diagramme de Ramachandran, plus de 90% des résidus, quel que soit le modèle, sont localisés dans des régions soient favorables, soient permises. Les résultats concernant les modèles *Open Inward-Facing* confirment que seul le template P-gp de souris peut être pris en considération pour la reconstruction de modèle dans cette conformation malgré les bons scores obtenus avec le template bactérien : (i) les deux palettes de score sont plutôt bonnes alors que l'un des deux templates possède beaucoup plus de similarité de séquence que l'autre (87% versus 30%), (ii) les scores ne sont pas du tout mauvais pour le template bactérien à 30%, ce qui confirme la parenté structurale observée grâce à la phylogénie. (iii) seul le score ProSA montre qu'il y a meilleure adéquation entre le *fold* de la souris et la séquence humaine, qu'entre le *fold* bactérien et la séquence humaine, ce qui paraît logique.

Pour la conformation *Closed Inward-Facing*, l'évaluation est satisfaisante mais elle est la moins bonne de tous les modèles. Ce résultat est dû à l'utilisation d'un seul template, ne présentant pas une très bonne résolution et dont la qualité est dépendante des reconstructions complémentaires réalisées (la chaîne principale et les chaînes latérales).

Pour la conformation *Outward-Facing* de la P-gp, l'évaluation produit de bons scores. Ce modèle est celui dont la qualité est la plus proche de celle du modèle *Open Inward-Facing* basé sur la structure de la P-gp de souris et qui bénéficie de l'identité de séquence la plus élevée (87%). Ces scores très favorables montrent l'intérêt d'opter pour une construction de modèle basée sur deux templates (voir Tableau-C-2).

Pour la conformation *Outward-Facing* de SUR1, nous pouvons constater que dans ce cas aussi et ceci malgré une identité de séquences assez faible vis-à-vis des structures bactériennes, le modèle construit possède de bonnes qualités stéréochimiques.

Dans le paragraphe suivant, est présenté l'article Bessadok *et al.* paru dans JBC. Pour assurer la meilleure clarté des figures représentatives des résultats de modélisation comparative, j'inclus cet article dans son format de soumission.

Recognition of sulfonylurea receptor (ABCC8/9) ligands by the multidrug resistance transporter P-glycoprotein (ABCB1): Functional similarities based on common structural features between two multispecific ABC proteins

# Recognition of sulfonylurea receptor (ABCC8/9) ligands by the multidrug resistance transporter P-glycoprotein (ABCB1): Functional similarities based on common structural features between two multispecific ABC proteins

*Anis Bessadok<sup>§#</sup>, Elisabeth Garcia<sup>##</sup>, H el ene Jacquet<sup>‡#</sup>, Solenne Martin<sup>§</sup>, Alexia Garrigues<sup>§</sup>, Nicolas Loiseau<sup>||</sup>, Fran ois Andr e<sup>§</sup>, St ephane Orłowski<sup>§\*</sup> and Michel Vivaudou<sup>‡\*</sup>*

\* + \* + \*

<sup>‡</sup> Institut de Biologie Structurale (CEA, CNRS, UJF), 41 Rue Jules Horowitz, 38027 Grenoble, France; <sup>§</sup> Service de Bio nerg tique, Biologie Structurale et M canismes, URA 2096 CNRS, iBiTec-S, CEA-Saclay, 91191 Gif-sur-Yvette cedex, France; <sup>||</sup>Laboratoire de Pharmacologie – Toxicologie, INRA-UR66, 180 chemin de Tournefeuille, BP 93173, 31027 Toulouse, France.

# these authors had equal contribution

\*Address correspondence to: Dr. Michel Vivaudou, IBS, 41 rue Jules Horowitz, 38027, Grenoble, France. Tel.: +33-4 38 78 48 67; Fax: +33-4 38 78 54 87. E-mail: michel.vivaudou@ibs.fr; and Dr St ephane Orłowski, iBiTec-S, URA CNRS 2096, CEA Saclay, 91191 Gif-sur-Yvette cedex, France. Fax: +33-1 69 08 87 17. E-mail: stephane.orłowski@cea.fr.

Present address of AG: D partement Recherche S curit , Direction des Sciences du Vivant, L'Or al – Recherche Avanc e, 1 avenue Eug ne Schueller, BP22, 93601 Aulnay-sous-Bois cedex, France.

Present address of SM: Biopr dic, Parc d'affaires de la Bret che, Batiment B1, 35760 Saint-Gr goire, France.

Present address of HJ: CEA/Cadarache, IBEB - LEMS, 13108 St Paul Lez Durance, France.

Present address of EG: CEA/Grenoble, 38054 Grenoble, France.

RUNNING TITLE: Homologous drug-binding sites in SUR and P-gp

## ABSTRACT

ATP-sensitive  $K^+$  ( $K_{ATP}$ ) channels are the target of a number of pharmacological agents, blockers like hypoglycemic sulfonylureas, and openers like the hypotensive cromakalim and diazoxide. These agents act on the channel regulatory subunit, the sulfonylurea receptor (SUR), which is an ABC protein with homologies to P-glycoprotein (P-gp). P-gp is a multidrug transporter expressed in tumor cells and in some healthy tissues. Since these two ABC proteins both exhibit multispecific recognition properties, we have tested whether SUR ligands could be substrates of P-gp. Interaction with P-gp was assayed by monitoring ATPase activity of P-gp-enriched vesicles. The blockers glibenclamide, tolbutamide, and meglitinide increased ATPase activity, with a rank order of potencies which correlated with their capacity to block  $K_{ATP}$  channels. P-gp ATPase activity was also increased by the openers SR47063 (a cromakalim analogue), P1075 (a pinacidil analogue) and diazoxide. Thus, these molecules bind to P-gp (although with lower affinities than for SUR), and are possibly transported by P-gp. Competition experiments among these molecules as well as with typical P-gp substrates revealed a structural similarity between drug binding domains in the two proteins. To rationalize the observed data, we addressed the molecular features of these proteins, and compared structural models, computerized by homology from the recently solved structures of murine P-gp and bacterial ABC transporters MsbA and Sav1866. Considering the various residues experimentally assigned to be involved in drug binding, we uncovered several hotspots which organized spatially in two main binding domains, selective for SR47063 and for glibenclamide, in matching regions of both P-gp and SUR.

## INTRODUCTION

ABC proteins form a large superfamily of mostly membrane proteins. They present as common characteristics a nucleotide binding domain able to hydrolyze ATP which includes the sequence motifs Walker A and B and the signature C (1). As membrane active transporters, they are expressed in virtually all branches of the living reign, and they exhibit a very broad diversity of transported

substrates, handled in either direction of influx or efflux (2). In higher mammals, they are involved in various pathophysiological situations and genetic diseases (3).

The  $K_{ATP}$  channel results from the constitutive association of 4 pore-forming Kir6.x subunits and 4 regulatory SUR subunits [(4);(5)]. In various combinations of the SUR isoforms, SUR1, SUR2A, and SUR2B, and the Kir6 isoforms, Kir6.1 and Kir6.2, these channels are present in most excitable cells – including neuronal, cardiac muscle, smooth muscle, and endocrine cells – (6) where they serve to couple membrane electrical properties to intracellular metabolism.

SUR, a member of the ABCC/MRP subfamily of ABC proteins, is the site of action of numerous drugs that cause either closing ( $K_{ATP}$  channel blockers) or opening ( $K_{ATP}$  channel openers) of the Kir6.x potassium pore (7). Blockers include antidiabetic drugs, sulfonylureas like glibenclamide, as well as non-sulfonylureas like meglitinide. Openers present interesting therapeutic opportunities, in particular for their antihypertensive and cardioprotective properties: a few, like diazoxide, pinacidil, nicorandil, and minoxidil are in clinical use as treatment against hypertension, angina pectoris, and alopecia, respectively (8).

Among ABC proteins of higher organisms, SUR stands out because its only known function is that of a channel regulator while most others are active (or passive in the case of CFTR) transporters. Yet, SUR possesses strong sequence homologies with other eukaryotic ABC proteins, like the multidrug-resistance transporters MRP1 and P-gp (P-glycoprotein), suggesting that these proteins could share some structural and pharmacological properties (9). Indeed, SUR is the target of a panel of structurally unrelated ligands, like P-gp and other multidrug transporters. Multispecific drug recognition by P-gp is still a misunderstood molecular property, in spite of a mass of published data including extensive mutational analyses (10), as well as a pharmacophore model of some of its known ligands (11). As binding of SUR ligands has also been linked to some particular residues [(12);(13)], it is of interest to compare the ligand recognition features of these two ABC proteins, with the aim of gaining new insights in both of them.

In addition, regulation of several ABC transporters by  $K_{ATP}$  channel modulators has been reported: CFTR is inhibited by glibenclamide, and

less potently by a few other  $K_{ATP}$  channel blockers and openers (14). Glibenclamide also inhibits MRP1 (15) and P-gp (16), and it appears to be a substrate of the latter. These observations are of interest since they can complete the pharmacological toolbox of researchers – glibenclamide is now the most widely-used blocker of CFTR channels – and they can point to potential side-effects and altered biodisposition. This is specially important for P-gp, which is an essential gate-keeper at the blood-brain barrier as well as in cancer cells [(17);(18)].

In this work, we have therefore focused on P-gp and examined whether representative compounds among  $K_{ATP}$  channel blockers and openers could interact with P-gp. As blockers, we used glibenclamide, as a reference, and its two moieties, tolbutamide and meglitinide; as openers, we used diazoxide, SR47063, a derivative of cromakalim, and P1075, a derivative of pinacidil. Our findings demonstrate that these molecules bind to and are possibly transported by P-gp. They imply a possible structural similarity between SUR and P-gp, which could extend to a mechanistic similarity, in line with our recent work [(19);(12)], suggesting that  $K_{ATP}$  channel openers and blockers interact with SUR much like transport substrates would, although intrinsic active transport by SUR has not yet been demonstrated.

In an effort to rationalize the observed data, and in the absence of high resolution crystallographic data, we built new homology models of human P-gp and SUR, based on multiple alignments with the bacterial ABC transporters MsbA and Sav1866, whose structures have been solved by X-ray crystallography in different conformations [(20);(21)], and on mouse P-gp whose structure has been recently solved in one conformation (22). Highlighting the various residues known to be implicated in drug binding specificity helped to delineate different hotspots for drug binding in the P-gp and SUR structures. Their 3D arrangement revealed two main binding domains for the drugs verapamil/SR47063 and for vinblastine/glibenclamide in P-gp, and their counterpart for binding of SR47063 and of glibenclamide in SUR.

In this work, we used structural models derived from comparative modeling, because it offers the advantage over crystallographic data to interpret a huge amount of binding data and to address the multispecificity character of these ABC proteins.

Homology modeling also allowed us to interpret the results in various protein conformations, in relation to the enzymatic turn-over.

## EXPERIMENTAL PROCEDURES

***P-glycoprotein-containing membrane vesicles preparation.*** The MDR cell line used, DC-3F/ADX, was selected from spontaneously transformed Chinese hamster lung fibroblast DC-3F on the basis of resistance to actinomycin D. This resistance is due to overexpression of the *pgp1* gene. The DC-3F/ADX cells and DC-3F cells, their drug-sensitive parental counterparts, were cultured, harvested by scraping and washed in PBS in the presence of antiproteases. The cells in suspension were then disrupted by sonication. The resulting centrifuged supernatant was layered on a sucrose cushion to isolate the total membrane vesicle fraction. Membrane protein concentrations were determined by Bradford's method. In vesicles prepared from DC-3F/ADX, P-glycoprotein accounts for about 12-15% of total membrane proteins, whereas P-gp cannot be detected in the control vesicles prepared from the DC-3F cells (23). Experiments were performed using DC-3F cells as a control and DC-3F/ADX cells as a P-gp test. Because SUR displays ATPase activity modulated by its ligands (19), our results could be tainted by the unlikely presence of different amounts of SUR in DC-3F and DC-3F/ADX cells. The effect would be however negligible because the ATPase activity of SUR is >100-fold less than P-gp [(24);(25);(26)].

***Measurement of ATPase activity.*** The experiments are based on the enzymatic determination of ATPase activity of native membrane vesicles containing high amounts of P-glycoprotein, measured at 37°C by using a standard coupled-enzyme assay, comprising an ATP-regenerating system and continuous spectrophotometric detection of NADH absorbance at 340 nm, which allows the monitoring of NADH consumption in the reaction medium, that is stoichiometric to ADP production.

The reaction medium contained 30 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM KCl, 2 mM  $MgCl_2$ , 1 mM dithiothreitol, 1 mM MgATP, supplemented with 0.1 mg/ml pyruvate kinase, 1 mM phosphoenol pyruvate, 0.1 mg/ml lactate dehydrogenase and 0.5 mM NADH. ATPase activity was determined with about 20  $\mu$ g/ml of membrane proteins. The assay

medium was supplemented with 10 mM sodium azide, 0.5 mM ouabain and 1 mM EGTA, to inhibit the H<sup>+</sup>-ATPases, Na<sup>+</sup>/K<sup>+</sup>-ATPase, and Ca<sup>2+</sup>-ATPases, respectively. The residual ATPase activity measured in the presence of the specific inhibitors of these enzymes and in the absence of any added drug was essentially due to the basal activity of P-glycoprotein. MgATP concentration was kept throughout the experiments at the sub-saturating level of 1 mM since we have previously reported that the Km value of P-gp for MgATP remains essentially unchanged and independent of transported substrates [(23);(27);(28)]. Shown error bars represent standard error of the mean, computed from measurements performed in 5 to 14 experiments using up to 4 distinct vesicle preparations. Model fitting to the ATPase activity values normalized to the basal values was done with Origin 5.0 software using a standard Hill equation:

$$f(|\text{Ligand}|) = A_0 + V_{\text{max}} \cdot |\text{Ligand}|^h / [|\text{Ligand}|^h + K_{1/2}^h]$$

where A<sub>0</sub> is the basal activity in absence of ligand, K<sub>1/2</sub> the concentration for half-maximal activation (or inhibition), and h (or -h) the Hill coefficient.

Although activity vs concentration curves often show an inhibition at the highest concentrations, we fitted these curves using only one high affinity specific binding site because : (i) this inhibition may be due to a non-specific membrane perturbation by amphiphilic drugs, rather than to a low-affinity specific binding site; (ii) curve fitting using a two-site binding model would require twice as many parameters as a single-site model, producing more uncertainty and no more qualitative information about the mutual relationships exhibited by the compounds tested. [See for example the parameters determined in Fig.4 compared with those in supplemental Fig.S1A and B, the two last fits depending on the level of the maximal inhibition expected (experimentally unattainable in the usable drug concentration range)]. Overall, the single-site binding model yielded estimates of the apparent affinities of various P-gp substrates consistent with both the K<sub>1/2</sub> values characterizing basal ATPase activity stimulation and the K<sub>i</sub> values characterizing the inhibition of ATPase activity stimulated by reference P-gp substrates [(27);(28);(29);(11)].

**Chemicals.** To test a range of concentrations for a given compound, aliquots were sequentially added to the same optical cuvette under continuous stirring. Dilution effects were taken into account for the final ATPase activity calculations. Chemicals

(from Sigma except where stated) and stock solution concentrations were: diazoxide, 100 mM; glibenclamide, 100 mM; meglitinide (4-[2-(5-chloro-2-methoxybenzamido)ethyl]benzoic acid; Aventis, Frankfurt, Germany), 100 mM; P1075 [N-cyano-N'-(1,1-dimethylpropyl)-N'-3-pyridylguanidine; Leo Pharmaceutical Products, Copenhagen, Denmark], 150 mM; progesterone, 30 mM; SR47063 [4-(2-cyanimino-1,2-dihydro-1-pyridyl)-2,2-dimethyl-6-nitrochromene; Sanofi Recherche, Montpellier, France], 150 mM; tolbutamide, 100 mM; verapamil, 150 mM; vinblastine, 100 mM. All stocks were prepared in DMSO, except for progesterone which was dissolved in ethanol. Compounds were added to the reaction medium directly from stock solutions or from dilutions either in DMSO or ethanol, or in distilled water. The maximal amount of DMSO added was ~2% for the last point of the dose-response measurements of the lowest affinity compounds. The effect of such concentrations of vehicle was assessed by testing increasing amounts of DMSO on control vesicles (with and without the non-P-gp ATPase inhibitors mentioned above) and on P-gp-containing/enriched vesicles (with and without the P-gp substrate verapamil) (data not shown). In all cases, DMSO, at the concentrations reached in our experiments, did not modify ATPase activity in any significant manner. At higher concentrations, the effects remained moderate, since 6% DMSO augmented P-gp ATPase activity by 6% (n=4).

Absorbance spectra for all tested compounds (data not shown) revealed that only SR47063, which has an absorbance peak at 328 nm, absorbed significantly at our assay wavelength of 340 nm. The resulting interference decreased the signal-to-noise ratio at the highest SR47063 concentrations used but would not alter the slope of the signal which is the parameter used to measure the ATPase rate.

**Phylogenetic analyses.** Amino acid sequences of 42 human and 2 bacterial, Sav1866 (accession number Q99T13) and MsbA (accession number A1A9J1), transmembrane ABC proteins were used for a phylogenetic reconstruction. The Protein Knowledgebase (UniProtKB: [www.uniprot.org](http://www.uniprot.org); (30)) and the TCDB (Transport Classification Database) ([www.tcdb.org](http://www.tcdb.org); (31)) were used to retrieve fasta files, Swiss-Prot accession codes,

synonyms and TCDB classification numbers (see Table S1).

Since some ABC proteins are monomeric (full-size transporters) whereas others are dimeric (half-size transporters), and since some of them exhibit an additional N-terminal TMD0 domain, we used the “Conserved Domain Database” (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> (32)) of NCBI to detect the respective consensus regions in the proteins. This roughly corresponds to one monomer of the dimeric proteins (+/- a dozen of residues), and amino acid sequences with a size of about 600-650 residues were selected for the phylogenetic study, thereby excluding the TMD0 sequences (Table S1).

Multiple amino acid sequence alignment was performed with DIALIGN (33).

A maximum likelihood tree was constructed using PhyML (34) and we made 100 bootstrap replicates [(35,36)] to assess confidence on the results. Indeed, this statistic technique is used to improve the robustness of a constructed tree and to inform about its reliability. As a general rule, if the bootstrap value for a given branch is found higher than 90%, then the topology of that branch is considered as “correct”. The phylogenetic tree was edited using TreeViewX

(<http://taxonomy.zoology.gla.ac.uk/software>

(37)): in this representation, each branch length reflects a statistical distance between the considered sequence and its “common ancestor”.

**Homology Modeling.** The alignment of SUR, P-gp, Sav1866 and MsbA, shown in Figure S2, was edited using Jalview Software (38). The secondary structures, derived from the 3D model of P-gp, were added manually above the alignment.

- *Template selection:* PSI-BLAST (39) against Protein Data Bank (PDB) was used to select the templates for the homology modeling of P-gp and SUR1. This analysis confirmed that MsbA and Sav1866 have the closest sequences to these two mammalian ABC proteins. MsbA and Sav1866 share 28-31% identity with P-gp and 20-21% with SUR1 (see legend of Fig.S2). After the withdrawal of the first crystallographic MsbA structures, the new structures deposited were proposed to represent three distinct putative functional conformations of this representative ABC transporter (20), the open inward-facing, the closed inward-facing and the outward-facing conformations. They were thus used as templates to rebuild three structural models for

the human P-gp (see Table 1); the outward-facing conformation also used Sav1866 as a template, allowing a two-template homology modeling. The recent crystallographic structure of mouse P-gp allowed us to rebuild another model for the human P-gp in the open inward-facing conformation based on a single pairwise alignment (87% sequence identity between human and mouse P-gp). For SUR1, only the outward-facing conformation was built using a two-template reconstruction from both St-MsbA and Sav1866 (Table 1), hence giving a higher confidence than a mono-template homology modeling.

- *Model generation:* For each modeling procedure, ten models were generated with the Modeller 9v6 software (40), with refinement settings ‘slow’ and the “automodel” procedure. The templates, which were either the N- and C-terminal halves of mouse P-gp structure or half-size bacterial transporters, do not contain the domain TMD0, the first intracellular loop and the linker region. Hence these regions were excluded from modeling.

- *Model selection:* A scoring program was applied to each generated model and the best scored model among the ten was selected. We have used a combination of the scoring programs available under Modeller, GA341 score and DOPE (Discrete Optimized Protein Energy) score.

- *Structural model assessment:* The stereochemical qualities of the models were checked using programs for analyzing and validating protein structures: *Procheck* (41), *Whatcheck* (42), *ProSA web* (43), *Molprobit* (44) and *QMEAN* (45).

- *Rebuilding structural models in a lipid environment:* The three modeled P-gp conformations were placed in a “native-like” environment in order to analyze the interactions of the protein with model phospholipids arranged in a double layer membrane. We chose POPC (1-palmitoyl-2-oleoyl-phosphatidylcholine) because it is the most prominent lipid in mammalian cell membranes. The procedure included a first step of solvation, realized by the Solvate program <http://www.mpibpc.mpg.de/home/grubmueller/downloads/solvate/index.html>, and then successive steps allowed by the VMD (Visual Molecular Dynamics) software (46), using the “solvation box” and the “membrane builder” procedures.

**Ligand molecular modeling and alignment.** The structures of glibenclamide and SR47063 were

minimized and aligned under SYBYL (Tripos, St. Louis, MO) with the structures of vinblastine and verapamil, respectively, using procedures and data from Garrigues et al (11). Hydrophobicity and H-bond potential areas were calculated using MOLCAD routine of SYBYL. The compounds were superimposed manually and consensus groups selected. The compounds were then aligned using the multifit command of SYBYL.

## RESULTS

**Experimental assay validation.** Throughout this work, the ATPase activity of P-gp-enriched vesicles was used as a reporter of the function of native P-gp. This assay has been validated thoroughly in earlier works [(28);(11)]. To further ascertain its suitability in our experimental settings, tests were performed using the P-gp substrates verapamil, progesterone, and vinblastine (Fig.1A). Our results are fully consistent, qualitatively and quantitatively, with the expected effects of these substrates on P-gp ATPase activity (28). Verapamil and progesterone increased ATPase activity with  $K_{1/2}$  of 1.3 and 17  $\mu$ M, respectively, with no obvious cooperativity (Hill coefficients nearing 1). This activation declined at higher concentrations, a phenomenon that could be consistent with the presence of both high and low-affinity sites as well as with a non-specific perturbation due to the insertion of an amphiphilic compound into the membrane bilayer. Vinblastine inhibited the basal ATPase activity of P-gp-enriched vesicles down to a level close to that of the control vesicles. As control vesicles were unaffected by these 3 compounds, it appears that the ~70% additional ATPase activity measured in P-gp-enriched vesicles arose mainly from the presence of P-gp proteins.

**$K_{ATP}$  channel blockers stimulate P-gp ATPase.** Inhibitors of  $K_{ATP}$  channels include sulfonylureas as well as non-sulfonylureas compounds, all known to act through direct binding to SUR (47). Evidence has been published that the prototypical sulfonylurea glibenclamide is a P-gp substrate able to block, at 100  $\mu$ M, transport of colchicine by P-gp (16). The data from Fig.1B are in agreement with this finding as they show that glibenclamide augments the ATPase activity with a maximum attained near 100  $\mu$ M. The dose-effect curve resembles that obtained in Fig.1A with verapamil.

Two other blockers were examined: meglitinide, which represents the non-sulfonylurea half of glibenclamide, and tolbutamide, the sulfonylurea half of glibenclamide. Both increased ATPase activity, although with lower affinities (Fig.1B). Although the maximum activation could not be reached as this would have required excessive concentrations and excessive levels of the vehicle DMSO, Hill equation fits to the data gave maximum activation levels of ~70% for the 2 compounds, a value similar to that achieved with the other ATPase activators, glibenclamide, verapamil and progesterone.

**$K_{ATP}$  channel openers stimulate P-gp ATPase.**  $K_{ATP}$  channel openers form an extended family of seemingly unrelated molecules that bind to SUR at a site distinct from the sulfonylurea binding site and allosterically cause channel opening (7). We tested three representative openers: diazoxide; SR47063, a derivative of cromakalim; and P1075, a derivative of pinacidil. Diazoxide targets both SUR1 and SUR2 isoforms (48), while SR47063 and P1075 target only SUR2 isoforms through a common site separate from that of diazoxide.

All three compounds caused an increase in ATPase activity of P-gp-containing vesicles but not control vesicles (Fig.1C). P1075 caused the largest increase ( $V_{max} = 101\%$ ) but had the lowest affinity ( $K_{1/2}=1.5$  mM). Diazoxide had also a low affinity of 0.6 mM. The affinity of SR47063 was highest at 36  $\mu$ M but its effect was more modest ( $V_{max} = 55\%$ ). These latter values likely underestimate the actual effects of SR47063 since the higher concentrations of SR47063 had an inhibitory effect on the contaminant sources of ATPase activity that appeared when the compound was tested in the control vesicles devoid of P-gp (Fig.1C). It should be noted that the effect of SR47063 was modest in the experiments using the full concentration range, but could be more sizeable in separate single-concentration experiments as in the experiment of Fig.3 and 4.

**Cumulative effects of glibenclamide and other  $K_{ATP}$  channel modulators.** Binding assays on SUR using radiolabelled glibenclamide have shown that most  $K_{ATP}$  channel modulators displace glibenclamide binding (9). We therefore performed experiments designed to reveal interaction between glibenclamide and our other test compounds. Full concentration-effect curves for glibenclamide were repeated in the presence of mM concentrations of

the  $K_{ATP}$  channel blockers tolbutamide and meglitinide and the openers diazoxide and P1075. Fig.2 shows that the glibenclamide curve was merely shifted up by these compounds without increase in apparent affinity (i.e.,  $K_{1/2}$  value). Thus, effects on P-gp were cumulative without hints of the kind of interactions observed with SUR [(49);(50)]. Surprisingly, neither meglitinide nor tolbutamide competed with glibenclamide in spite of their very similar chemical structures.

**Interactions between glibenclamide and established P-gp substrates.** The compounds verapamil, progesterone, and vinblastine are typical P-gp substrates (Fig.1A) presumed to bind at distinct position within the P-gp binding pocket [(28);(51)]. As such, they constitute useful test drugs to characterize the binding of novel P-gp ligands such as those reported above. Each drug was therefore assayed in combination with selected  $K_{ATP}$  channel modulators.

Experiments combining glibenclamide and verapamil, shown in Fig.3A and B, demonstrate that glibenclamide inhibited the verapamil-induced increase in ATPase activity, and vice versa. This inhibition was not accompanied by a significant increase in  $K_{1/2}$  value of the control substance, implying a non-competitive interaction. Indeed, a “pure” competitive inhibition by 10  $\mu$ M glibenclamide or by 1  $\mu$ M verapamil, concentrations corresponding to their respective affinities, should have induced a 2-fold increase of  $K_{1/2}$  value for verapamil and glibenclamide, respectively. The dose-dependency of the inhibition correlated well with the respective  $K_{1/2}$  values of glibenclamide and verapamil observed for basal ATPase stimulation (Fig.1A and B). Indeed, the  $K_{1/2}$  value for verapamil activation of basal ATPase was close to 1  $\mu$ M, and the same concentration was sufficient to reduce by half the glibenclamide-induced increase in ATPase activity (Fig.3B). Conversely, the  $K_{1/2}$  value for glibenclamide activation of basal ATPase was near 10  $\mu$ M and Fig.3A shows that, in the presence of that concentration of glibenclamide, the verapamil-induced increase in ATPase activity was roughly half what it was in control. This correlation suggests that a single glibenclamide binding site mediates both activation of basal ATPase and inhibition of verapamil-induced ATPase.

A similar observation was made when glibenclamide was tested against progesterone as shown in Fig.4A: glibenclamide inhibited the effects

of progesterone with a half-maximal concentration close to 10  $\mu$ M. Curve fitting of the ATPase activity vs. progesterone concentration data yielded  $K_{1/2}$  values decreasing from 17  $\mu$ M in absence of glibenclamide to a non-significantly different 12  $\mu$ M in 10  $\mu$ M glibenclamide. Lower  $K_{1/2}$  values of 1.8 and 1  $\mu$ M were obtained in 100 and 300  $\mu$ M glibenclamide, respectively, but the shallow nature of the curves makes these values too uncertain to consider this decrease as significant and infer cooperativity rather than non-competitive inhibition.

The effects of vinblastine on the ATPase vs. glibenclamide concentration curve (Fig.5A) were also inhibitory, with a dose-dependent increase in  $K_{1/2}$  (from 11  $\mu$ M in control to 19  $\mu$ M at 1  $\mu$ M vinblastine) suggestive of a weak competitive interaction. Vinblastine inhibited the ATPase stimulation elicited by glibenclamide with full inhibition achieved at ~10  $\mu$ M vinblastine, a concentration which, by itself, did not affect basal activity of P-gp (see Fig.1A). Half-maximal effect was observed at ~1  $\mu$ M, which matches values previously determined from the inhibition by vinblastine of other drug-stimulated ATPase activities: verapamil (28), nicardipine (29), indinavir (52).

**Interactions between  $K_{ATP}$  channel openers and established P-gp substrates.** The  $K_{ATP}$  channel openers diazoxide and P1075, unlike glibenclamide, shifted upward the ATPase vs. verapamil concentration curve without modifying significantly the  $K_{1/2}$  of verapamil (Fig.3C). We therefore conclude that the ATPase stimulation by these openers and verapamil were mostly additive, suggesting separate binding sites on P-gp. SR47063 increased the  $K_{1/2}$  of verapamil from 1.1 to 1.9  $\mu$ M, a slight change considering the data points scattering, and it caused a decrease in  $V_{max}$  from 76% to 41%. Although this  $V_{max}$  decrease could reflect saturation of the verapamil effect due to the large stimulation already elicited by SR47063 in those particular experiments, these results suggest a weak competitive inhibition between verapamil and SR47063. Experiments with progesterone (Fig.4B) revealed that stimulation by diazoxide and progesterone were additive, whereas P1075 and SR47063 produced a decrease in the  $V_{max}$  of progesterone from 66% to 40% with a reduction in  $K_{1/2}$  with P1075 that appeared less significant because of experimental data uncertainties. This

indicates a difference in the mechanism of action of diazoxide and P1075 or SR47063.

Finally, as with glibenclamide, 10  $\mu$ M vinblastine fully inhibited the ATPase stimulation produced by SR47063 (Fig.5B), although the data with lower concentrations of vinblastine (not shown) did not permit to quantify changes in  $K_{1/2}$  indicative of competitive or non-competitive interactions.

**Analysis of molecular recognition elements in glibenclamide and SR47063 structures.** Using SYBYL software, we identified hydrophilic (i.e., electron donor and acceptor groups) and hydrophobic (i.e., alkyl and aromatic groups) elements contained in the molecular structures of the two P-gp ligands, glibenclamide and SR47063. SR47063, which exhibits a competitive interaction with verapamil, has in common with verapamil an electron-donor group (CN) and an aromatic group (Fig.6A). Glibenclamide, which exhibits a competitive interaction with vinblastine, presents five parts in common with vinblastine when considering their 3D-alignment: two aromatic cycles, two hydrogen donor groups (NH) and one electron donor group (CO and SO<sub>2</sub>) (Fig.6B). We have previously shown that multispecific ligand interaction with P-gp relies on the recognition of a 3D pattern of consensus hydrophilic and hydrophobic elements presented by these ligands, independently of their precise chemical structures (11). Furthermore, multidrug recognition capacity is amplified by the existence of different 3D patterns, each defining a pharmacophore on P-gp. Indeed, according to the multiparmacophoric model of the multidrug recognition domain of P-gp established on the basis of mutual interactions between several P-gp ligands, verapamil and vinblastine were identified as representatives of two distinct pharmacophores, pharmacophores I and II, respectively (11). The combined enzymological and structural evidence presented here suggests that SR47063 and glibenclamide belong to the pharmacophores I and II, respectively (Fig.6C). Indeed, the aromatic and the electron-donor groups of SR47063 and of glibenclamide correspond to the consensus elements defining these two pharmacophores.

These data about pharmacophoric-type recognition processes by P-gp have prompted us to investigate the structural features of the P-gp protein involved in the selective molecular recognition of its various ligands.

#### **Phylogenetic study of membrane ABC proteins.**

The phylogenetic tree of a set of membrane ABC proteins composed of the membrane human ABC transporters and two bacterial ABC transporters of known structure is shown in Fig.7. Its branches split along the five human subfamilies, A, B, C, D and G, previously defined on the basis of sequence similarities signing divergent evolution (3). Using either the N- or C-terminal sequence of the monomeric full-size ABC transporters in the alignment did not change the topology of the tree (not shown). Based on the high bootstrap values, the phylogenetic tree could be interpreted with a good confidence. Two distinct “clusters” are found. They represent ABCA-G and ABCB-C-D. It appears that P-gp (ABCB subfamily) and SUR1/2 (ABCC subfamily) belong to the same cluster. Moreover, inside this cluster, the ABCC branch is closer to the ABCB than the other branches in term of sequence evolution, as indicated by the length of the (horizontal) line defining these branches.

Interestingly, the two multispecific bacterial transporters MsbA and Sav1866 belong to the ABCB branch, in close proximity to P-gp (ABCB1), as well as ABCB4 (MDR3) and ABCB11 (BSEP). This confirms that MsbA and Sav1866 can be considered as suitable templates for modeling P-gp structure by homology.

#### **Molecular modeling of P-gp.**

- *Selection and assessment of the final models.* All the modeled structures generated, for P-gp as well as for SUR, were validated by GA341 scores >0.9. The selection of the “best” models (one for each conformation) was performed using the DOPE score, even though we are aware that it was initially designed for globular soluble proteins, and is applied to membrane proteins for lack of more suitable tools. The quantitative assessment of the selected models is presented in Table S2. Even for the two inward-facing conformations, constructed from the Ec- and Vc-MsbA low-resolution C $\alpha$ -positions-only structures, all models show very satisfying quality, as evidenced by Ramachandran plots that show in all cases that >90% of the residues are in favorable or allowed regions.

- *Structural model of P-gp.* From the three conformations of the bacterial ABC transporters (2 nucleotide-free “apo conformations”: “open inward-facing” and “closed inward-facing”; 1 nucleotide-bound: “outward-facing”), we generated three different P-gp structural models, that could reflect

different enzymological states of the active transporter during its catalytic turn-over (see coordinates files in pdb format in supplemental data).

The open inward-facing conformation model (Fig.S3), derived from the template Ec-MsbA (20), displays the following features: (i) the two NBDs are far apart (distance between P-loop1 [Ser429] and P-loop2 [Ser1072]  $\approx 48.4$  Å, mean distance between P-loop and LSGGQ segment [Leu531/Leu1176]  $\approx 65.4$  Å); (ii) the transmembrane helices TM1-2 and 4-5 are not exactly symmetrical compared with their C-terminal counterparts, helices TM7-8 and 10-11, respectively; (iii) the TM4-5/10-11 are intertwined between the two halves of the protein; (iv) all the helices of the transmembrane domain are in continuity with their corresponding intracellular helices; (v) these helices, oblique with respect to the membrane plane, form an internal cavity (“chamber”) lined by the TM4, 6, 7, 10 and 12 at the level of the cytosolic leaflet of the membrane.

During the course of this work, the publication of an X-ray structure for murine P-gp (87% of identity with human P-gp) in the open inward-facing conformation (22), with a better resolution than that of Ec-MsbA and that of the previously published structures obtained by electron microscopy [(53);(54)], although solved “by halves” and phased on MsbA, allowed us to generate another model of this conformation with a better reliability (Fig.8A). For this reason, we will only consider this model in the following. Compared with P-gp modeled from Ec-MsbA, the main difference is that the two NBDs are closer (distance between P-loop1 and P-loop2  $\approx 35.0$  Å, mean distance between P-loop and LSGGQ segment  $\approx 23.7$  Å).

The closed inward-facing conformation model (Fig.8B), derived from the template Vc-MsbA (20), differs from the open inward-facing models by the following features: (i) the two NBDs are in close proximity (distance between P-loop1 and P-loop2  $\approx 8.0$  Å; but, mean distance between P-loop and LSGGQ segment  $\approx 36.5$  Å); (ii) the transmembrane helices are perpendicular to the membrane plane, and no longer form a “chamber”.

The outward-facing conformation models (Fig.8C&D), built from templates St-MsbA (20) and Sav1866 (21), differ from the inward-facing models: (i) the two NBDs are associated (distance between P-loop1 and P-loop2  $\approx 23.5$  Å; mean distance

between P-loop and LSGGQ segment  $\approx 11.6$  Å, indicating a rotation of the NBDs from their position in the closed inward-facing conformation); (ii) the transmembrane helices are oblique with respect to the membrane plane, with an orientation opposite to that in the open inward-facing conformation, and form an extracellularly-accessible pocket in the phospholipid membrane, that is best seen upon rotation of the model (Fig.8D). At variance with our model, the model of the same conformation elaborated by O'mara & Tieleman (55) presents a region TM1-EC1-TM2 which extends in the extracellular medium with a marked asymmetry between the 2 halves of the protein; The model of Ravna et al (56) also predicts a slight asymmetry between the two halves of the protein, which is not apparent in our model. For an unobstructed view of the protein structure, the models are presented devoid of their membrane embedding in Fig.S4.

- *3D localization of P-gp residues involved in substrate recognition.* The P-gp models permit spatial localization of 52 residues (Table 3) reported to be involved in substrate recognition, as previously collected from point mutagenesis and chemical labeling data both for human and for mice P-gp [(10);(57)] as well as from crystallographic data (22). In Fig.9A displaying P-gp model in the open inward-facing conformation, the top-view representation reveals that the 52 residues are not distributed uniformly around the internal chamber, suggesting that the different helices do not participate equally in drug recognition. It is noteworthy that all important residues (i) are placed at the level of the phospholipid membrane, which is determined with a good confidence since it results from a simulation of the molecular interactions between the lipids and the protein (and not from the prediction from a simple hydropathy plot), and (ii) are oriented towards the interior of the protein cavity, the so-called internal chamber. In addition, in a given TM helix (except TM11), important residues are not uniformly distributed but appear clustered at the level of either the cytosolic or the exoplasmic leaflet. This non-uniform distribution reinforces the notion of “hotspots for drug binding” previously developed by Shilling et al. (10) (see below). Finally, the open inward-facing conformation of P-gp being the drug binding competent form of the protein, the drug recognition specificity of all the experimentally documented residues allows to identify binding hotspots for verapamil (TM 4E, 5E,

6E, 7E, 10C, 11C and 12E) and for vinblastine (TM 1C, 6E, 11C, 11E), where the E and C suffixes designate the exoplasmic and cytosolic halves of the transmembrane helices, respectively.

In Fig.9B, the same residues are highlighted in the model of the closed inward-facing conformation. As a consequence of helices motion, they now even more clearly segregate between “lower” and “upper” hotspots reminiscent of the “hotspot model” proposed by Shilling et al. (10), especially when considering the top-view of the structure.

In the outward-facing conformation of Fig.9C, the important residues are also distributed between “lower” and “upper” hotspots. We can thus divide these hotspots into 4 classes: (i) “cytosolic N-terminal”, located in TM1, (ii) “cytosolic C-terminal” located in TM10-11, (iii) “exoplasmic N-terminal”, located in TM4-5-6, (iv) “exoplasmic C-terminal”, located in TM7-11-12. These 4 hotspot classes include 49 out of the 52 selected residues. The 3 outsider residues (354, 937 and 938) are rather isolated and located at the membrane interface. From the inward-facing conformations to the outward-facing one, there is a significant reorientation of the residues of helices TM1-6-7-12 due to a rotation of these helices consistent with recent interpretation of crystallographic data (58).

**Molecular modeling of SUR.** SUR1 in the outward-facing conformation was modelled with the St-MsbA and Sav1866 templates (see coordinates files in pdb format in supplemental data). This model (Fig.10A) based on ~20% sequence identity was of good quality in terms of ProSA and QMEAN scores and Ramachandran analysis.(Table S2). It reveals the position in the transmembrane domains of the residues previously shown to be involved in the binding of  $K_{ATP}$  channel openers (12), namely Thr1285 and Met1289, and blockers (13), namely Ser1237 (Fig.10B). Their orientation towards the interior of the protein appears consistent with their presumed role. As already apparent in the alignment of Fig.S2, these residues, located in the upper (TM12E) and the lower part of the transmembrane domain (Fig.10C), correspond to P-gp residues of the above-characterized hotspots.

## DISCUSSION

Evidence has been reported that the sulfonyleurea glibenclamide is a substrate of P-gp (16). We show here, using an ATPase measurement assay, that

other ligands of the ABC protein SUR that act as  $K_{ATP}$  channel blockers and openers are also able to interact with the ABC protein P-gp. P-gp is a multidrug transporter responsible for the MDR phenotype of some tumor cells, and as such it is able to recognize and transport a broad panel of structurally unrelated molecules. The multispecific recognition of various substrates by P-gp constitutes the first step in the enzymatic cycle leading to the ATP hydrolysis-coupled expelling of the drugs out of the cell. However, multidrug recognition by P-gp, as well as by SUR, remains mysterious at the molecular level even after publication of several structures of multidrug transporters. Here we show that useful insight can be gained by comparing the binding properties and structures of these two rather homologous ABC proteins. Our strategy was to gather new information on shared ligands, develop models based on the latest high-resolution ABC proteins structures, and analyze the large body of existing functional, biochemical, and mutagenetic data in order to identify the structural determinants of drug recognition by both P-gp and SUR. This permitted to evidence and characterize hotspots for drug binding consistent with a multipharmacophoric model with a similar architecture for P-gp and SUR.

**$K_{ATP}$  channels blockers and P-glycoprotein.** As representative  $K_{ATP}$  channel blockers, we tested glibenclamide, tolbutamide and meglitinide. The sulfonyleurea glibenclamide was the most potent ( $K_{1/2} \sim 11 \mu\text{M}$ ) at stimulating ATPase activity of P-gp. Meglitinide which represents the non-sulfonyleurea half of glibenclamide had also a clear effect ( $K_{1/2} \sim 300 \mu\text{M}$ ). Tolbutamide, the sulfonyleurea half of glibenclamide, was the weakest agonist ( $K_{1/2} \sim 3 \text{mM}$ ). The Hill coefficients for all three molecules were close to 1, suggesting a single binding site linked to ATPase activation.

Although concentrations for half-maximal stimulation of P-gp ATPase were lower by 2 to 3 orders of magnitude than those necessary to bind to SUR and block  $K_{ATP}$  channels, the rank order of potency of these compounds was the same for P-gp and SUR1 [(49);(47)]. There appeared to be no obvious interactions between these 3 compounds in modulating P-gp ATPase since when tolbutamide or meglitinide – at high concentrations – were combined with glibenclamide, the resulting increase in ATPase activity was merely the sum of the increases elicited by each drug with no hint of competition. This finding could be considered as

surprising since it rules out any overlap between the binding sites of these 3 molecules in spite of the large degree of resemblance of their chemical structures, since the benzoic acid derivative meglitinide and the sulfonylurea tolbutamide roughly represent the 2 halves of the glibenclamide molecule. This observation is reminiscent of what has been previously reported on P-gp when studying steroids (27) and dihydropyridines(29). It reveals the versatility of the possible interactions of P-gp drug binding domain with the various hydrophilic/hydrophobic recognition elements of its ligands. To explain the much higher channel blocking affinity of glibenclamide on SUR1 and its very slow dissociation rate, it has been postulated that SUR1 possesses a benzoic acid site and a sulfonylurea site, both sites being separate but close enough to allow the binding of a single molecule of glibenclamide [(49);(13)]. Such a scheme cannot apply to P-gp as it predicts a competitive behavior between glibenclamide and tolbutamide or meglitinide that we have not observed. Instead, evidence points to separate non-interacting sites on P-gp for each of the 3 molecules. This conclusion implies that one should not discard the hypothesis that SUR could also harbors distinct, though mutually exclusive sites for these molecules (see below).

To examine the relations between the site of action of glibenclamide on P-gp and the sites of established ligands of P-gp, we performed experiments combining both glibenclamide and either verapamil, progesterone, or vinblastine. Results are recapitulated in Fig.6C. Verapamil reduced the glibenclamide-induced increase in ATPase activity in a dose-dependent manner without significantly shifting the glibenclamide  $K_{1/2}$ . Antagonism of activation by glibenclamide and increase in basal ATPase activity were both half-maximal at about 1  $\mu$ M verapamil, indicating that the same verapamil binding site mediates the two phenomena. The opposite experiment - effect of glibenclamide on verapamil-induced activity - similarly supports the existence of a single binding site for glibenclamide on P-gp. These mutually non-competitive inhibitions show that glibenclamide and verapamil bind to different, though allosterically connected, sites. The same could be concluded from the observed non-competitive antagonism of the progesterone-induced ATPase activity by glibenclamide.

Vinblastine also antagonized the glibenclamide-stimulated ATPase activity much like it did for the verapamil-stimulated ATPase [(28);(51)]. Full inhibition was obtained at a vinblastine concentration of 10  $\mu$ M that had by itself little impact on P-gp basal ATPase. Half-inhibition of the glibenclamide-stimulated ATPase activity was reached at ~1  $\mu$ M vinblastine whereas half-inhibition of the basal activity was at 30  $\mu$ M. There was also a moderate (about 2 fold) increase by 1  $\mu$ M vinblastine of the  $K_{1/2}$  of glibenclamide, indicative of a competition between vinblastine and glibenclamide. Assuming a 1:1 stoichiometry for this mutual exclusion leads to a  $K_i$  value for vinblastine antagonism of glibenclamide stimulation of about 1  $\mu$ M, which reinforces the existence of a high-affinity vinblastine binding site on P-gp, consistent with previous observations on the inhibition by vinblastine of P-gp ATPase stimulated by various P-gp substrates [(28);(29);(52)]. The observed inhibition of the basal ATPase activity by higher vinblastine concentrations is thus probably of a similar nature as the inhibition observed for verapamil above 50  $\mu$ M or progesterone above 100  $\mu$ M. The competition between glibenclamide and vinblastine could arise from a partial site overlap, leading to mutually exclusive steric constraints, since they possess the common hydrophilic/hydrophobic recognition elements of pharmacophore II (11), and likely bind to the same structural determinants of P-gp (see below). The absence of competition between glibenclamide and verapamil, whereas verapamil binds close to vinblastine and displays with it mutual exclusion by steric constraints without common recognition element (11), can be explained by the fact that glibenclamide molecule is smaller than vinblastine (Fig.6C).

***K<sub>ATP</sub> channels openers and P-glycoprotein.*** Three prototypical  $K_{ATP}$  channel openers were tested: SR47063, an analogue of cromakalim, P1075, an analogue of pinacidil, and diazoxide. The former two are known ligands of the SUR2 isoform, probably sharing overlapping binding sites (12) while the latter targets another site on both SUR1 and SUR2 isoforms (48). SR47063, P1075, and diazoxide augmented P-gp ATPase activity with  $K_{1/2}$  values of 36  $\mu$ M, 1.5 mM and 0.6 mM, respectively, and Hill coefficients near 1, indicative of a single binding site on P-gp for each of them. In comparison, radioligand binding assays on the

SUR2B isoform (26) have yielded dissociation constants of ~300 nM for levromakalim, a molecule close in structure and activity to SR47063, ~10 nM for P1075 and ~15  $\mu$ M for diazoxide. Values for the SUR2A isoforms were about fivefold greater, while for the SUR1 isoform a dissociation constant of ~150  $\mu$ M was registered for diazoxide and in the 1 mM range for the other 2 openers. Whatever the SUR isoform considered, there appears to be therefore little correlation between SUR and P-gp in terms of affinities for openers and rank order of potencies.

Binding experiments on various SUR isoforms have shown an apparent competition between  $K_{ATP}$  channel openers and blockers [for review see (7)]. Such competition was not seen with P-gp between the two openers tested, P1075 and diazoxide, and glibenclamide, since the effects of these openers on the glibenclamide-stimulated ATPase activity were strictly cumulative. We infer that sites for openers and blockers are located in different regions of P-gp, as they probably are for SUR (50).

Combined experiments with verapamil indicate no evidence of competition between verapamil and P1075 or diazoxide. This suggests no overlap between the sites for P1075 and diazoxide and verapamil. Further tests showed a non-competitive inhibition of progesterone-stimulated activity by P1075 or SR47063 but no interaction between progesterone and diazoxide. Therefore, diazoxide and P1075/SR47063 interact differently with P-gp than they do with SUR. The experiment concerning SR47063 points to a moderate competitive inhibition between verapamil and SR47063 and to a clear inhibition by vinblastine of the SR47063-stimulated activity, although the type of inhibition between these two drugs could not be determined. The existence of common recognition elements in the molecular structures of SR47063 and verapamil indicates that SR47063 may belong to the pharmacophore I, according to Garrigues et al (11). This is reinforced by the close structural resemblance of SR47063 with rhodamine 123, a compound for which recognition by P-gp has been determined to be competitive with verapamil (SM & SO, unpublished results) and to correspond to pharmacophore I (NL & SO, unpublished results). Due to the incomplete overlap with verapamil, SR47063 could sterically interfere with vinblastine, but this is difficult to ascertain experimentally. Fig.6C summarizes these observations in relation to

the two pharmacophores previously described on P-gp. The data suggest a striking functional homology between SUR and P-gp for recognition of the tested ligands, especially glibenclamide and SR47063.

***Implications for the molecular mechanisms of multispecific ligand recognition.*** Our results demonstrate a correlation between SUR and P-gp in terms of shared ligands. Does this correlation arise from the promiscuous nature of these ligands, which have a tendency to lack specificity (glibenclamide, for instance, has been postulated to be a universal ABC protein ligand (16)), or from the promiscuous nature of the target sites, or from a true structural homology between the target proteins, SUR and P-gp? Because the structural data on eukaryotic ABC proteins are limited to the recently published structure of mouse P-gp in one conformation (22), we performed an in-silico structural modeling of P-gp and SUR, based on the various crystallographic structures of the bacterial ABC transporters MsbA and Sav1866. To validate this choice of template homologues, we performed a phylogenetic analysis of a large set of membrane ABC proteins using a global analysis method based on a maximum likelihood algorithm, that is less dependent on varying sequence lengths and that provides a quantitative measure of protein homology (59). This analysis positioned MsbA and Sav1866 close to P-gp in the ABCB branch of the phylogenetic tree. In addition, the SUR-containing ABCC branch was the closest to the ABCB branch. This supports the choice of MsbA and Sav1866 as structural homologues. Furthermore, the various satisfactory quantitative assessments of the models obtained for P-gp as well as for SUR confirmed the pertinence of this choice. Indeed, the models correctly predict (i) the intramembrane position of the various residues known to be involved in substrate recognition and (ii) the orientation of their side chains towards the interior of the protein where they delimit a variable-size “chamber”, depending on the conformation of the protein.

In the case of P-gp, we modeled three different conformations, open inward-facing, closed inward-facing and outward-facing, which are supposed to represent successive states of the enzyme during its transport cycle (20). The comparison between the open conformation and the two others revealed that residues involved in substrate recognition are not spatially distributed evenly, but segregate between several hotspots mainly located either in the “lower”

/ “cytosolic” or “upper” / “exoplasmic” region of the phospholipid membrane. Analysis of the abundant mutagenetic and chemical labeling data for verapamil and vinblastine binding shows that the set of hotspots 4E/5E/6E/7E/10C/11C/12E could constitute anchoring points for verapamil and the ligands belonging to its competition class; similarly, the set of hotspots 1C/6E/11C/11E could constitute anchoring points for vinblastine and its competition class. Thus, these two sets of hotspots correspond to pharmacophores I and II, respectively (11), which provides a valuable link between the enzymological analysis and the structural analysis that reinforces the multisite model for the molecular mechanisms involved in multispecific drug recognition. This rough description of the drug recognition by P-gp cannot obviously replace a deep and careful analysis of ligand binding using in-silico docking techniques, which would require a massive effort out of the scope of this report. However, this allows to propose the following interpretations. (i) The two hotspots, 6E and 11C, shared by verapamil and vinblastine, could represent the common point between the two pharmacophores that we have previously reported (11). (ii) The higher number of hotspots forming the verapamil recognition pattern could indicate that P-gp is able to bind simultaneously two verapamil molecules, of limited size, as observed for QZ-59SSS (22), but only one vinblastine molecule, which is much larger. (iii) The residues that interact with the two peptides QZ-59RRR and QZ-59SSS in the crystallographic structure of P-gp are mostly located outside the hotspots; not surprisingly, since these peculiar ligands do not share the hydrophilic/hydrophobic recognition elements of either verapamil or vinblastine (not shown). (iv) By comparing the three P-gp conformations, it appears that some of the hotspots do not move within the membrane plane (TM 6, 7, 8, 12) while others do (TM 1, 4, 5, 10, 11); this is true for both sets of hotspots linked to verapamil and vinblastine, and is consistent with the required affinity alteration for the transported substrate during the enzymatic turn-over of a typical active transporter. (v) The location of the hotspots within the membrane bilayer, found both at the level of the cytosolic leaflet and the exoplasmic leaflet, is consistent with the two dominant functional models proposed in the literature for P-gp, i.e. a membrane floppase, and a “vacuum-cleaner” expelling the transported substrate directly out of the membrane.

In the case of SUR1, we built a structural model for the outward-facing conformation, which is supposed to be the physiologically relevant enzymatic state since it is the nucleotide-bound form. SR47063 binding to SUR1 involves a protein region that corresponds to the “upper sites” of P-gp, namely residues Thr1285 and Met1289 [(12);(9)] that are equivalent to P-gp residues Ser979 and Phe983. Glibenclamide binding to SUR1 involves Ser1237 (13), equivalent to P-gp Ser931 located at the membrane interface just beyond the “lower sites” which are compatible with the size of the average conformation of free glibenclamide (18 Å). Thus, the two ligands present binding sites located in corresponding regions of SUR and P-gp when aligned. As a whole, these two multispecific ABC membrane proteins appear to harbor a common architecture of binding sites for a pattern of structurally diverse ligands.

Because the openers can augment the ATPase activity of SUR2 (19), we have hypothesized that these ligands could act as “pseudo-transport” substrates of SUR2 (12). The finding that the openers SR47063 and P1075 stimulate the ATPase activity of P-gp, suggests not only structural but also functional homology between P-gp and SUR.

Beyond these similarities, SUR and P-gp display distinct characteristics for drug recognition and binding: (i) The variety of ligands of P-gp exceeds that of SUR [(60);(61);(62)]; (ii) P-gp binds larger ligands (up to 1250 Da (60)) than SUR (up to 500 Da for glibenclamide (62)); (iii) the smallest P-gp ligands (in the Mw range of SUR ligands) have poor affinities (beyond the  $\mu$ M range (11)) whereas the smallest SUR ligands can have high affinity; (iv) P-gp ligands often display non-competitive interactions except if they have sufficient size (>ca 700 Da) (11), whereas SUR ligands are generally always competitive one with each other, even when they bind distinct sites [(63);(47)]. All these features can be reconciled in a model in which the volume of the “cavity” forming the P-gp multidrug binding domain is much larger than that presented by SUR. Schematically, in P-gp, the multidrug binding domain forms a continuous multispecific pocket, whereas in SUR it appears rather restrained to (at least) two discrete multispecific sites, even though the two multidrug binding domains appear built around common hotspots.

The phylogenetic analysis of the membrane ABC proteins, supported by the high values of bootstrap

at several branches, shows the relative proximity of the various subfamilies. This analysis supports a close structural and functional relationship between P-gp, representative of the ABCB branch, and SUR, representative of the ABCC branch. In contrast, this cannot be extrapolated readily to the D branch, and a fortiori to the more remote A and G branches.

**Therapeutical implications.** Several  $K_{ATP}$  channel blockers and openers are in clinical use for a variety of conditions including diabetes and hypertension [(64);(8)]. Since P-gp is present and active at the intestinal and blood-brain barriers as well as of excretory systems [(17);(18)], knowing if and how a molecule is transported by P-gp is useful in anticipating pharmacokinetics and tissue distribution, and potential side-effects due to drug-drug interactions. In this work, we have measured the effects of blockers and openers on P-gp ATPase activity. The correlation between ATPase activity and transport is complex and not yet fully resolved [(65);(66)]. Molecules that augment P-gp ATPase activity are likely to be transported although ATPase activation does not necessarily imply a net transport (67). Of the molecules tested here, direct evidence of transport has only been supplied for glibenclamide (16). Such evidence would be difficult to obtain for the other molecules which are either not available in labeled form or have weaker affinities. If transport indeed occurs, one can infer that the bioavailability of these  $K_{ATP}$  channel modulators could be altered and that interferences with P-gp transport of co-administered drugs could lead to drug-drug interaction and eventual side effects. This is expected to be a significant factor for compounds that have comparable affinities for P-gp and their intended target, SUR. This is the case for SR47063, which activates muscle  $K_{ATP}$  channels at  $\mu$ M concentrations (68) and interacts with P-gp at the same concentrations. In contrast, the other compounds tested have much greater affinities for  $K_{ATP}$  channels than for P-gp, as exemplified by glibenclamide which blocks pancreatic  $K_{ATP}$  channels at nanomolar concentrations and activates P-gp ATPase activity at micromolar concentrations. We may anticipate that the capacity of these compounds to act across barriers should be high.

## REFERENCES

1. Higgins, C. F. (1992) *Annu Rev Cell Biol* **8**, 67-113
2. Holland, I. B., and Blight, M. A. (1999) *J Mol Biol* **293**(2), 381-399
3. Vasiliou, V., Vasiliou, K., and Nebert, D. W. (2009) *Hum Genomics* **3**(3), 281-290
4. Inagaki, N., Gono, T., Clement, J. P. t., Namba, N., Inazawa, J., Gonzalez, G., Aguilar-Bryan, L., Seino, S., and Bryan, J. (1995) *Science* **270**(5239), 1166-1170
5. Antcliff, J. F., Haider, S., Proks, P., Sansom, M. S., and Ashcroft, F. M. (2005) *Embo J* **24**(2), 229-239
6. Shi, N. Q., Ye, B., and Makielski, J. C. (2005) *J Mol Cell Cardiol* **39**(1), 51-60
7. Moreau, C., Prost, A. L., Derand, R., and Vivaudou, M. (2005) *J Mol Cell Cardiol* **38**(6), 951-963
8. Jahangir, A., and Terzic, A. (2005) *J Mol Cell Cardiol* **39**(1), 99-112
9. Moreau, C., Gally, F., Jacquet-Bouix, H., and Vivaudou, M. (2005) *Mol Pharmacol* **67**(4), 1026-1033
10. Shilling, R. A., Venter, H., Velamakanni, S., Bapna, A., Woebking, B., Shahi, S., and van Veen, H. W. (2006) *Trends Pharmacol Sci* **27**(4), 195-203
11. Garrigues, A., Loiseau, N., Delaforge, M., Ferte, J., Garrigos, M., Andre, F., and Orłowski, S. (2002) *Mol Pharmacol* **62**(6), 1288-1298
12. Moreau, C., Jacquet, H., Prost, A. L., D'Hahan, N., and Vivaudou, M. (2000) *Embo J* **19**(24), 6644-6651
13. Ashfield, R., Gribble, F. M., Ashcroft, S. J., and Ashcroft, F. M. (1999) *Diabetes* **48**(6), 1341-1347
14. Sheppard, D. N., and Welsh, M. J. (1992) *J Gen Physiol* **100**(4), 573-591
15. Payen, L., Delugin, L., Courtois, A., Trinquart, Y., Guillouzo, A., and Fardel, O. (2001) *Br J Pharmacol* **132**(3), 778-784
16. Golstein, P. E., Boom, A., van Geffel, J., Jacobs, P., Masereel, B., and Beauwens, R. (1999) *Pflugers Arch* **437**(5), 652-660
17. Fromm, M. F. (2004) *Trends Pharmacol Sci* **25**(8), 423-429
18. Ambudkar, S. V., Dey, S., Hrycyna, C. A., Ramachandra, M., Pastan, I., and Gottesman, M. M. (1999) *Annu Rev Pharmacol Toxicol* **39**, 361-398
19. Bienengraeber, M., Alekseev, A. E., Abraham, M. R., Carrasco, A. J., Moreau, C., Vivaudou, M., Dzeja, P. P., and Terzic, A. (2000) *Faseb J* **14**(13), 1943-1952
20. Ward, A., Reyes, C. L., Yu, J., Roth, C. B., and Chang, G. (2007) *Proc Natl Acad Sci U S A* **104**(48), 19005-19010
21. Dawson, R. J., and Locher, K. P. (2006) *Nature* **443**(7108), 180-185
22. Aller, S. G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P. M., Trinh, Y. T., Zhang, Q., Urbatsch, I. L., and Chang, G. (2009) *Science* **323**(5922), 1718-1722
23. Garrigos, M., Belehradek, J., Jr., Mir, L. M., and Orłowski, S. (1993) *Biochem Biophys Res Commun* **196**(3), 1034-1041
24. de Wet, H., Fotinou, C., Amad, N., Dreger, M., and Ashcroft, F. M. (2010) *Febs J* **277**(12), 2654-2662
25. de Wet, H., Mikhailov, M. V., Fotinou, C., Dreger, M., Craig, T. J., Venien-Bryan, C., and Ashcroft, F. M. (2007) *Febs J* **274**(14), 3532-3544
26. Schwanstecher, M., Sieverding, C., Dorschner, H., Gross, I., Aguilar-Bryan, L., Schwanstecher, C., and Bryan, J. (1998) *Embo J* **17**(19), 5529-5535
27. Orłowski, S., Mir, L. M., Belehradek, J., Jr., and Garrigos, M. (1996) *Biochem J* **317** ( Pt 2), 515-522
28. Garrigos, M., Mir, L. M., and Orłowski, S. (1997) *Eur J Biochem* **244**(2), 664-673
29. Pascaud, C., Garrigos, M., and Orłowski, S. (1998) *Biochem J* **333** ( Pt 2), 351-358
30. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. (2005) *Nucleic Acids Res* **33**(Database issue), D154-159

31. Le, T., Tseng, T. T., and Saier, M. H., Jr. (1999) *Mol Membr Biol* **16**(2), 173-179
32. Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., Liebert, C. A., Liu, C., Madej, T., Marchler, G. H., Mazumder, R., Nikolskaya, A. N., Panchenko, A. R., Rao, B. S., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Vasudevan, S., Wang, Y., Yamashita, R. A., Yin, J. J., and Bryant, S. H. (2003) *Nucleic Acids Res* **31**(1), 383-387
33. Morgenstern, B., Dress, A., and Werner, T. (1996) *Proc Natl Acad Sci U S A* **93**(22), 12098-12103
34. Guindon, S., and Gascuel, O. (2003) *Syst Biol* **52**(5), 696-704
35. Felsenstein, J. (1985) *Evolution* **39**, 783-791
36. Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics
37. Page, R. D. (1996) *Comput Appl Biosci* **12**(4), 357-358
38. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) *Bioinformatics* **20**(3), 426-427
39. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res* **25**(17), 3389-3402
40. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) *Proteins* **23**(3), 318-326
41. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J Biomol NMR* **8**(4), 477-486
42. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**(6580), 272
43. Sippl, M. J. (1993) *Proteins* **17**(4), 355-362
44. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) *Nucleic Acids Res* **32**(Web Server issue), W615-619
45. Benkert, P., Tosatto, S. C., and Schomburg, D. (2008) *Proteins* **71**(1), 261-277
46. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J Mol Graph* **14**(1), 33-38, 27-38
47. Dorschner, H., Brekardin, E., Uhde, I., Schwanstecher, C., and Schwanstecher, M. (1999) *Mol Pharmacol* **55**(6), 1060-1066
48. D'Hahan, N., Moreau, C., Prost, A. L., Jacquet, H., Alekseev, A. E., Terzic, A., and Vivaudou, M. (1999) *Proc Natl Acad Sci U S A* **96**(21), 12162-12167
49. Gribble, F. M., Tucker, S. J., Seino, S., and Ashcroft, F. M. (1998) *Diabetes* **47**(9), 1412-1418
50. Gribble, F. M., and Reimann, F. (2002) *Biochem Soc Trans* **30**(2), 333-339
51. Litman, T., Zeuthen, T., Skovsgaard, T., and Stein, W. D. (1997) *Biochim Biophys Acta* **1361**(2), 169-176
52. Megard, I., Garrigues, A., Orlowski, S., Jorajuria, S., Clayette, P., Ezan, E., and Mabondzo, A. (2002) *Brain Res* **927**(2), 153-167
53. Rosenberg, M. F., Kamis, A. B., Callaghan, R., Higgins, C. F., and Ford, R. C. (2003) *J Biol Chem* **278**(10), 8294-8299
54. Lee, J. Y., Urbatsch, I. L., Senior, A. E., and Wilkens, S. (2008) *J Biol Chem* **283**(9), 5769-5779
55. O'Mara, M. L., and Tieleman, D. P. (2007) *FEBS Lett* **581**(22), 4217-4222
56. Ravna, A. W., Sylte, I., and Sager, G. (2007) *Theor Biol Med Model* **4**, 33
57. Loo, T. W., Bartlett, M. C., and Clarke, D. M. (2006) *Biochem J* **399**(2), 351-359
58. Gutmann, D. A., Ward, A., Urbatsch, I. L., Chang, G., and van Veen, H. W. (2009) *Trends Biochem Sci* **35**(1), 36-42
59. Felsenstein, J. (1981) *J Mol Evol* **17**(6), 368-376
60. Avendano, C., and Menendez, J. C. (2002) *Curr Med Chem* **9**(2), 159-193
61. Bryan, J., Crane, A., Vila-Carriles, W. H., Babenko, A. P., and Aguilar-Bryan, L. (2005) *Curr Pharm Des* **11**(21), 2699-2716
62. Coghlan, M. J., Carroll, W. A., and Gopalakrishnan, M. (2001) *J Med Chem* **44**(11), 1627-1653
63. Loffler-Walz, C., and Quast, U. (1998) *Br J Pharmacol* **123**(7), 1395-1402
64. Nagashima, K., Takahashi, A., Ikeda, H., Hamasaki, A., Kuwamura, N., Yamada, Y., and Seino, Y. (2004) *Diabetes Res Clin Pract* **66** Suppl 1, S75-78
65. Sharom, F. J. (1997) *J Membr Biol* **160**(3), 161-175

66. Al-Shawi, M. K., Polar, M. K., Omote, H., and Figler, R. A. (2003) *J Biol Chem* **278**(52), 52629-52640
67. Polli, J. W., Wring, S. A., Humphreys, J. E., Huang, L., Morgan, J. B., Webster, L. O., and Serabjit-Singh, C. S. (2001) *J Pharmacol Exp Ther* **299**(2), 620-628
68. Forestier, C., Pierrard, J., and Vivaudou, M. (1996) *J Gen Physiol* **107**(4), 489-502

## FOOTNOTES

**Financial support:** HJ was supported by studentships from the « Ligue Nationale de Recherche contre le Cancer » and « Les Amis des Sciences ». SM was supported by studentships from SPI-Bio and ARC.

**Acknowledgements:** We are grateful to Dr. P. Gautier (Sanofi Recherche, Montpellier, France) for the generous gift of SR 47063, Dr. L. Billerup (Leo Pharmaceutical Products, Copenhagen, Denmark) for P1075, and Dr. J. Pünter (Aventis Pharma, Frankfurt, Germany) for meglitinide.

**Abbreviations:**  $K_{ATP}$  channels, ATP-sensitive potassium channels; SUR, sulfonylurea receptor; ABC, ATP-binding cassette; TM(D), transmembrane (domain); NBD, nucleotide binding domain; MRP, multidrug-resistance (associated) protein; MDR, multidrug resistance; P-gp, P-glycoprotein; ABCC, subfamily C of ABC proteins; VRP, verapamil; VBL, vinblastine; PRG, progesterone; GLB, glibenclamide, DZX, diazoxide; MGL, meglitinide; TLB, tolbutamide.

**Keywords:**  $K_{ATP}$  channels, potassium channel openers, potassium channel blockers, ABC transporters, sulfonylurea receptor, MDR1/P-glycoprotein, molecular modeling, hotspots.

## FIGURE LEGENDS

**Figure 1.** Drug concentration-dependence of stimulated P-gp ATPase activity.

**Panel A.** Effect of standard P-gp ligands on ATPase activity using P-gp-containing vesicles. The ATPase activity of P-gp-devoid (control, open symbols) and P-gp-enriched (filled symbols) native membrane vesicles was measured over a range of concentrations for 3 well-characterized P-gp substrates: verapamil (squares), progesterone (triangles), and vinblastine (circles). Dashed lines represent fits of the Hill equation to the initial ligand-induced changes in ATPase activity of P-gp-enriched vesicles. Concentrations for half-maximal effect ( $K_{1/2}$ ), Hill coefficient (h), and  $V_{max}$  were: 1.3  $\mu$ M, 0.94, and 73% for verapamil; 31  $\mu$ M, 1.35, and 53% for vinblastine; and 17  $\mu$ M, 2.1, and 66% for progesterone. Basal activity of P-gp-containing vesicles (i.e., 100% after normalization) was 120-300 nmol/min per mg of membrane protein while it was ~30 nmol/min/mg in control vesicles. When not visible, error bars were smaller than the symbols.

**Panel B.** Effect of SUR ligands that block  $K_{ATP}$  channels on ATPase activity using P-gp-containing vesicles, showing that SUR blockers are P-gp substrates with the same order of potency. ATPase activity of control (open symbols) and P-gp-enriched (filled symbols) vesicles was measured over a range of concentrations for the  $K_{ATP}$  channel blockers, glibenclamide (squares), tolbutamide (triangles), and meglitinide (circles). The chemical structures of these compounds are illustrated on the right. Dashed lines represent fits of the Hill equation to the initial ligand-induced increases in ATPase activity of P-gp-enriched vesicles.  $K_{1/2}$ , h, and  $V_{max}$  were: 11  $\mu$ M, 1.2, and 73% for glibenclamide; 276  $\mu$ M, 0.75, and 70% for meglitinide; 3 mM, 0.68, and 70% for tolbutamide. Basal activities of P-gp enriched vesicles used for the glibenclamide, tolbutamide, and meglitinide assays were on average 126, 96, and 114 nmol/min per mg of membrane protein, respectively.

**Panel C.** Effect of SUR ligands that open  $K_{ATP}$  channels on ATPase activity using P-gp-containing vesicles, showing that SUR openers are low-affinity P-gp substrates. ATPase activity of control (open symbols) and P-gp-enriched (filled symbols) vesicles was measured over a range of concentrations for the  $K_{ATP}$  channel openers, SR47063 (squares), P1075 (triangles), and diazoxide (circles). The chemical structures of these compounds are illustrated on the right. Dashed lines represent fits of the Hill equation to the ligand-induced increases in ATPase activity of P-gp-enriched vesicles.  $K_{1/2}$ , h, and  $V_{max}$  were: 36  $\mu$ M, 1.3, and 55% for SR47063; 610  $\mu$ M, 1, and 63% for diazoxide; 1.5 mM, 1.2, 101% for

P1075. Basal activities of P-gp enriched vesicles used for the SR47063, P1075, and diazoxide assays were on average 150, 138, and 114 nmol/min per mg of membrane protein, respectively.

**Figure 2.** Effect of various SUR ligands on the glibenclamide-induced stimulated P-gp ATPase activity, showing a lack of “cross-talk” between glibenclamide and other  $K_{ATP}$  blockers and openers. The effects of glibenclamide on P-gp ATPase activity were measured in the absence (squares) and presence of other  $K_{ATP}$  channel modifiers, meglitinide (1 mM, down triangles), tolbutamide (3 mM, diamonds), P1075 (1 mM, up triangles), or diazoxide (1 mM, circles). Dashed lines represent fits of the Hill equation to the glibenclamide-induced increases in ATPase activity.  $K_{1/2}$ ,  $h$ , and  $V_{max}$  were: 11  $\mu$ M, 1.2, and 73% in control; 11  $\mu$ M, 0.9, and 78% with meglitinide added; 6.7  $\mu$ M, 1.1, and 94% with tolbutamide; 6.5  $\mu$ M, 1.1, and 81% with P1075; 9.8  $\mu$ M, 1.3, and 69% with diazoxide.

**Figure 3.** Interactions between verapamil and  $K_{ATP}$  channel modulators. **Panel A.** Verapamil-induced ATPase activity measured in the absence (squares) or presence of 2 (triangles) or 10  $\mu$ M (diamonds) glibenclamide. Dashed lines represent fits of the Hill equation to the data points.  $K_{1/2}$ ,  $h$ , and  $V_{max}$  were: 1.3  $\mu$ M, 0.94, and 76% in control; 1.4  $\mu$ M, 0.9, and 49% with 2  $\mu$ M glibenclamide; 1.6  $\mu$ M, 1.7, and 27% with 10  $\mu$ M glibenclamide added. **Panel B.** Glibenclamide-induced ATPase activity measured in absence (squares) and presence of 0.1 (circles), 1 (diamonds) or 10  $\mu$ M (triangles) verapamil. Dashed lines represent fits of the Hill equation to the experimental data.  $K_{1/2}$ ,  $h$ , and  $V_{max}$  were: 11  $\mu$ M, 1.2, and 73% in control; 10.7  $\mu$ M, 1.3, and 62% with 0.1  $\mu$ M verapamil; 6.4  $\mu$ M, 1.8, and 28% with 1  $\mu$ M verapamil. **Panel C.** Verapamil-induced ATPase activity measured in the absence (squares) or presence of the  $K_{ATP}$  channel openers SR47063 (50  $\mu$ M, circles), P1075 (1 mM, triangles), or diazoxide (1 mM, diamonds). Fits shown as dashed lines yielded  $K_{1/2}$ ,  $h$ , and  $V_{max}$  of 1.3  $\mu$ M, 0.94, and 76% in control; 1.9  $\mu$ M, 1.4, and 41% with SR47063; 1.2  $\mu$ M, 0.9, and 72% with P1075; 0.95  $\mu$ M, 1.2, and 77% with diazoxide.

**Figure 4.** Modulation of the progesterone-induced ATPase activity by  $K_{ATP}$  channel openers and blockers. **Panel A.** Effects of progesterone in the absence (squares) or presence of 10 (diamonds), 100  $\mu$ M (down triangles), or 300  $\mu$ M (up triangles) glibenclamide. Fits shown as dashed lines yielded  $K_{1/2}$ ,  $h$ , and  $V_{max}$  of 17  $\mu$ M, 2.1, and 66% in control; 12  $\mu$ M, 3, and 32% with 10  $\mu$ M glibenclamide; 1.8  $\mu$ M, 1, and 32% with 100  $\mu$ M glibenclamide; 1  $\mu$ M, 1.5, and 11% with 300  $\mu$ M glibenclamide. **Panel B.** Effects of progesterone in the absence (squares) or presence of 1 mM diazoxide (circles), 1 mM P1075 (triangles), or 40  $\mu$ M SR47063 (diamonds). Fits shown as dashed lines yielded  $K_{1/2}$ ,  $h$ , and  $V_{max}$  of 17  $\mu$ M, 2.1, and 66% in control; 12  $\mu$ M, 2.5, and 76% with diazoxide; 7  $\mu$ M, 1.5, and 40% with P1075; 15  $\mu$ M, 2, and 39% with SR47063.

**Figure 5.** Inhibition by vinblastine of ATPase stimulation produced by glibenclamide and SR47063. **Panel A.** Effects of glibenclamide on P-gp ATPase activity measured in the absence (squares) or presence of 0.1  $\mu$ M (triangles), 1  $\mu$ M (diamonds), or 10  $\mu$ M (circles) vinblastine. Dashed lines represent fits of the Hill equation to the glibenclamide-induced increases in ATPase activity.  $K_{1/2}$ ,  $h$ , and  $V_{max}$  were: 11  $\mu$ M, 1.2, and 73% in control; 15  $\mu$ M, 0.9, and 70% with 0.1  $\mu$ M vinblastine; 19  $\mu$ M, 1.2, and 39% with 1  $\mu$ M vinblastine; a fit was not possible for 10  $\mu$ M vinblastine, but  $V_{max}$  was estimated as ~5%. **Panel B.** Effects of SR47063 on P-gp ATPase activity measured in the absence (squares) and presence of 10  $\mu$ M (circles) vinblastine.

**Figure 6.** Pharmacophoric model of glibenclamide and SR47063. **Panel A.** 3D-alignment of SR47063 (blue sticks representation) and verapamil (red sticks) with surfaces highlighting the aligned aromatic groups (green) and the electron-donor CN group (blue) on both molecules. **Panel B.** 2D-structures of verapamil and SR47063 evidencing the common aromatic groups in green and electron-donor groups in blue. **Panel C.** 3D-alignment of glibenclamide (yellow sticks) and vinblastine (cyan sticks) with surfaces showing the aligned aromatic groups (green), the hydrogen-donor NH/OH groups (red) and the electron-donor CO/CO<sub>2</sub> groups (blue) on both molecules. **Panel D.** 2D-structures of glibenclamide and vinblastine evidencing the common aromatic groups in green, hydrogen-donor groups in red, and electron-donor groups in blue. **Panel E.** Schematic representation of the functional interactions between the binding sites on P-gp for  $K_{ATP}$  channel blockers (GLB, glibenclamide; MGL, meglitinide; TLB, tolbutamide; in dark grey), openers (SR, SR47063; DZX, diazoxide; P1075; in light grey), and

known P-gp substrates (VRP, verapamil; VBL, vinblastine; PRG, progesterone; in white). Solid arrow lines represent mutual exclusiveness indicated by the competitive inhibition observed between glibenclamide and vinblastine, and between SR47063 and verapamil. Connecting lines represent binding to separate sites: either mutual destabilization reflected by observed non-competitive inhibition, or mutual stabilization reflected by observed positive allosteric effect, or a lack of recorded interactions reflected by additive effects. The interaction between SR47063 and vinblastine is shown exclusively by a simple proximity since our data are insufficient to assess the competitive or non-competitive nature of the observed inhibition. The dotted boxes represent the two pharmacophores previously characterized (11).

**Figure 7.** A phylogenetic tree depicting the relationships between the human membrane ABC transporters and two bacterial ABC transporters. This maximum likelihood tree is based on 44 amino acid sequences (42 human and 2 bacterial). Five principal branches are distinguishable: ABCB Branch (red hook), ABCA Branch (blue hook), ABCC Branch (brown hook), ABCD Branch (green hook) and ABCG Branch (orange hook). P-gp (ABCB1), MsbA and Sav1866 are highlighted in a red box. SUR1 (ABCC8) is shown in a brown box. Only, the bootstrap values higher or equal to 90% are shown.

**Figure 8.** Side views of structural models of P-gp in a POPC membrane. **Panel A.** Open inward-facing conformation structural model, rebuilt from murine P-gp. **Panel B.** Closed inward-facing conformation structural model, rebuilt from *Vibrio cholera* MsbA. **Panel C.** Outward-facing conformation structural model, rebuilt from *Salmonella typhimurium* MsbA and *Staphylococcus aureus* Sav1866. **Panel D.** Same as 8C after 90° rotation around the symmetrical axis of the protein.

**Figure 9.** Magnification and top-view of the membrane region of P-gp conformations. Color code: black, hotspots of the transmembrane helix 1C, orange, hotspots of the transmembrane helix 11C, brown, hotspots of the transmembrane helix 10C, red, hotspots of the transmembrane helix 12E, marine, hotspots of the transmembrane helix 11E, magenta, hotspots of the transmembrane helix 7E, green, hotspots of the transmembrane helix 6E, blue, hotspots of the transmembrane helix 5E, purple blue, hotspots of the transmembrane helix 4E. **Panel A.** Open inward-facing conformation. **Panel B.** Closed inward-facing conformation. **Panel C.** Outward-facing conformation.

**Figure 10.** SUR1 structural model and 3D alignment of outward-facing models of P-gp and SUR1. **Panel A.** Outward-facing conformation model of SUR1. **Panel B.** Top-view of 3D transmembrane location of residues involved in the binding of  $K_{ATP}$  channel openers (Thr1285 and Met1289) and blockers (Ser1237). **Panel C.** Structural alignment of TM11 (P-gp)/ TM16 (SUR1) and TM12 (P-gp)/ TM17 (SUR1). SUR1 residues involved in openers and blockers binding are represented and labeled in green, P-gp TM12E hotspots in red, and P-gp TM11 residue in blue.

**Table 1:** Reconstructed P-gp models and their corresponding templates.

Vc: *Vibrio cholera*, Ec: *Escherichia coli*, St: *Salmonella typhimurium*, Sa: *Staphylococcus aureus*

<b>Templates (PDB code)</b>	<b>Resolution (Å)</b>	<b>Models</b>
Ec-MsbA (3B5W)	5.3	Pgp (Open Inward-Facing Conformation)
Vc-MsbA (3B5X)	5.5	Pgp (Closed Inward-Facing Conformation)
Sav1866 (2HYD)	3.0	Pgp (Outward-Facing Conformation)
St-MsbA (3B60)	3.7	
mouse Pgp (3G5U)	3.8	Pgp (Open Inward-Facing Conformation)
Sav1866 (2HYD)	3.0	SUR1 (Outward-Facing Conformation)
St-MsbA (3B60)	3.7	

**Table 2:** Enzymological parameters from P-gp ATPase activity measurements shown in Figs 1-5.

$K_{1/2}$ : the half-activating concentration of the tested compound on P-gp basal ATPase activity; (-) indicates that it is the half-inhibiting concentration since there is no apparent activation.

$K_i$ : the half-inhibiting concentration of the tested compound on P-gp stimulated ATPase activity, as determined from the concentration-dependence curves of the substrate indicated in brackets.

nd: not determined.

ME: mutually exclusive, that is the mutual relationship between two drug substrates revealed by a competitive inhibiting effect on P-gp ATPase activity;

NME: non-mutually exclusive, that is the mutual relationship between two drug substrates revealed by either a non-competitive inhibiting effect or an additive effect or a positive allosteric effect on P-gp ATPase activity.

	Drug studied at increasing concentrations	$K_{1/2}$ ( $\mu\text{M}$ )	$K_i$ ( $\mu\text{M}$ )	Mutual relationships (ME versus NME)
reference Pgp substrates	Verapamil (VRP)	1.3	$\approx 1$ (from GLB)	NME with GLB
	Progesterone (PRG)	17	nd	nd
	Vinblastine (VBL)	31 (-)	$\approx 1$ (from GLB)	ME with GLB
SUR blockers	Glibenclamide (GLB)	11	$\approx 10$ (from VRP) $\geq 10$ (from PRG)	NME with VRP, PRG
	Meglitinide (MGL)	280	nd	NME with GLB
	Tolbutamide (TLB)	3000	nd	NME with GLB
SUR openers	SR47063	36	$\geq 50$ (from VRP) $> 40$ (from PRG)	ME with VRP NME with PRG
	P1075	1500	nd	NME with GLB, VRP, PRG
	Diazoxide (DZX)	610	nd	NME with GLB, VRP, PRG

**Table 3:** Transmembrane localization of P-gp residues involved in substrate recognition.

P-gp residues identified from point mutagenesis, chemical labelling [(10);(57)] or crystallographic data (22). The equivalent residues belonging to MsbA, Sav1866 and SUR1 are identified thanks to the primary sequences alignment represented Fig.S2.

	P-gp	MsbA	SUR1	Sav1866	References
N-Terminal					
TM1	H61	I28	G16	F17	(10)
	G64	G31	C19	I20	(10)
	L65	I32	I20	I21	(10)
	M69	L36	V24	I25	(22)
TM4	S222	A175	I467	F171	(22)
TM5	L304	L257	A549	T253	(22)
	I306	A259	P551	T255	(10)
	Y307	S260	I552	D256	(22)
TM6	F335	F288	L580	V284	(10)
	F336	S289	S581	G285	(22)
	V338	M291	F583	L287	(10)
	L339	I292	H584	E288	(22)
	I340	A293	I585	L289	(22)
	G341	L294	L586	L290	(10)
	A342	M295	V587	F291	(22)
	F343	R296	T588	G292	(22)
	A354	Q307	S599	T303	(10)
C-Terminal					
TM7	Q725	N37	L1028	P34	(57)
	P726	A38	V1027	L35	(57)
	F728	S40	I1031	I37	(57)
	A729	D41	D1030	K38	(57)
	F732	M44	L1035	I41	(22)
TM8	L762	L54	L1070	F70	(22)
TM9	T837	V147	S1143	W141	(22)
	A841	A151	L1147	I145	(10)
	L843	I153	C1149	I147	(10)
TM10	P866	P176	I1172	P172	(10)
	I868	V178	C1174	Y174	(22)
	A869	S179	Y1175	I175	(10)
	G872	I182	Q1178	V178	(22)
TM11	I937	A247	A1243	W243	(10)
	F938	S248	N1244	N244	(10)
	F942	D252	E1248	F248	(10)
	S943	P253	V1249	A249	(10)
	T945	I255	M1251	F251	(10)
	Q946	Q256	E1252	N252	(10)
	Y950	S260	A1256	D256	(10)
	S952	A262	V1258	G258	(10)
	Y953	L263	V1259	P259	(10)
F957	L267	A1263	I263	(10)	
TM12	L975	T285	G1281	A281	(10)
	F978	F288	L1284	V284	(10)
	S979	S289	T1285	G285	(22)
	V981	M291	A1287	L287	(10)
	V982	I292	L1288	E288	(10)
	F983	A293	M1289	L289	(10)
	G984	L294	V1290	L290	(22)
	A985	M295	S1291	F291	(22)
	M986	R296	N1292	G292	(22)
	G989	K299	N1295	R295	(22)
	Q990	S300	W1296	R296	(22)
	S993	N303	R1299	A299	(22)

Fig.1

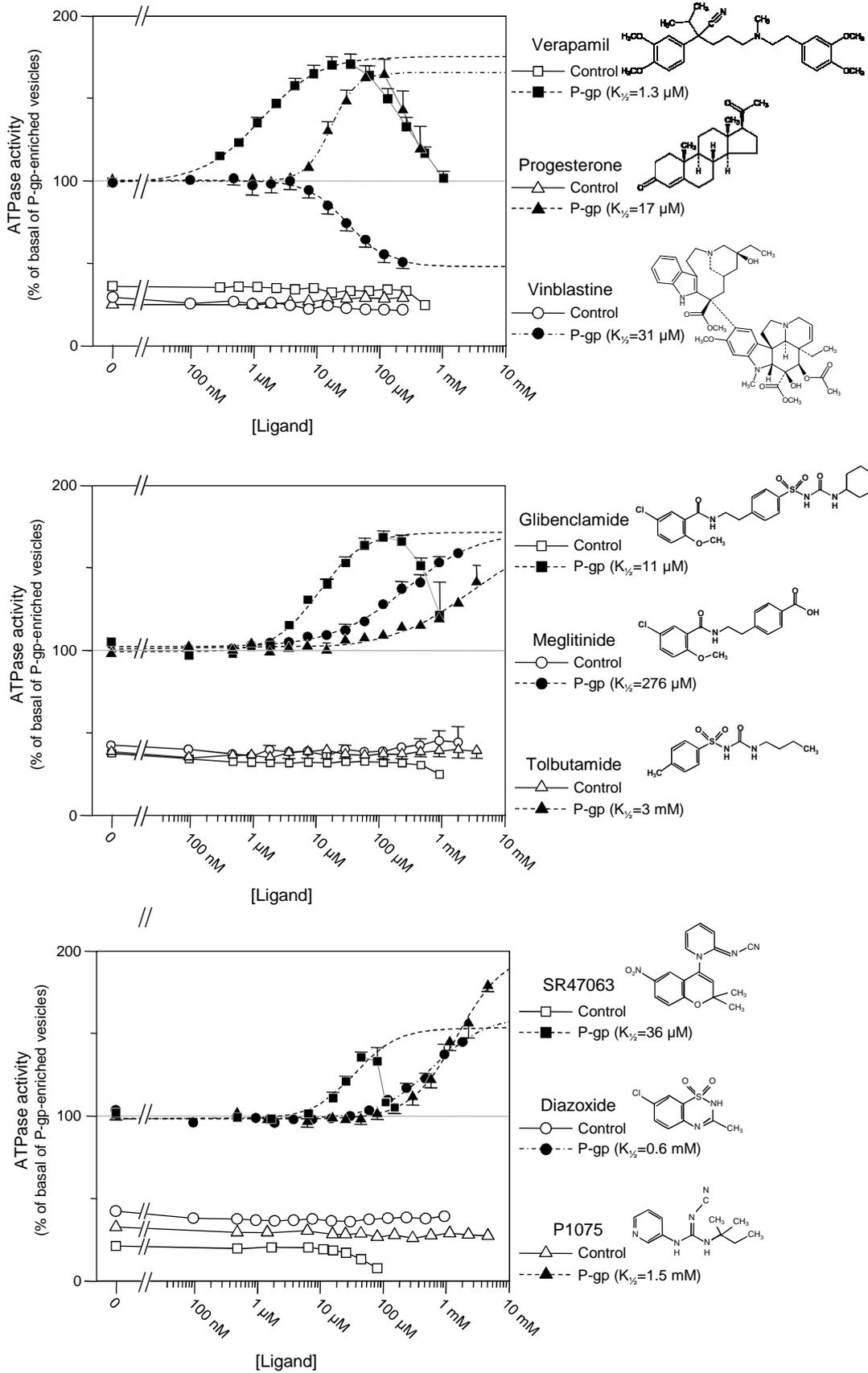


Fig.2

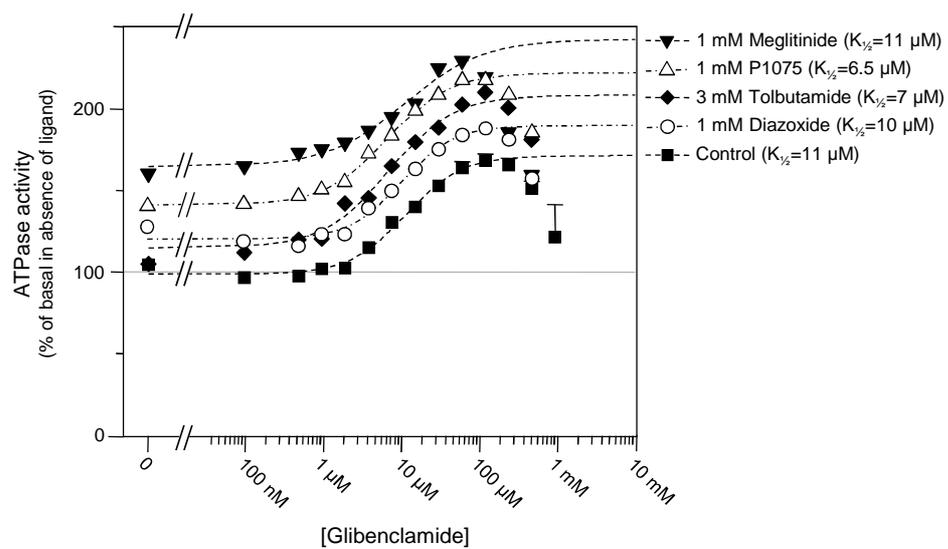


Fig.3

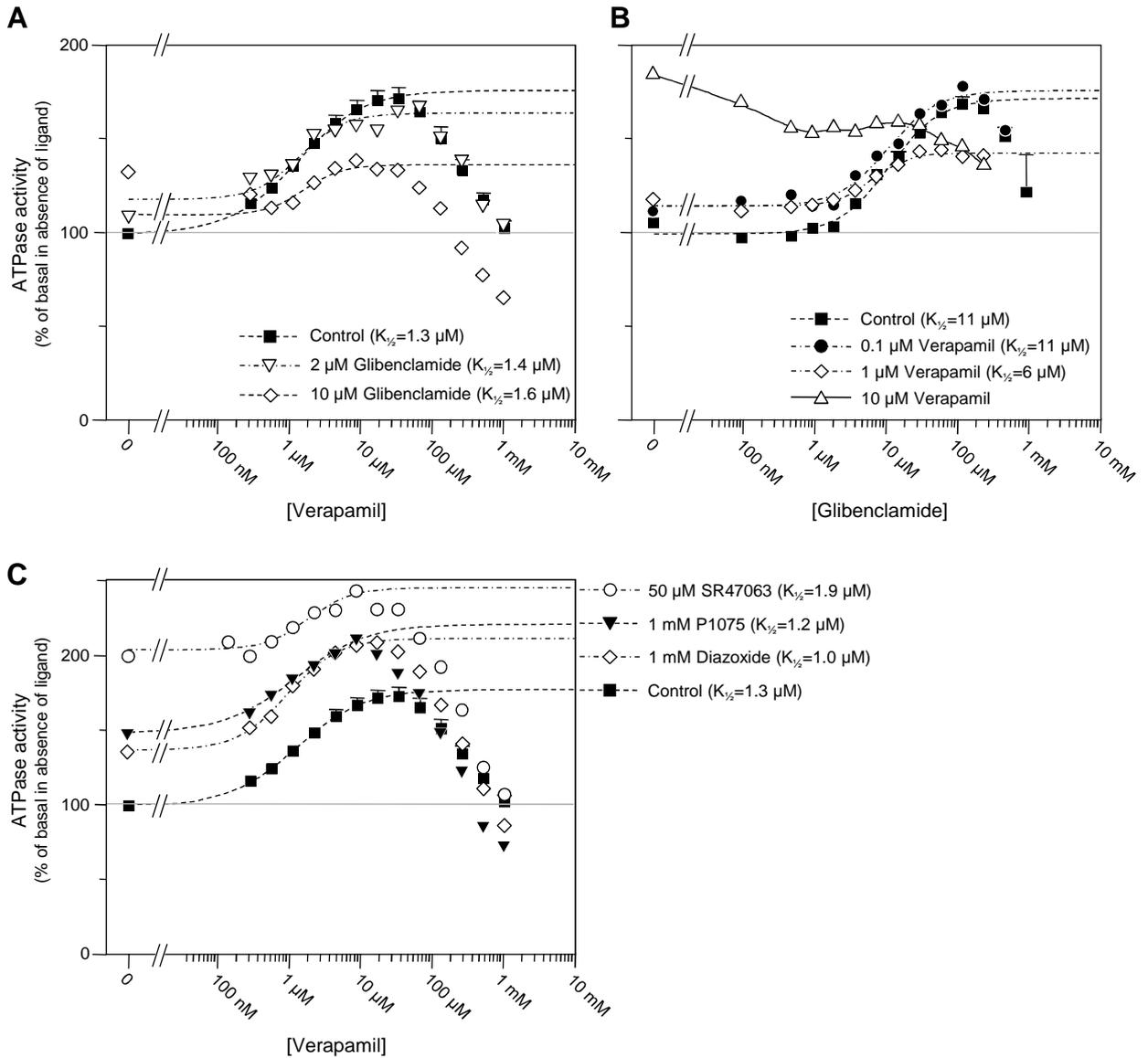


Fig.4

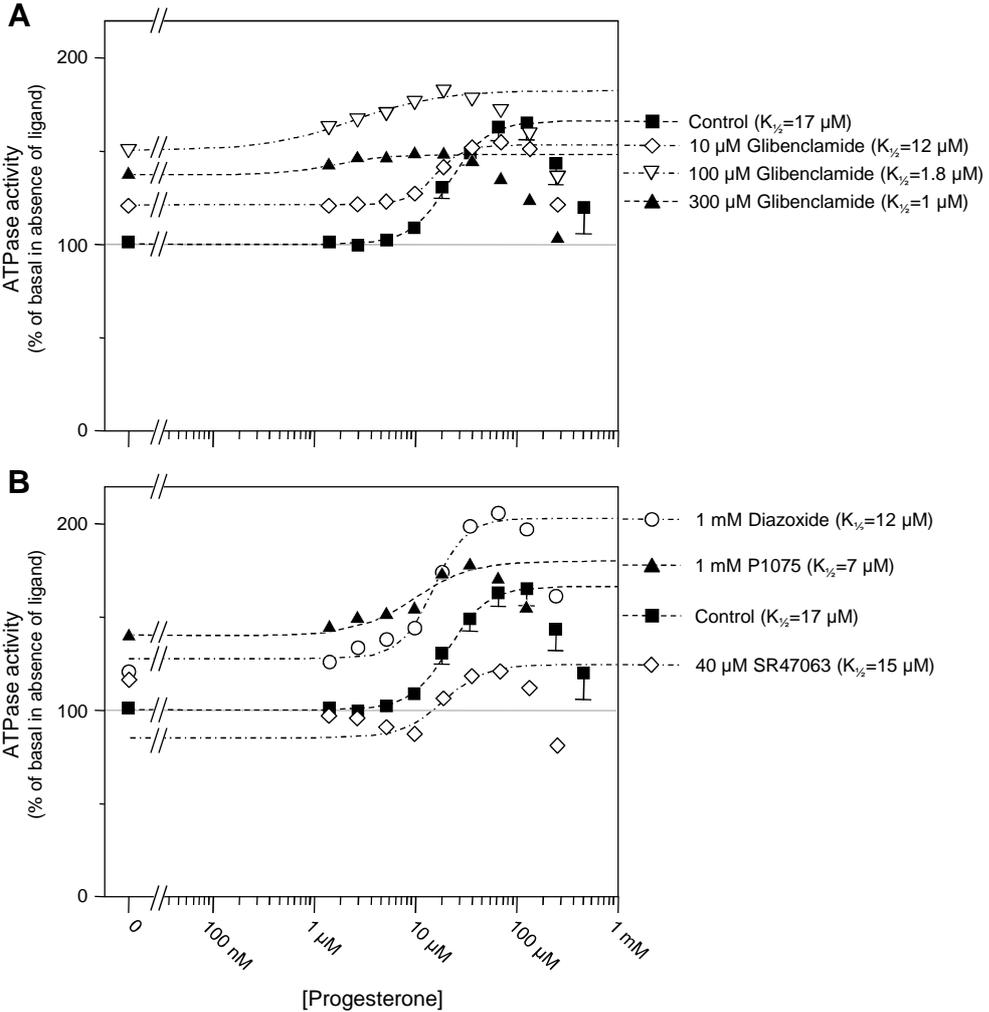


Fig.5

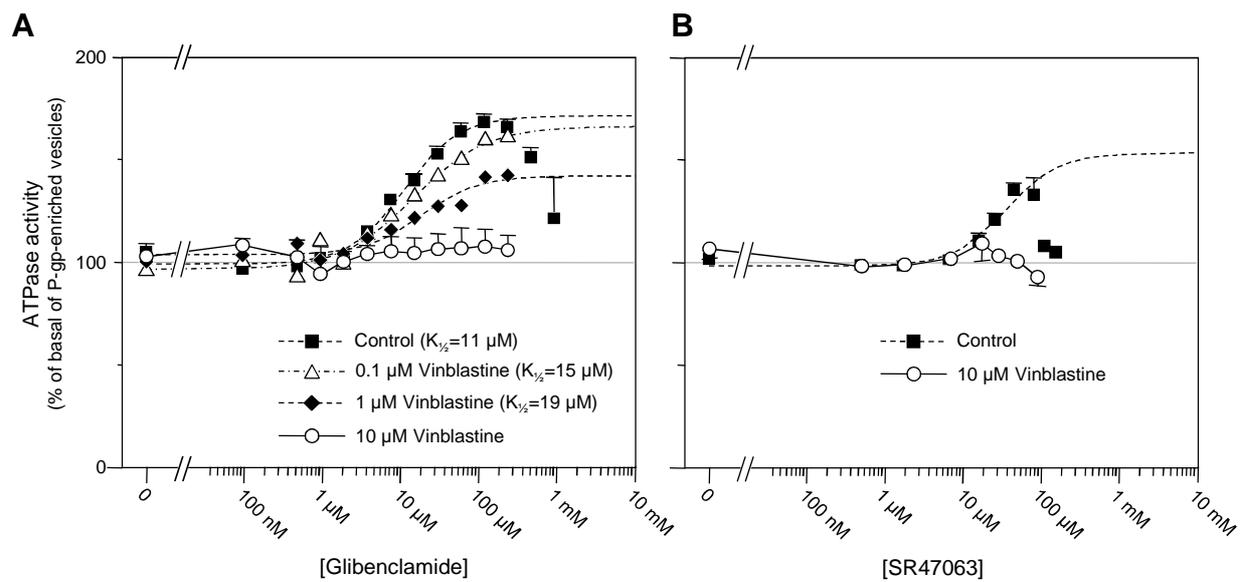


Fig.6

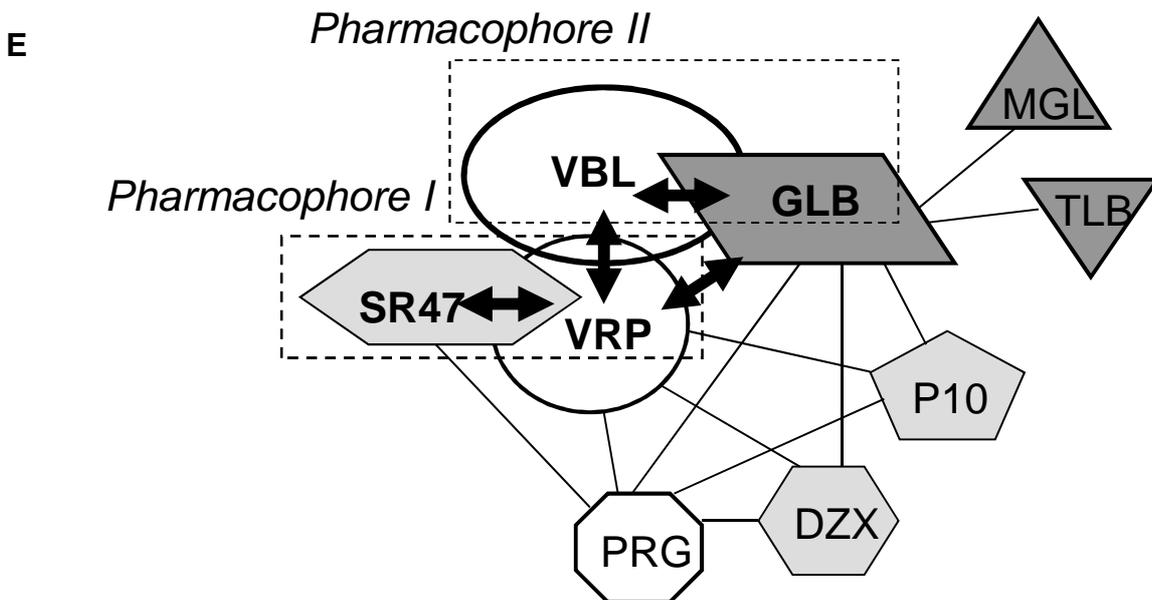
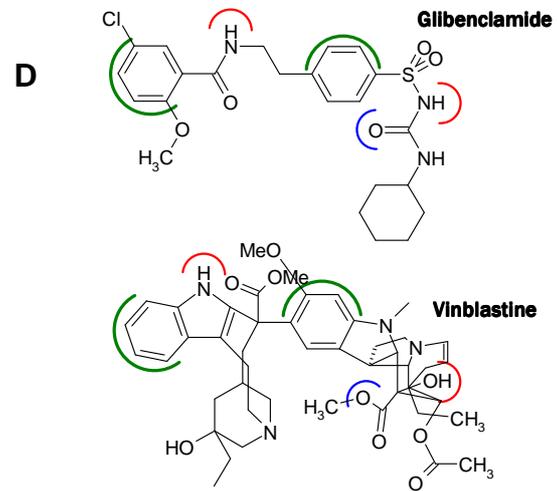
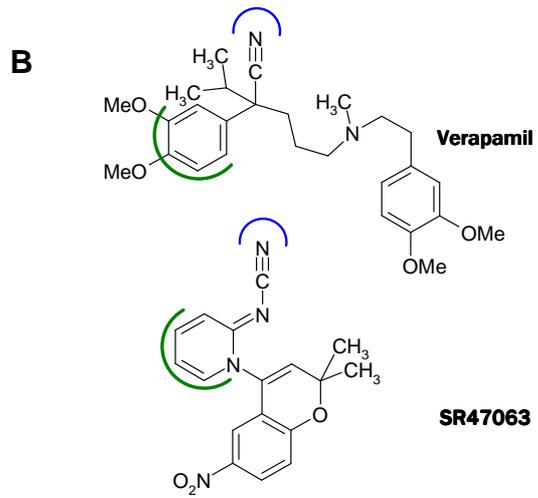
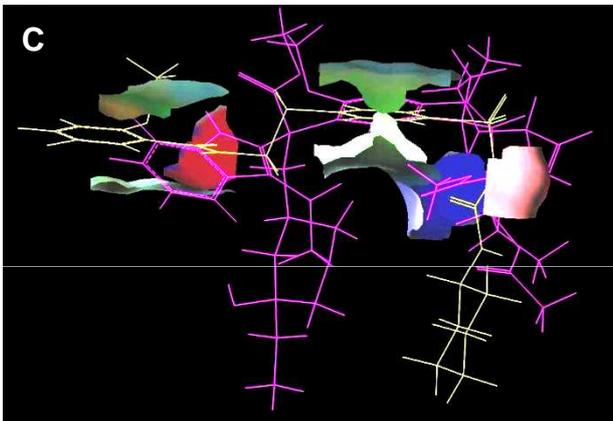
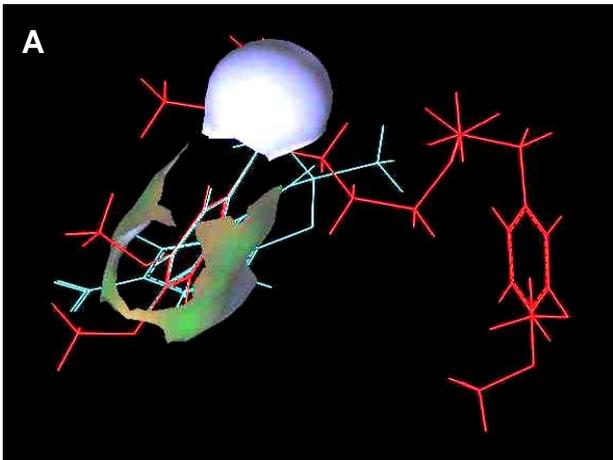


Fig.7

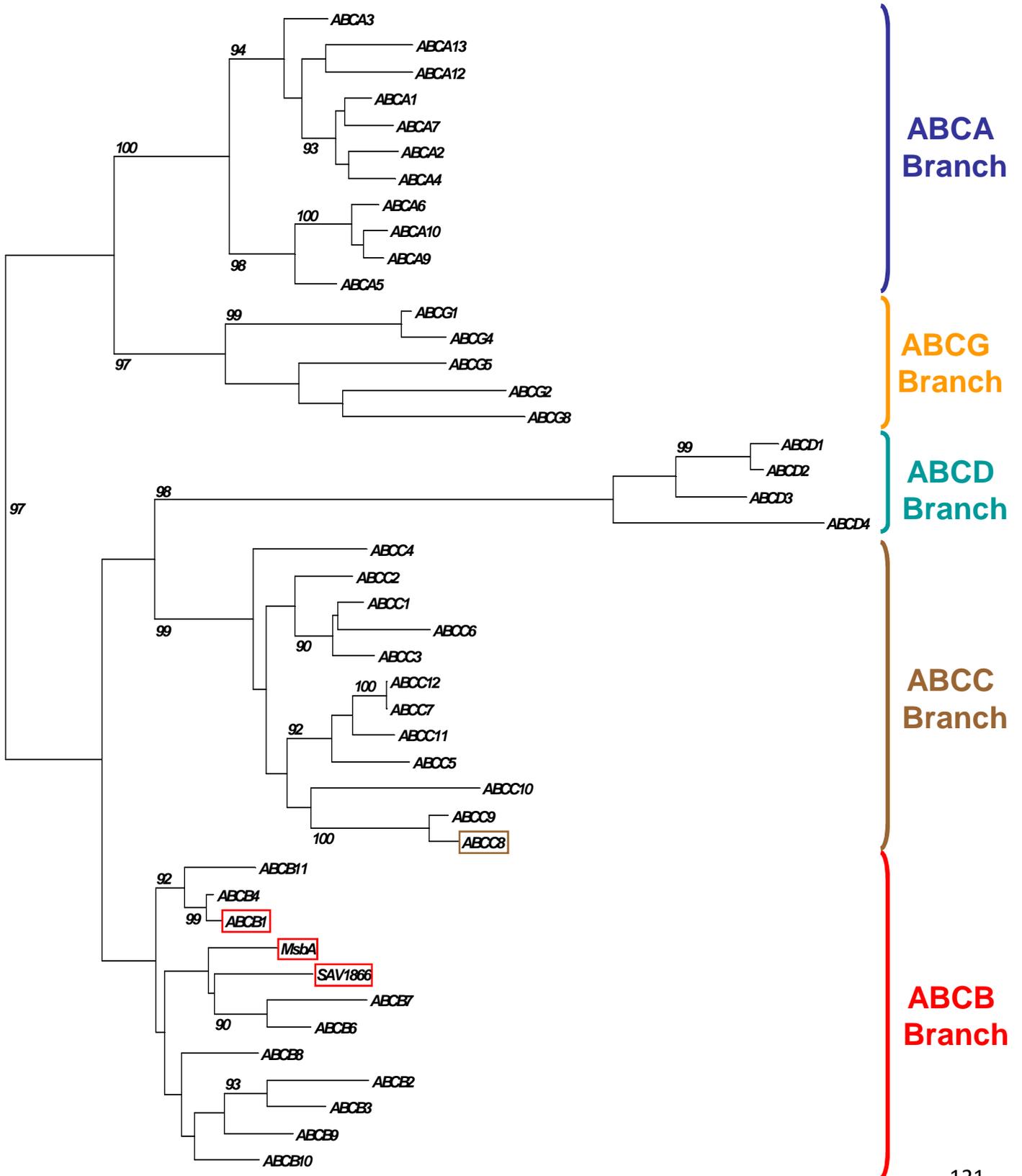


Fig.8

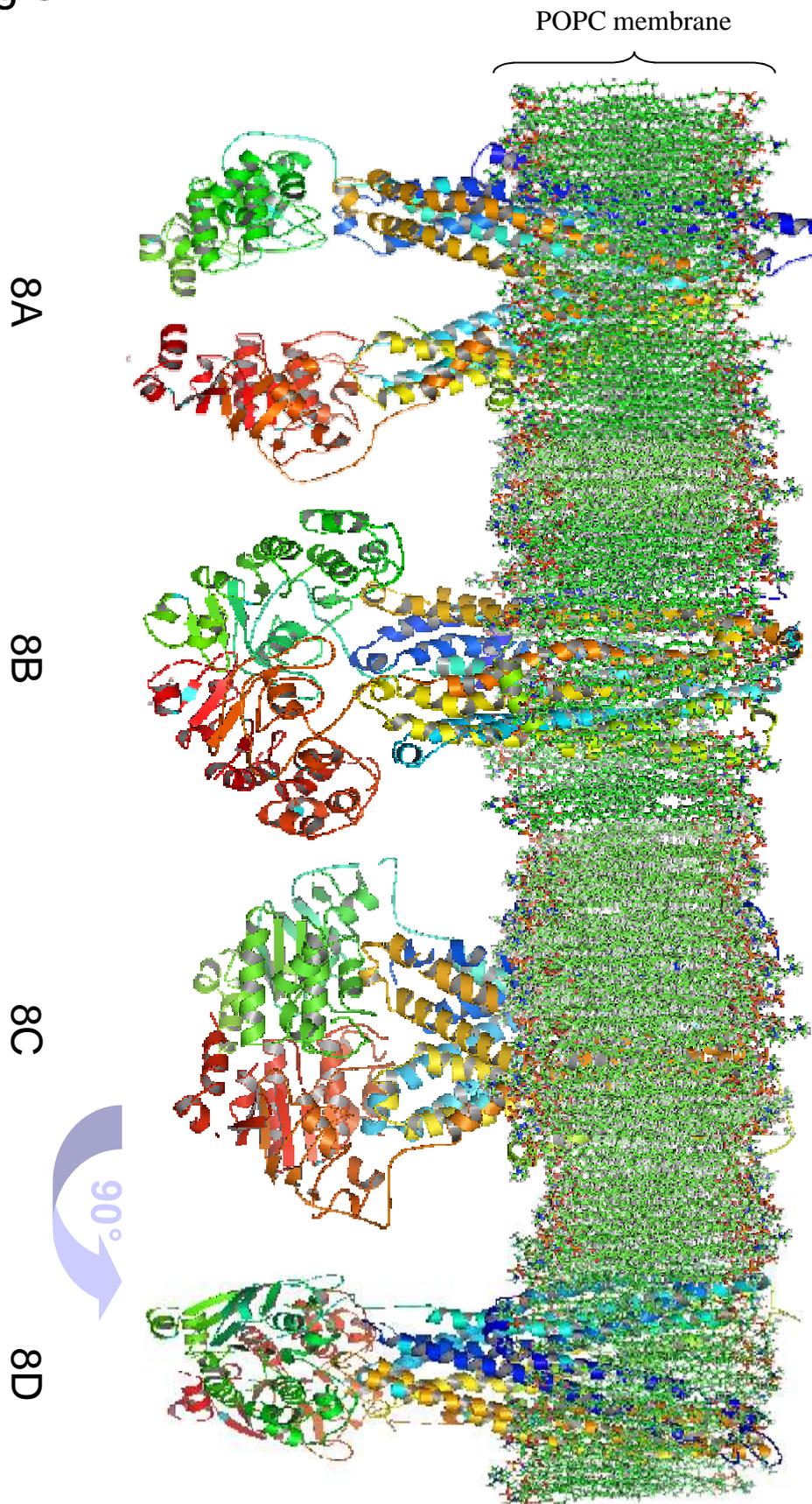


Fig. 9

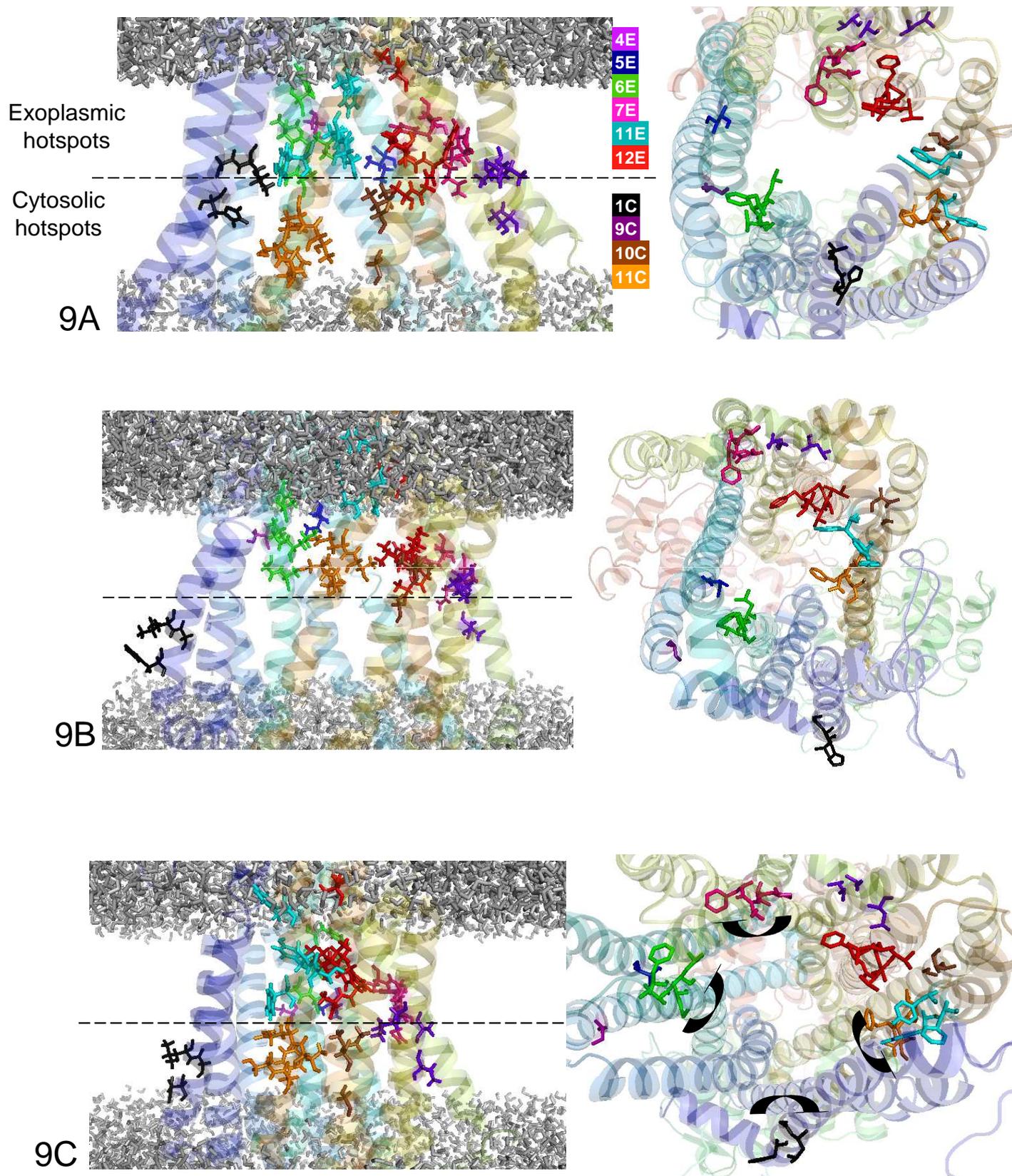


Fig.10

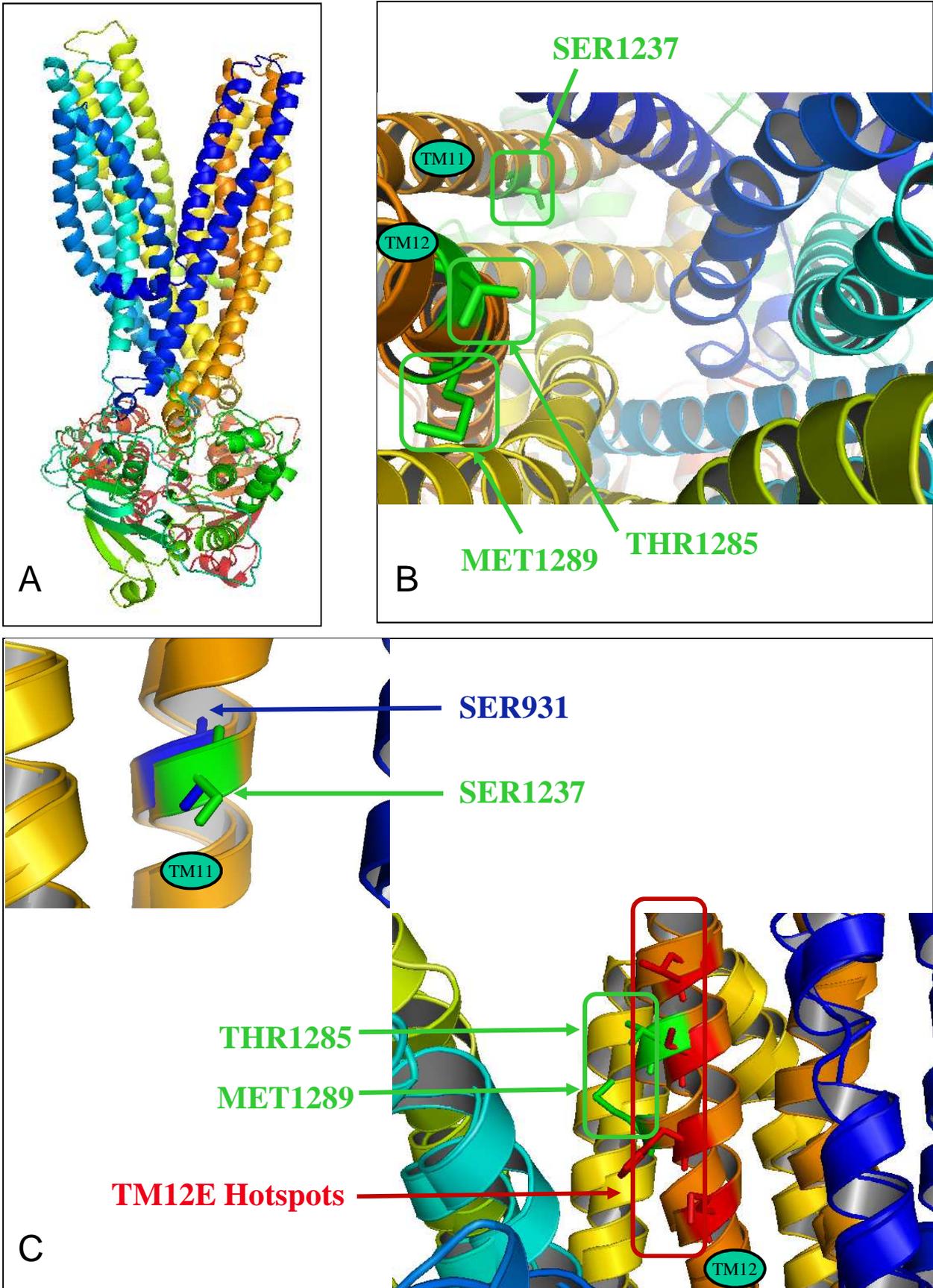


Fig.S1

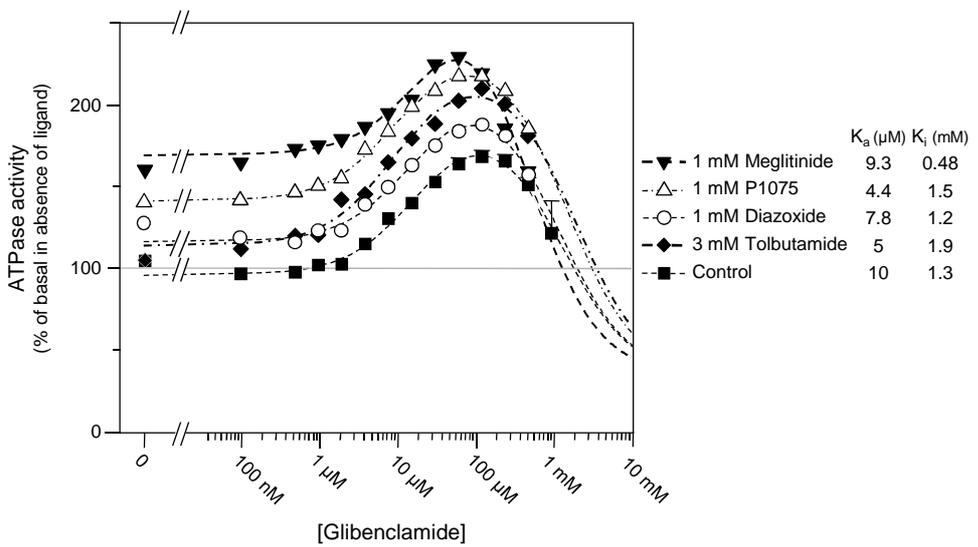
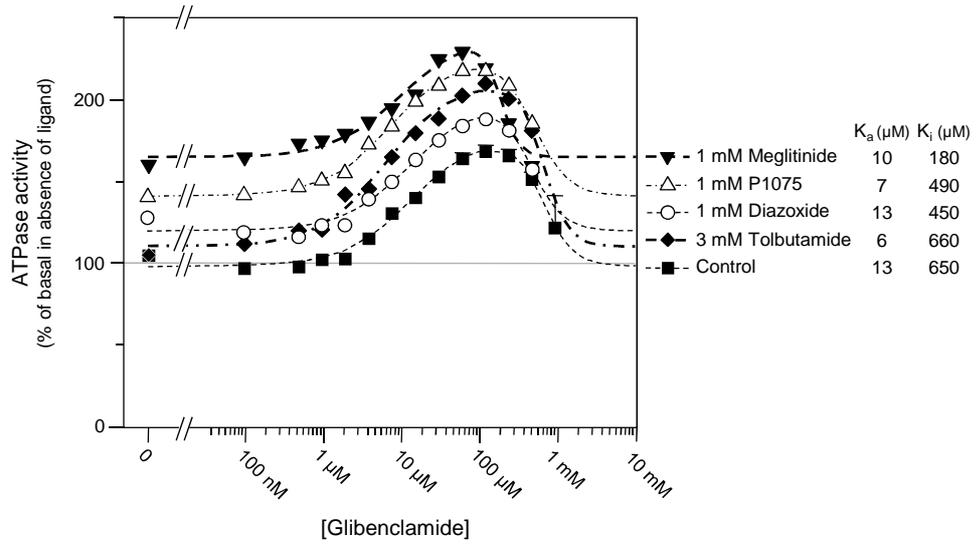
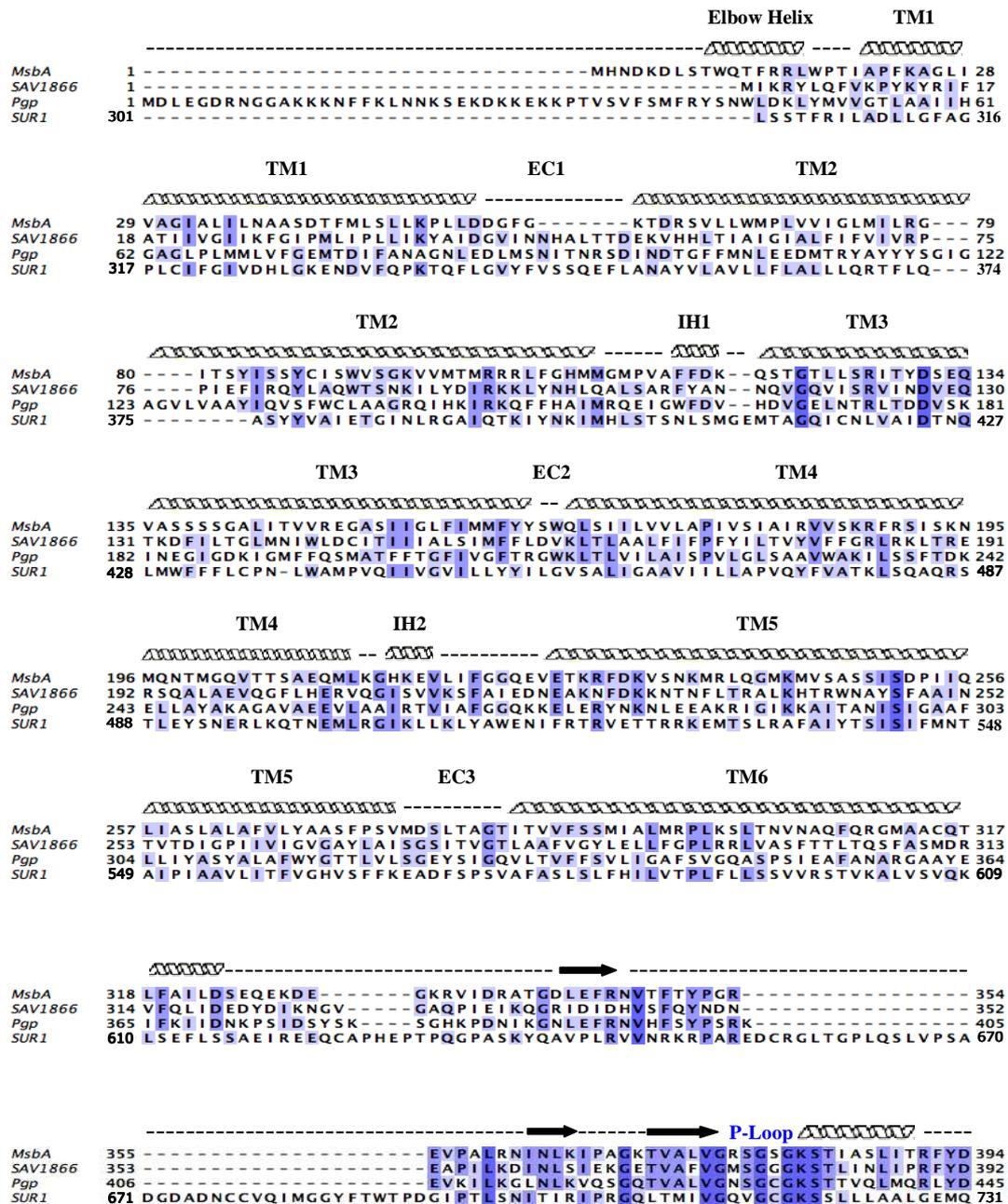


Fig. S2



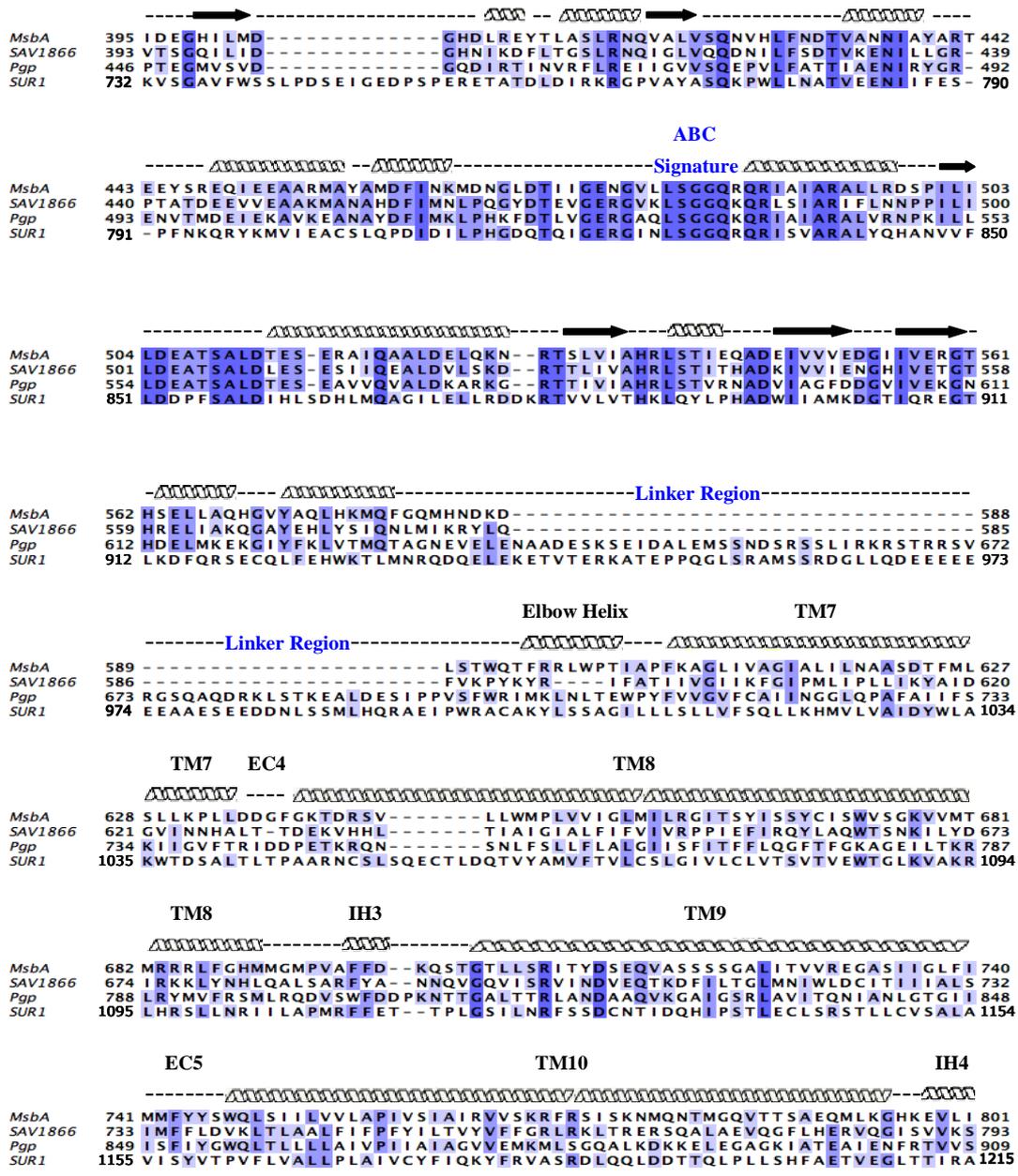
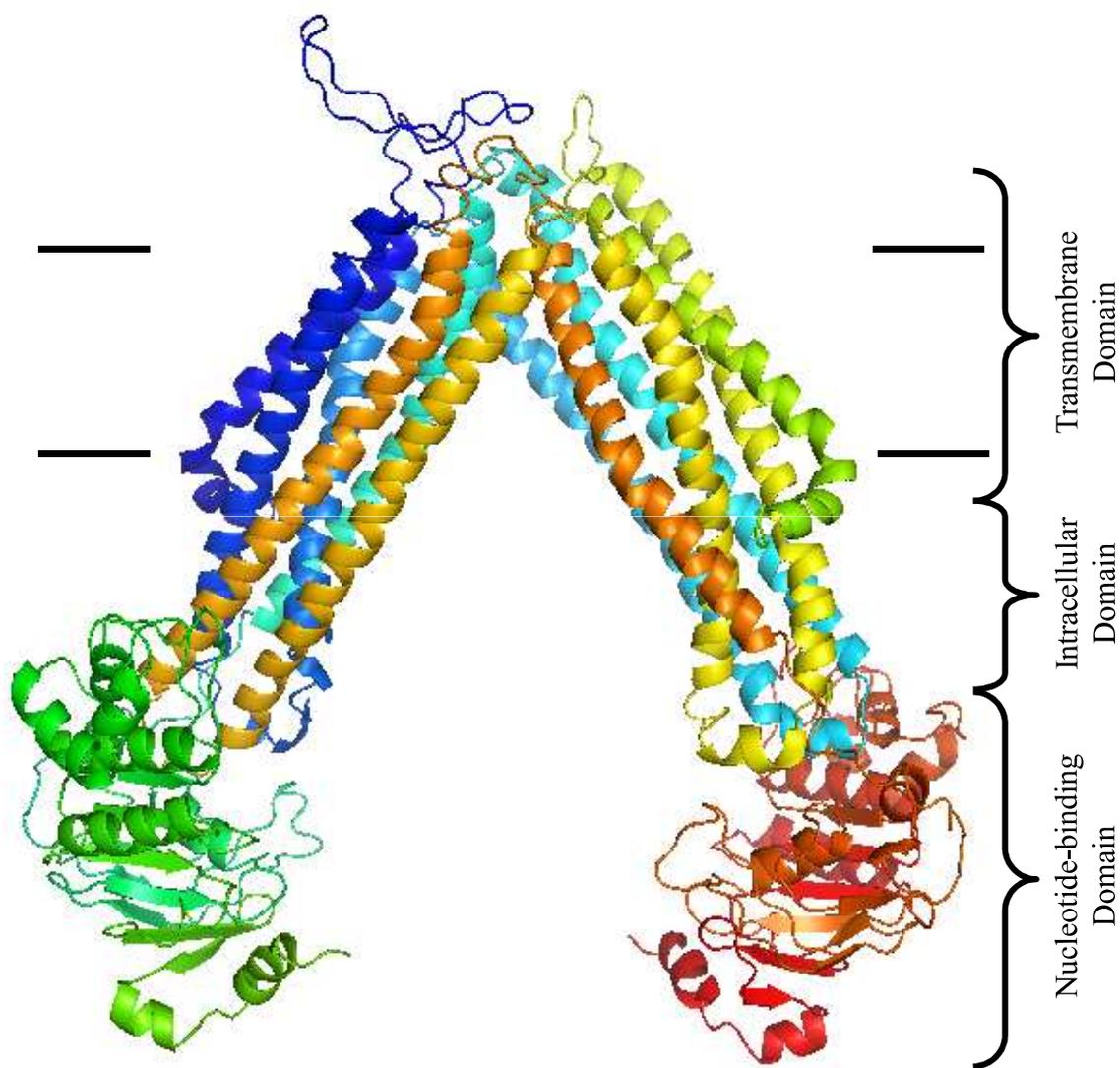


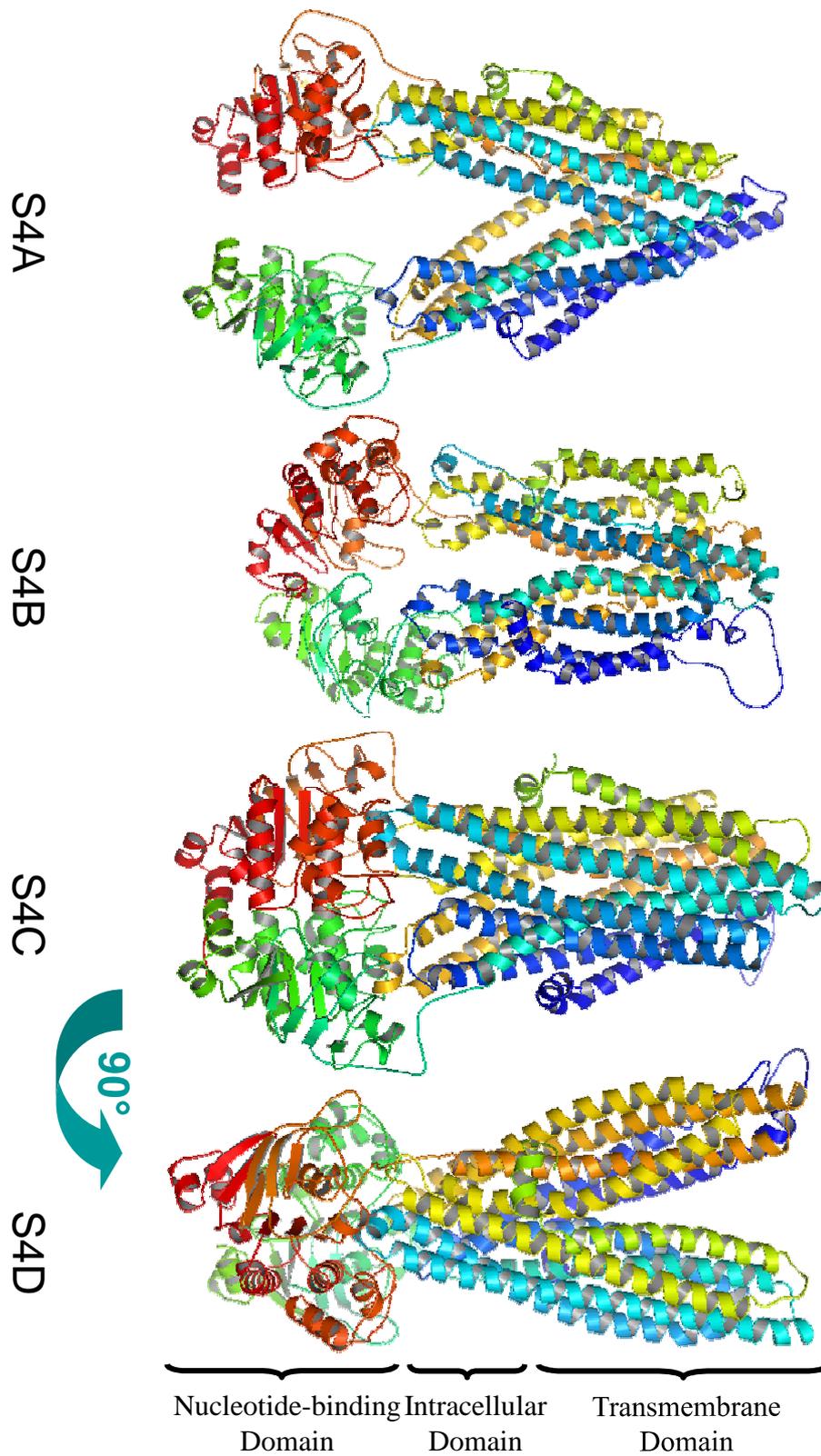


Fig.S3



**Figure S3.** Side view of an open inward-facing conformation model (nucleotide free) of P-gp. This model was rebuilt based on Escherichia Coli MsbA. The three domains composing the P-gp are represented by braces

Fig.S4



**Figure S4.** Side views of P-gp in three structural models: identical to Fig. 8 but POPC molecules are not represented. The three domains composing the P-gp are represented by braces

Table S1: Human ABC-type efflux transporters; with their amino acid sequences selected for the phylogenetics analyses, their Swiss-Prot accession codes, synonyms and TCDB classification numbers.

Sub-family	Member	Proteins (a.a)	Residues (a.a)	Acc. Nb (SwissProt)	Synonyms (Swissprot/TCDB)	TCDB Classification	
ABCA	ABCA1	2261	450-1150	O95477	ABC-1, Cholesterol efflux regulatory protein, Cholesterol/phospholipid flippase	3.A.1.211.1	
	ABCA2	2435	450-1150	Q9BZC7	ABC2, Multidrug resistance pump	3.A.1.211.3	
	ABCA3	1704	200-800	Q99758	ABC-C	3.A.1.211.1	
	ABCA4	2273	500-1200	P78363	RIM ABC transporter, RIM protein, RmP, Stargardt disease protein, The retinal-specific ABC transporter	3.A.1.211.2	
	ABCA5	1642	150-750	Q8WWZ7	ATP-binding cassette sub-family A member 5	3.A.1.211.1	
	ABCA6	1617	150-750	Q8N139	ATP-binding cassette sub-family A member 6	3.A.1.211.1	
	ABCA7	2146	450-1050	Q8IZY2	ATP-binding cassette sub-family A member 7 (Macrophage ABC transporter) (Autoantigen SS-N) (ABCA-SSN)	3.A.1.211.1	
	ABCA9	1524	250-850	Q8IUA7	None	3.A.1.211.3	
	ABCA10	1543	100-700	Q8WWZ4	ATP-binding cassette sub-family A member 10	3.A.1.211.1	
	ABCA12	2595	950-1550	Q86UK0	ATP-binding cassette 12	3.A.1.211.1	
	ABCA13	5058	450-1150	Q86UQ4	ATP-binding cassette sub-family A member 13	Not present	
	ABCB	ABCB1	1280	-	P08183	Multidrug resistance protein 1 (MDR1), P-glycoprotein 1, Broad specificity multidrug resistance (MDR) efflux pump	3.A.1.201.1
		ABCB2	808	-	Q03518	APT1, Peptide transporter TAP1, Peptide transporter PSF1, Peptide supply factor 1, PSF-1, MHC heterodimeric peptide exporter	3.A.1.209.1
ABCB3		686	-	Q03519	APT2, Peptide transporter TAP2, Peptide transporter PSF2, Peptide supply factor 2, PSF-2, MHC heterodimeric peptide exporter	3.A.1.209.1	
ABCB4		1286	1-700	P21439	Multidrug resistance protein 3 (MDR3), P-glycoprotein 3, Short chain fatty acid phosphatidylcholine translocase	3.A.1.201.3	
ABCB5		812	-	Q2M3G0	ATP-binding cassette sub-family B member 5		
ABCB6		842	-	Q9NP58	Mitochondrial ABC transporter 3, ABC transporter umat, Iron transporter	3.A.1.210.6	
ABCB7		752	-	O75027	ABC transporter 7 protein (mitochondrial), ABC7 iron transporter, (X-linked sideroblastia anemia protein)	3.A.1.210.4	
ABCB8		735	-	Q9NUT2	Mitochondrial ATP-binding cassette 1, M-ABC1	3.A.1.201.1	
ABCB9		766	-	Q9NP78	TAP-like protein, TAPL, hABCB9	3.A.1.209.1	
ABCB10		738	-	Q9NRK6	Mitochondrial ATP-binding cassette 2, M-ABC2	3.A.1.212.1	
ABCB11		1321	100-700	O95342	Bile salt export pump, BSEP or SPGP	3.A.1.201.2	
ABCC	ABCC1	1531	250-900	P33527	MRP1, Leukotriene C(4) transporter, LTC4 transporter, Drug resistance pump	3.A.1.208.8	
	ABCC2	1545	250-900	Q92887	Canalicular multispecific organic anion transporter 1, MRP2, Canalicular multidrug resistance protein, Hepatic canalicular conjugate exporter	3.A.1.208.2	
	ABCC3	1527	250-900	O15438	Canalicular multispecific organic anion transporter 2, MRP3, Multi-specific organic anion transporter-D, MOAT-D	3.A.1.208.9	
	ABCC4	1325	1-700	O15439	MRP4, Multi-specific organic anion transporter-B, MOAT-B	3.A.1.208.7	
	ABCC5	1437	150-800	O15440	MRP5, Multi-specific organic anion transporter-C, MOAT-C, pABC11, SMRP	3.A.1.208.13	
	ABCC6	1503	250-1000	O95255	MRP6, Anthracycline resistance-associated protein, Multi-specific organic anion transporter-E, MOAT-E, Multidrug (anthracycline) resistance organic anion efflux pump, The pseudoxanthoma elasticum disease protein)	3.A.1.208.10	
	ABCC7	1480	50-700	P13569	Cystic fibrosis transmembrane conductance regulator, CFTR, cAMP-dependent chloride channel	3.A.1.202.1	
	ABCC8	1581	250-1000	Q09428	Sulfonylurea receptor 1, SUR1	3.A.1.208.4	
	ABCC9	1549	250-1000	O60706	Sulfonylurea receptor 2, SUR2	3.A.1.208.4	
	ABCC10	1492	250-900	Q5T3U5	MRP10	3.A.1.208.9	
	ABCC11	1382	100-750	Q96J66	MRP8, Cyclic nucleotide (cAMP and cGMP) efflux pump	3.A.1.208.13	
	ABCC12	1359	1-750	Q96J65	Multidrug resistance-associated protein 9	3.A.1.208.13	
ABCD	ABCD1	745	-	P33897	Adrenoleukodystrophy protein, ALDP, Long chain fatty acid (LCFA) transporter	3.A.1.203.3	
	ABCD2	740	-	Q9UBJ2	Adrenoleukodystrophy-related protein., hALDR, Adrenoleukodystrophy-like 1	3.A.1.203.3	
	ABCD3	659	-	P28288	70 kDa peroxisomal membrane protein, PMP70, Peroxisomal long chain fatty acyl (LCFA) transporter	3.A.1.203.1	
	ABCD4	606	-	O14678	Peroxisomal membrane protein 69, PMP69, Peroxisomal membrane protein 1-like, PXMP1-L, P70R	3.A.1.203.1	
ABCG	ABCG1	678	-	P45844	White protein homolog, ATP-binding cassette transporter 8	3.A.1.204.1	
	ABCG2	655	-	Q9UNQ0	Placenta-specific ABC transporter, Breast cancer resistance protein, Mitoxantrone resistance-associated protein, CDw338 antigen	3.A.1.204.2	
	ABCG4	646	-	Q9H172	WHITE2	3.A.1.204.1	
	ABCG5	651	-	Q9H222	Sterolin-1	3.A.1.204.5	
	ABCG8	673	-	Q9H221	Sterolin-2	3.A.1.204.5	

**Table S2. Assessment of structural models:** P-gp Open Inward-Facing A: a reconstructed P-gp model based on Ec-MsbA structure as a template, P-gp Open Inward-Facing B: a reconstructed P-gp model based on mouse P-gp structure as template. For the others modelled structures, templates are summarized Table 1

Methods	Structural models	P-gp Open Inward-Facing A	P-gp Open Inward-Facing B	Pgp Closed Inward-Facing	Pgp Outward-Facing	SURI Outward-Facing	
ProSA	Z-Score	-9,62	-12,09	-4,21	-11,41	-9,59	
Procheck	Ramachandran (%)	97,8	99,4	94,7	99,5	98,7	
	G-Factors	Dihedral	0	-0,15	0	-0,13	0
		Covalent	-0,62	-0,25	-0,27	-0,73	-0,27
		Overall	-0,03	-0,18	-0,1	-0,04	-0,1
Whatcheck	Structure Z-scores	1st generation packing quality	0,644	0,656	1,109	-0,293	1,23
		2nd generation packing quality	-2,645	-2,148	-2,187	-3,451	-2,341
		Ramachandran Plot appearance	-2,823	-3,827	-2,115	-3,39	-2,182
		chi-1/chi-2 rotamer normality	-1,164	-2,866	-1,999	-1,567	-2,846
		Backbone conformation	0,984	1,18	2,682	-0,719	1,948
	RMS Z-scores	Bond lengths	0,969	0,955	0,962	0,972	0,978
		Bond angles	1,339	1,269	1,291	1,421	1,279
		Omega angle restraints	0,13	0,138	0,126	0,159	0,126
		Side chain planarity	0,575	0,35	0,465	0,407	0,326
		Improper dihedral distribution	1,121	1,224	1,109	1,272	1,147
		Inside/outside distribution	1,145	1,149	1,171	1,209	1,222
MOLPROBITY	Residues with bad bonds (%)	0,00	0,00	0,00	0,00	0,00	
	Residues with bad angles (%)	1,52	0,68	3,25	0,68	1,07	
QMEAN	QMEANscore	0,512	0,607	0,256	0,742	0,525	

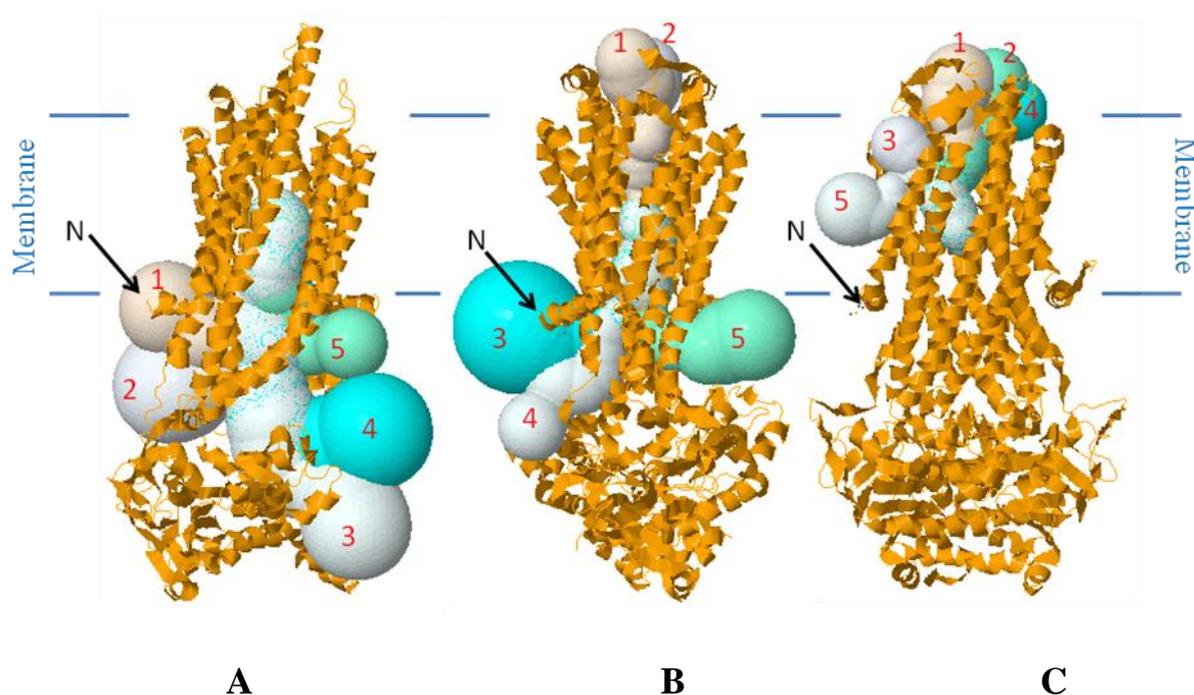
## II.4. Etude du cycle de translocation de la P-gp

Les modèles 3D réalisés de la P-gp et de SUR1 sont présentés et étudiés dans le cadre de l'article inclus dans ce manuscrit. Nous nous intéressons dans cette section à l'analyse du cycle de translocation que nous avons réussi à représenter chez la P-gp humaine en modélisant les trois conformations supposées catalytiques citées précédemment, et pour lesquelles nous recherchons des arguments en faveur de leur implication effective au cours du turn-over enzymatique de la P-gp. Nous pouvons ainsi étudier les canaux d'accès du substrat dans chaque conformation et mettre en relation la transformation des canaux au cours du cycle avec le mécanisme d'expulsion des substrats dans le milieu extracellulaire. Il faut souligner qu'avant de procéder à cette étude, nous avons réalisé une étape d'affinement et d'optimisation supplémentaires (en plus de celle effectuée grâce à Modeller) de nos modèles 3D grâce au programme SYBYL™ (Tripos Inc, St Louis, USA). Ce programme permet par des calculs énergétiques, à partir d'un champ de force défini (nous avons utilisé le champ de force Tripos, plus adapté aux molécules de masses molaires élevées), de déterminer les conformations de plus faibles énergies, qui sont les plus probables. La méthode de minimisation utilisée est celle de *conjugate gradient*. L'optimisation réalisée a été locale et itérative (1000 itérations) ; elle a concerné essentiellement les chaînes latérales dans le but de limiter les « clashes » stériques.

### II.4.1. Analyse des cavités et des canaux d'accès de la P-gp

La surface d'une protéine est d'une forme irrégulière révélant des poches et des cavités éventuellement connectées avec la surface de cette protéine par des tunnels, des canaux ou des pores. Dans le cas de la P-gp, les changements conformationnels imposent un changement drastique de la taille des cavités et de leurs tunnels d'accès. Les substrats peuvent investir les cavités de la protéine à partir de l'environnement extérieur -ou bien à partir du milieu membranaire- selon la forme et la taille de ces tunnels connectant le cœur de protéine avec le solvant. L'anatomie des tunnels change au cours du temps en raison de la dynamique de la protéine et de la membrane.

Dans le cadre de notre analyse des canaux d'entrée et de sortie de la P-gp, nous avons utilisé la version en ligne du programme Mole<sup>18</sup> [19]. Ce programme sert à localiser de manière automatisée les canaux moléculaires dans les structures protéiques. Mole nous a ainsi permis de révéler, pour chaque conformation, 5 canaux majeurs permettant l'accès au cœur de la région membranaire au niveau de laquelle seraient localisés les sous-sites de liaison des substrats de la P-gp. Ces tunnels d'accès sont présentés Figure-C-7.



**Figure-C-7 :** Représentation schématique des tunnels d'accès de la P-gp dans ses trois conformations catalytiques. A, Conformation *Open Inward-Facing*, B, Conformation *Closed Inward-Facing*, C, Conformation *Outward-Facing*. La P-gp est représentée en « cartoon » orange pour les trois conformations. Chaque modèle a été représenté selon l'orientation révélant au mieux les tunnels d'accès, le côté N-terminal est distingué par la lettre N, le milieu membranaire est représenté en bleu. Les tunnels d'accès sont numérotés en rouge, dans l'ordre de leur détection par Mole.

Nous pouvons constater les différences de trajectoires entre les tunnels d'accès révélés au niveau des trois conformations. Ainsi, pour la conformation *Open Inward-Facing*, ces tunnels relient majoritairement le cœur du milieu membranaire de la P-gp avec le milieu

<sup>18</sup> <http://mole.chemi.muni.cz>: Ce programme demande à l'utilisateur un point de départ des tunnels, ainsi qu'un nombre maximal de canaux, les résultats sont donc dépendants de ces paramètres. Pour ce calcul, nous avons choisi n = 5 tunnels maxi, et comme point de départ d'exploration un point au cœur de la membrane au niveau de chaque conformation.

intracellulaire en distinguant néanmoins quelques accès latéraux se positionnant au niveau du feuillet intracytoplasmique de la bicouche lipidique. Ces accès sont également observés au niveau de la conformation *Closed Inward-Facing* mais des accès du côté extracellulaire sont également observés. Nous nous focaliserons dans cette conformation sur le point commun entre ces tunnels qui est le cœur de la membrane. En effet, la conformation *Closed Inward-Facing* est imaginée comme la conformation pour laquelle la protéine a la plus grande affinité pour le substrat. Les accès révélés vers le milieu extracellulaire montrent que le changement conformationnel a bien eu lieu et a été suivi par une ouverture (un tunnel de sortie) non observée quand la P-gp est dans la conformation *Open Inward-Facing*. Enfin, en analysant les tunnels d'accès révélés au niveau de la conformation *Outward-Facing*, nous pouvons évoquer uniquement des « chemins » de sortie pour les substrats. Dans cette configuration, les canaux sont quasi unidirectionnels vers le milieu extracellulaire, l'accès au milieu intracellulaire n'étant plus possible.

Sachant que chaque chemin d'accès est décrit par un certain nombre de résidus, nous avons essayé, pour chaque conformation, d'analyser si des résidus faisant partie des hotspots (voir Tableau 3 de [1]) participaient à la description des cinq canaux identifiés :

- Dans la conformation *Open Inward-Facing*, 4 résidus hotspots font partie des acides aminés délimitant les cinq tunnels révélés : au niveau de TM6E (Phe 343), de l'hélice TM11C (Phe 942 et Gln 946) et de TM12E (Ala 985).
- Dans la conformation *Closed Inward-Facing*, 11 résidus hotspots sont détectés : au niveau de TM6E (Phe 336, Leu 339 et Phe 343), de TM11C (Phe 938 et Phe 942), de TM11E (Tyr 950 et Phe 957) et de TM12E (Leu 975 et Val 982, Met 986 et Gln 990).
- Dans la conformation *Outward-Facing*, 9 résidus hotspots sont détectés : au niveau de TM6E (Ile 340 et Phe 343) et de TM12E (Leu 975, Ser 979, Val 982, Phe 983, Met 986, Gln 990 et Ser 993).

Nous pouvons remarquer qu'il y a beaucoup de résidus hydrophobes dans ces canaux/hotspots, majoritairement des grandes chaînes latérales (Phe, Leu) dans les canaux de la forme A et B (*Inward-Facing*, Figure-C-7). Par contre, dans la forme « expulsive » (*Outward*), on trouve davantage de résidus polaires avec les résidus hydrophobes (Ser, Met, Gln) ce qui participe à la perte d'affinité pour les substrats hydrophobes lors de leur relargage.

La plupart des résidus hotspots identifiés dans les canaux appartiennent au même groupe d'hélices TM : TM6, TM11 et TM12. Le plus grand nombre de résidus appartenant à

des hotspots est détecté dans la conformation *Closed Inward-Facing*, sorte de « conformation hybride » entre les deux extrêmes, pendant laquelle la P-gp aurait la plus grande affinité envers ses substrats.

On remarque également au cours du cycle un glissement des « hotspots » appartenant aux canaux. De la conformation *Inward-Facing* vers la conformation *Outward-Facing*, les hotspots sont d'abord majoritairement de la classe cytoplasmique, puis majoritairement de la classe exoplasmique. Ainsi, dans la conformation *Open Inward-Facing*, les résidus hotspots appartiennent aux deux classes cytosolique (TM11) et exoplasmique (TM6 et TM12). Dans la conformation *Closed Inward-Facing*, en plus de résidus cytosoliques représentatifs de l'hélice TM11, des résidus exoplasmiques de la même hélice sont détectés. Dans la conformation *Outward-Facing*, les résidus hotspots de l'hélice transmembranaire TM11 ne sont plus détectés mais essentiellement des résidus de l'hélice TM12, qui sont tous de la classe exoplasmique. Ce résultat est intéressant, car il est compatible avec l'idée que dans cette conformation, la P-gp est dans la phase d'expulsion de son substrat.

Alors que la localisation membranaire plutôt constante des résidus hotspots, révélée dans Bessadok *et al.*, semble favoriser le modèle de transport : « floppase », notre étude des canaux d'accès met en évidence la possibilité que l'autre modèle de transport supposé « aspirateur hydrophobe ou Vacuum cleaner » puisse aussi être utilisé par la P-gp. Ainsi, nous remarquons qu'au niveau de la conformation *Closed Inward-Facing* supposée représenter un état intermédiaire, le nombre de résidus hotspots exoplasmiques est nettement supérieur à celui des résidus hotspots cytosoliques. L'ouverture vers le feuillet exoplasmique de la membrane semble être favorisée.

#### II.4.2. Comparaison de deux conformations

Dans le but d'évaluer les changements conformationnels au cours du cycle de translocation supposé pour la P-gp, et leurs éventuelles répercussions fonctionnelles, nous avons procédé à une comparaison entre les structures à des niveaux caractéristiques : les domaines TMD, NBD ainsi que les hélices transmembranaires liées aux hélices intracellulaires. Nous avons opté pour la méthode de comparaison la plus utilisée, basée sur les mesures de distance entre deux structures: le RMSD (*Root Mean Square Deviation*). Ce paramètre mesure la distance entre deux structures superposées par la racine carrée de la

moyenne des carrés des distances entre les atomes mis en correspondance dans les deux structures comparées - décrits par leurs coordonnées cartésiennes- :

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2}$$

Où  $\delta$  est la distance entre un nombre d'atomes équivalent dans les deux structures. Cette mesure de distance est calculée après l'alignement structural d'une structure sur une autre de sorte à ce que le RMSD soit minimal.

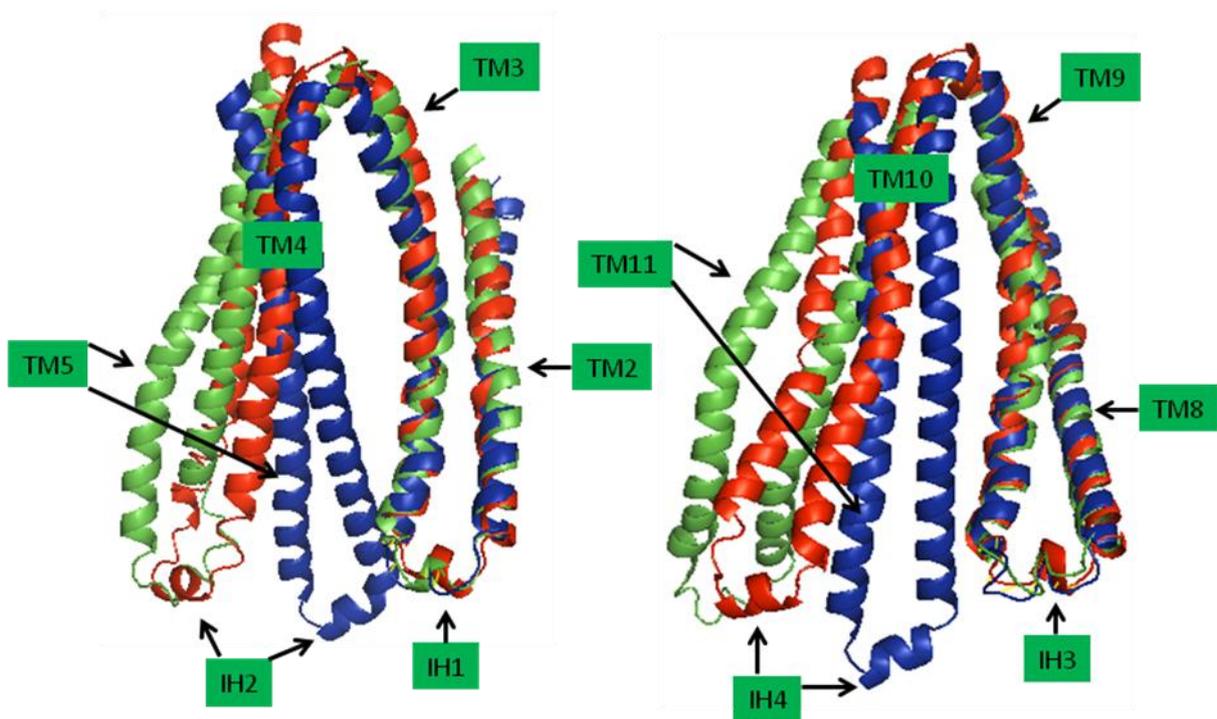
Nous avons réalisé une première comparaison de structures qui a concerné les domaines TMD et NBD des trois structures modélisées de la P-gp. Pour cette comparaison, nous avons utilisé le programme VMD [20], qui a aussi servi à réaliser les alignements structuraux entre les différents domaines. Les valeurs de RMSD obtenues sont résumées dans le Tableau-C-4.

RMSD (Å)		Pgp <i>Open Inward-Facing</i>				Pgp <i>Closed Inward-Facing</i>			
		TMD1	TMD2	NBD1	NBD2	TMD1	TMD2	NBD1	NBD2
<b>Pgp Closed Inward- Facing</b>	<b>TMD1</b>	6,48				x			
	<b>TMD2</b>		5,78				x		
	<b>NBD1</b>			2,73				x	
	<b>NBD2</b>				2,86				x
<b>Pgp Outward- Facing</b>	<b>TMD1</b>	9,89				7,22			
	<b>TMD2</b>		8,21				6,47		
	<b>NBD1</b>			1,75				2,54	
	<b>NBD2</b>				1,89				2,58

**Tableau-C-4 : Les distances RMSD entre les domaines transmembranaires et nucléotidiques de la P-gp dans trois conformations catalytiques différentes.** Les régions comparées sont composées des acides aminés compris entre : 51-357 pour TMD1, 711-1000 pour TMD2, 392-628 pour NBD1, 1035-1273 pour NBD2.

Nous pouvons constater que les distances moyennes sont nettement plus importantes au niveau des domaines transmembranaires par rapport aux domaines nucléotidiques. Ceci traduit un mouvement plus important affectant les hélices transmembranaires, ce qui correspond bien au fait que ces hélices sont au cœur de l'activité de translocation de substrats très variés, par opposition aux domaines NBD qui sont globulaires, ne sont pas déformés au

cours de la catalyse et reconnaissent un seul substrat, l'ATP. Il est aussi intéressant de constater que les valeurs de RMSD entre deux conformations sont sensiblement proches entre les deux TMD comme entre les deux NBD. En comparant les structures 2 à 2, nous pouvons remarquer que les degrés de similarités entre les TMD des conformations *Open Inward-Facing* et *Outward-Facing* sont moins importants qu'entre les TMDs des conformations *Open Inward-Facing* et *Closed Inward-Facing*. Cette constatation supporte l'hypothèse d'un important mouvement des hélices transmembranaires entre la conformation dédiée à la liaison de substrat (*Open Inward-Facing*) et celle dédiée à son expulsion en dehors de la cellule (*Outward-Facing*).



**Figure-C-8 : Comparaison des sous-unités de la P-gp.** Vues de face présentant les changements conformationnels de la P-gp et révélant la mobilité des segments TM4-IH2-TM5 et TM10-IH4-TM11. Pour la clarté de la figure, les domaines NBD ont été enlevés. Les sous-unités ont été alignées au niveau des hélices TM2-TM3 et TM8-TM9. Les conformations *Open Inward-Facing*, *Closed Inward-Facing* et *Outward-Facing* sont respectivement représentées en vert, rouge et bleu.

Dans le but d'infirmer ou de confirmer nos hypothèses concernant ce mouvement des hélices transmembranaires, nous avons procédé à des mesures de distance entre les régions : TM2-IH1-TM3 (120-209), TM4-IH2-TM5 (216-317), TM8-IH3-TM9 (757-853), TM10-IH4-TM11 (854-957) (Tableau-C-5), le mouvement le plus important observé entre les trois

conformations et celui de la fermeture de la « pince » représentée par les hélices précédemment citées (Figure-C-8).

On constate qu'un mouvement significatif est observé au niveau des segments TM4-IH2-TM5 et TM10-IH4-TM11 entre les deux conformations supposées de début et de fin du cycle de translocation (*Open Inward-Facing* et *Outward-Facing*). Les valeurs RMSD de la comparaison des segments décrits Figure-C-8 sont résumées dans le Tableau-C-5.

RMSD (Å)		<b>Pgp Open Inward-Facing</b>				<b>Pgp Closed Inward-Facing</b>			
		TM2-IH1-TM3	TM4-IH2-TM5	TM8-IH3-TM9	TM10-IH4-TM11	TM2-IH1-TM3	TM4-IH2-TM5	TM8-IH3-TM9	TM10-IH4-TM11
<b>Pgp Closed Inward-Facing</b>	TM2-IH1-TM3	2,42				x			
	TM4-IH2-TM5		5,58				x		
	TM8-IH3-TM9			2,27				x	
	TM10-IH4-TM11				5,22				x
<b>Pgp Outward-Facing</b>	TM2-IH1-TM3	2,25				3,17			
	TM4-IH2-TM5		4,37				2,97		
	TM8-IH3-TM9			3,06				2,97	
	TM10-IH4-TM11				3,22				2,19

**Tableau-C-5 : Les distances RMSD entre des segments sélectionnés de la P-gp dans ses trois conformations catalytiques.**

Les résultats obtenus sont en accord avec les constatations et les hypothèses émises quant à un mouvement spécifique des hélices transmembranaires pendant le cycle de

translocation. Nous pouvons remarquer que la dissimilarité la plus importante concerne le passage de la conformation *Open Inward-Facing* aux autres conformations, et plus spécialement au niveau du segment TM4-IH2-TM5 et de son équivalent coté C-terminal TM10-IH4-TM11. Pour les segments TM2-IH1-TM3 et TM8-IH3-TM9, on constate que les valeurs de RMSD sont moins importantes.

Ces observations sont en accord avec celles de Ward *et al.* [8] imaginant un mouvement plus important du segment TM4-IH2-TM5 (MsbA étant un monomère) avec comme charnière les boucles extracellulaires EL2-EL3. Ces observations sont également en accord avec l'hypothèse d'un mouvement de ce segment qui forme un ensemble rigide avec le domaine NBD qui lui est correspondant (NBD2). Ce mouvement serait assuré grâce à l'ancrage de l'hélice intracellulaire, en occurrence IH2, au niveau du domaine nucléotidique. Des mesures de distances réalisées au niveau des hélices intracellulaires (IH1-IH2-IH3-IH4) révèlent des valeurs de RMSD  $< 1\text{\AA}$ , confirmant l'idée que les changements conformationnels concerneraient essentiellement les hélices transmembranaires et que les hélices intracellulaires ancrées dans les domaines NBD (IH2 et IH4) ne seraient pratiquement pas affectées.

### III. Conclusion

Dans le cas de la P-gp, notre travail de modélisation comparative a permis l'obtention de trois modèles 3D représentant cette protéine d'intérêt dans trois conformations différentes : *Open Inward-Facing*, *Closed Inward-Facing* et *Outward-Facing*. Ces conformations sont proposées pour décrire le cycle de transport de la P-gp. Le placement des modèles dans un environnement membranaire a permis de révéler une localisation constante des résidus hotspots responsables du changement de résistance pour la P-gp au niveau de la bicouche lipidique.

Deux protéines ABC telles que la P-gp et SUR1 peuvent avoir des ligands communs sans nécessairement partager une même fonction. Des régions bien spécifiques des deux protéines semblent représenter des sous-sites de liaison pour des ligands reconnus par les deux protéines ABC.

A ce stade de notre étude, il est difficile de discriminer lequel des deux modèles alternatifs de transport, "floppase" ou "aspirateur hydrophobe", est utilisé par la P-gp. Une coexistence des deux modèles peut être imaginée dépendamment de la nature et de la taille du substrat transloqué.

Nos calculs des distances RMSD mettent en évidence de manière statistique l'importance des mouvements et des changements qu'affectent les domaines transmembranaires pendant le cycle de transport.

Suite à l'obtention de modèles 3D de la P-gp, nous avons entrepris des travaux de docking dont les résultats sont présentés dans le chapitre suivant.

## IV. Bibliographie

1. Bessadok, A., et al., *Recognition of sulfonyleurea receptor (ABCC8/9) ligands by the multidrug resistance transporter P-glycoprotein (ABCB1): functional similarities based on common structural features between two multispecific ABC proteins*. J Biol Chem, 2011. **286**(5): p. 3552-69.
2. Aller, S.G., et al., *Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding*. Science, 2009. **323**(5922): p. 1718-22.
3. Tusnady, G.E. and I. Simon, *The HMMTOP transmembrane topology prediction server*. Bioinformatics, 2001. **17**(9): p. 849-50.
4. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes*. J Mol Biol, 2001. **305**(3): p. 567-80.
5. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices*. J Mol Biol, 1999. **292**(2): p. 195-202.
6. Smith, T.F. and M.S. Waterman, *Identification of common molecular subsequences*. J Mol Biol, 1981. **147**(1): p. 195-7.
7. Dawson, R.J. and K.P. Locher, *Structure of a bacterial multidrug ABC transporter*. Nature, 2006. **443**(7108): p. 180-5.
8. Ward, A., et al., *Flexibility in the ABC transporter MsbA: Alternating access with a twist*. Proc Natl Acad Sci U S A, 2007. **104**(48): p. 19005-10.
9. Lupas, A., M. Van Dyke, and J. Stock, *Predicting coiled coils from protein sequences*. Science, 1991. **252**(5009): p. 1162-4.
10. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
11. Maupetit, J., R. Gautier, and P. Tuffery, *SABBAC: online Structural Alphabet-based protein Backbone reconstruction from Alpha-Carbon trace*. Nucleic Acids Res, 2006. **34**(Web Server issue): p. W147-51.
12. Gautier, R., A.C. Camproux, and P. Tuffery, *SCit: web tools for protein side chain conformation analysis*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W508-11.
13. Sali, A., et al., *Evaluation of comparative protein modeling by MODELLER*. Proteins, 1995. **23**(3): p. 318-26.
14. Sippl, M.J., *Recognition of errors in three-dimensional structures of proteins*. Proteins, 1993. **17**(4): p. 355-62.
15. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR*. J Biomol NMR, 1996. **8**(4): p. 477-86.
16. Hooft, R.W., et al., *Errors in protein structures*. Nature, 1996. **381**(6580): p. 272.
17. Davis, I.W., et al., *MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W615-9.
18. Benkert, P., S.C. Tosatto, and D. Schomburg, *QMEAN: A comprehensive scoring function for model quality assessment*. Proteins, 2008. **71**(1): p. 261-77.
19. Petrek, M., et al., *MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels*. Structure, 2007. **15**(11): p. 1357-63.
20. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

# Docking

# Chapitre D : Docking

---

## I. Introduction

En biologie structurale, on étudie le lien entre la structure des molécules et leur fonction biologique. Ce domaine recouvre des questions relevant, d'une part, de la pharmacologie (conception de médicaments) et, d'autre part, de la biologie cellulaire. Dans le premier cas, un récepteur étant connu, le but est de trouver un ligand, c'est-à-dire une (ou plusieurs) molécule(s) complémentaire(s). Dans le second, l'intérêt porte plutôt sur l'interaction entre macromolécules (protéines ou acides nucléiques) intervenant dans les cycles cellulaires. Une distinction est faite entre les interactions ligand-protéine et les interactions protéine-protéine. Une question centrale commune aux deux problématiques est le « docking<sup>19</sup> » (amarrage moléculaire). Les méthodes de docking s'attachent à prédire les interactions possibles lors de la formation de complexes moléculaires. En particulier, le docking permet la prédiction des meilleures orientations selon lesquelles deux molécules interagissent. Pour ce travail de thèse, nous nous sommes attachés davantage au docking ligand-protéine pour étudier la multispécificité de la P-glycoprotéine. Le docking repose sur l'hypothèse que les ligands formant des interactions favorables avec le récepteur sont ceux présentant une affinité de liaison élevée. Ces interactions entre molécules sont non covalentes, et incluent donc : (i) les interactions Van der Waals (interactions de faible intensité existant entre les atomes et dues à leurs charges électriques), (ii) les interactions électrostatiques, et (iii) les interactions hydrogène.

Dans le cas d'une protéine à intérêt clinique comme la P-gp, l'obtention de modèles 3D décrivant des états catalytiques différents représente une avancée importante dans le but de comprendre son processus de reconnaissance et d'export de substrats. Une étude pharmacologique, menée en 2002 au sein du laboratoire sur la P-gp de hamster [1] avait suggéré l'existence de sous-sites de liaison différents mais très proches pour expliquer le

---

<sup>19</sup> Terme que l'on peut traduire par amarrage moléculaire, ou arrimage moléculaire, mais nous adoptons le terme « docking » par commodité dans la suite de ce manuscrit.

mode de fixation de plusieurs molécules d'intérêt thérapeutique. Pour tester *in-silico* les hypothèses émises par cette étude, nous avons réalisé une étude de docking sur deux structures modélisées de P-gp de Hamster (isoforme sur laquelle ont été obtenues les données *in vitro* en 2002) dans les conformations *Inward* et *Outward*. Dans le but de comparer les processus de reconnaissance et d'export de substrats entre P-gp de différentes espèces animales, nous avons réalisé en parallèle des travaux de docking sur deux modèles 3D de la P-gp humaine, également dans les conformations *Inward* et *Outward*.

## **II. Le processus de docking**

### **II.1. Présentation**

Il existe différents algorithmes de docking. Les critères de choix sont le plus souvent le nombre de molécules à tester et la précision souhaitée, critères qui ont un impact sur le temps de calcul et aux moyens nécessaires. Les méthodes de docking les plus rapides sont les méthodes dites de docking rigide [où les molécules (le récepteur et le ligand) sont structurellement figées], historiquement les premières conçues. Avec l'amélioration des puissances de calcul, les méthodes de docking dites « semi-flexibles » où les partenaires d'interaction sont partiellement déformables (en général ligand flexible, récepteur rigide), se sont imposés, car elles permettent d'obtenir des résultats plus précis, même dans des approches couteuses de criblage virtuel haut débit.

Le docking semi-flexible est constitué de deux étapes principales. La première, dite de docking, permet aux petites molécules testées d'explorer le site d'interaction avec la protéine. La molécule adopte plusieurs conformations possibles au cours de cette recherche, et le programme génère plusieurs positions possibles d'amarrage du ligand dans la protéine et plusieurs positions autour de la protéine. La deuxième étape, dite de *scoring*, permet d'évaluer les poses rencontrées lors de la première étape pour ne retenir que la meilleure. Cette étape donne lieu à un calcul de score.

## II.2. L'étape de docking

Les programmes de docking effectuent une exploration de l'espace des configurations pour générer et évaluer un grand nombre de liaisons potentielles. Il existe trois méthodes d'exploration conformationnelle : systématique, stochastique et déterministe [2]

### II.2.1. La recherche systématique

Ces méthodes exploitent tous les degrés de liberté des angles de torsion du ligand sous certains critères permettant de limiter l'exploration si la conformation du ligand ne remplit pas certaines conditions. Ces algorithmes sont dits aussi de fragmentation / reconstruction. Dans un premier temps, la molécule est découpée en parties rigides et en liaisons flexibles. Les parties rigides sont positionnées en premier et le ligand est reconstruit de proche en proche, explorant diverses conformations lors de la mise en place des liaisons flexibles [3].

Parmi les programmes fonctionnant sur ce principe, figurent DOCK [4], FlexX [5], Surflex [6] et Glide [7-8]. Dans le cas de FlexX, le positionnement du premier fragment est basé non seulement sur la complémentarité de surface, mais tient également compte des liaisons hydrogène et interactions de type hydrophobe qui pourraient être formées. Surflex utilise en plus une sorte de ligand idéal composé de fragments comme NH, C=O et CH<sub>3</sub> placés par le module *protomol* pour sélectionner les poses de fragments qui formeront le plus de liaisons avec la protéine.

### II.2.2. La recherche stochastique

Le ligand est considéré dans son entier. Les mouvements de rotation et de translation et les variations des angles de torsion se font au hasard. Cette recherche aléatoire peut se faire par des approches de type Monte Carlo ou algorithmes génétiques.

L'algorithme de Monte Carlo permet de contrôler cette exploration aléatoire grâce au critère de Métropolis. Une nouvelle conformation d'énergie plus faible que la précédente sera acceptée. LigandFit [9], Prodock [10] et AutoDock [11] utilisent tous l'algorithme de Monte Carlo. AutoDock a été le premier à utiliser cet algorithme, mais offre d'autres approches stochastiques jugées plus performantes (algorithme génétique, algorithme hybride Larmarckian) c'est le programme que nous avons utilisé (dans sa version 4.2) [12] pour

réaliser les travaux de docking présentés dans ce mémoire. LigandFit fait varier les angles de torsion, et retient la molécule si la forme du ligand est compatible avec celle du site de liaison.

Les algorithmes génétiques génèrent une population de conformations et de positions du ligand au hasard, qui évolue selon des événements génétiques (mutations, crossing-over) régis par un taux de probabilité et une pression de sélection fixés. Un taux de survie définit le nombre d'individus passant d'une génération à l'autre. Le nombre de générations maximum est fixé, mais un critère de convergence peut arrêter l'algorithme avant. AutoDock propose ces algorithmes.

### II.2.3. La recherche déterministe

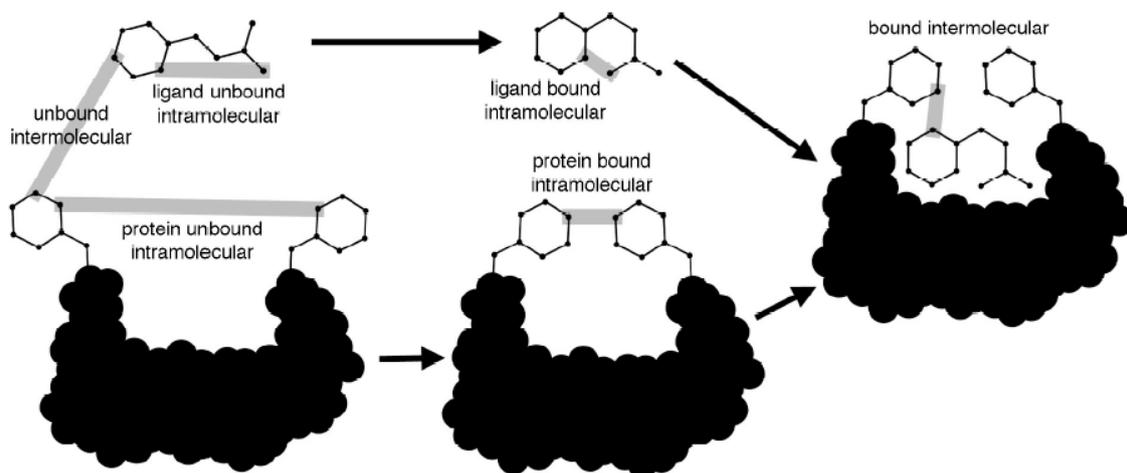
La dynamique moléculaire et la minimisation d'énergie sont utilisées pour trouver des conformations d'énergie la plus basse possible. Le résultat de ces méthodes est fonction de l'état initial. Le problème des minima locaux, qui ne correspondent pas à la plus basse énergie possible, peut être contourné par le franchissement de barrières d'énergie à plus haute température, comme le permet le recuit simulé.

## II.3. L'étape de scoring

Les résultats du docking consistent le plus souvent en différentes poses du ligand dans le site d'interaction ou au-delà du site d'interaction lorsque la zone de recherche comprend d'autres cavités ou des sites de surface. Ces poses correspondant à plusieurs conformations favorables que le ligand peut adopter. Les méthodes de « scoring » permettent d'évaluer l'énergie de liaison du complexe formé et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées.

La fonction de scoring est estimée généralement par l'évaluation des contributions entropiques et enthalpiques à la stabilité du complexe. Dans les programmes de docking, on trouve différents types de fonctions de score. Autodock utilise un champ de force basé sur l'énergie libre calculée de façon semi-empirique. Ce champ de force a été paramétré en utilisant un grand nombre de complexes protéine-inhibiteur pour lesquels les structures et les constantes d'inhibition sont connues. Il évalue la liaison protéine-ligand en deux étapes

(Figure-D-1) : le champ de force estime les énergies intramoléculaires du ligand et de la protéine lors de leur transition de l'état libre à leur état lié, puis estime les énergies intermoléculaires combinant le ligand et la protéine dans l'état lié.



**Figure-D-1 : Les étapes d'évaluation par AutoDock de la liaison protéine-ligand (Source : Guide d'utilisation d'AutoDock).**

L'équation décrivant le champ de force inclut six évaluations d'énergies par paires (V) et une estimation de l'entropie conformationnelle perdue pendant la liaison.

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf})$$

Intramoléculaire ligand	Intramoléculaire protéine	Intermoléculaire protéine-ligand	Entropie conformationnelle
----------------------------	------------------------------	-------------------------------------	-------------------------------

L : Ligand ; P : Protéine

### III. Etude de docking chez les transporteurs ABC multispécifiques

#### III.1. Contexte scientifique

L'étude pharmacologique de 2002 [1] a combiné des données enzymologiques et des approches de modélisation tridimensionnelle, et a permis d'analyser les interactions mutuelles

entre des substrats de la P-gp. Les interactions révélées par cette étude pour être résumées comme suit :

- Le vérapamil (VRP) est compétitif avec des molécules de plus grande taille que lui, comme la vinblastine (VBL), l'actinomycine D, la cyclosporine A, la bromocriptine et la pristinamycine IA, et partage des éléments hydrophiles/hydrophobes consensus avec toutes sauf VBL, mais est non compétitif avec une molécule de petite taille qui est la tentoxine (TTX).
- La VBL est, quant à elle, compétitive avec les molécules de grande taille mais également avec une molécule de petite taille comme la TTX, en partageant avec elle des éléments consensus.

Note : Une inhibition compétitive suggère que les deux substrats agissent au niveau du même site de liaison de la P-gp, un seul substrat pouvant s'y lier à la fois. Une inhibition non compétitive signifie que les deux substrats peuvent se lier simultanément à la P-gp au niveau de sites distincts. Des effets allostériques peuvent également être impliqués dans les interactions entre un substrat et un inhibiteur.

Les données de cette étude ont permis de décrire une carte topologique de la poche de liaison de la P-gp et de conclure à l'existence de deux pharmacophores distincts pour la reconnaissance des substrats transportés par la P-gp. Les molécules appartenant au même site (i.e. s'excluant mutuellement) ont ensuite été superposées en 3D, et alignées à la main de manière à révéler des points pharmacophoriques (de type alkyles, aromatiques, hydrogène donneur... etc). Ainsi, une représentation schématique a pu être établie décrivant ces molécules liées au niveau de deux sites spécifiques avec des distances et angles solides caractéristiques.

La Figure-D-2 fait apparaître l'organisation spatiale des groupements caractéristiques de chacun des sites et la zone de contact entre les deux pharmacophores, qui est postulée pour rendre compte de l'effet taille des ligands se traduisant par une exclusion mutuelle (par exclusion stérique) entre les plus gros.

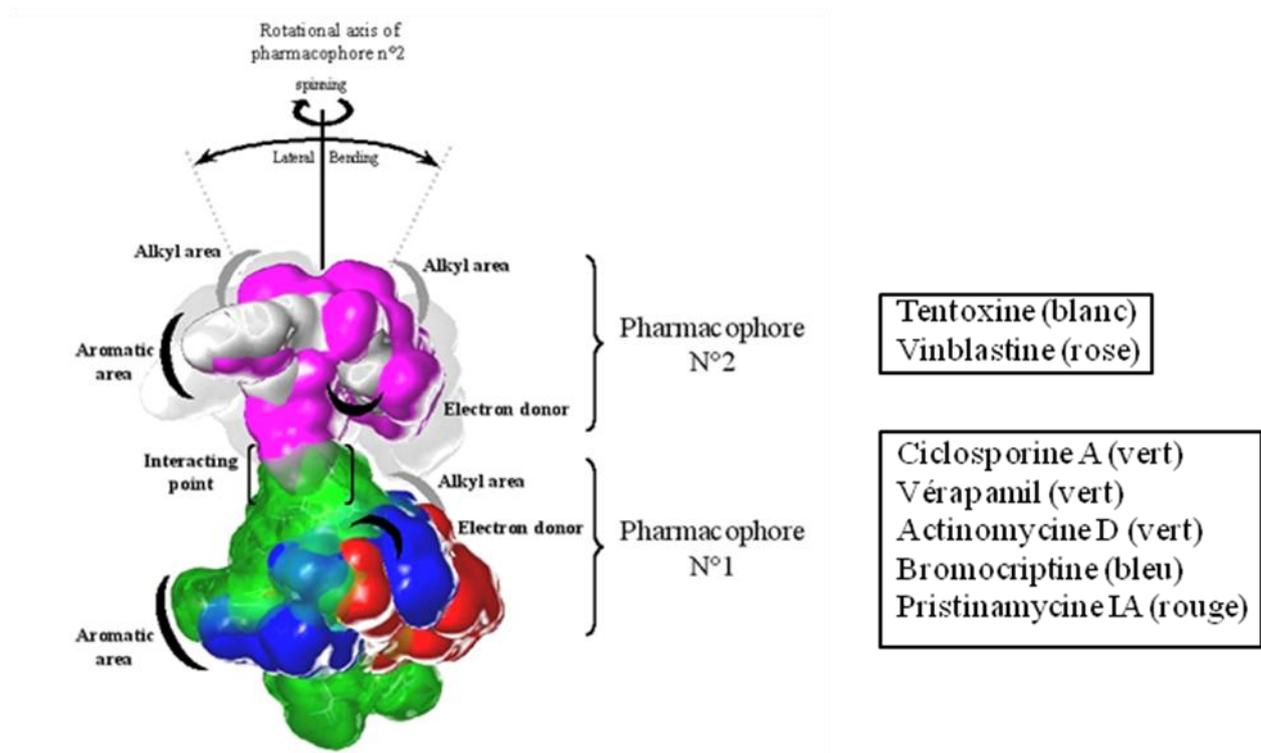


Figure-D-2 : Modèle pharmacophorique de reconnaissance multispécifique de la P-gp. D'après

[1]

### III.2. Les protagonistes considérés

#### III.2.1. Le récepteur

Les modèles 3D disponibles au laboratoire (chapitre C) sont ceux de la P-gp humaine, mais les résultats de l'étude pharmacologique ont été obtenus sur des vésicules enrichies en P-gp de hamster. Nous avons donc dû réaliser un travail supplémentaire de modélisation comparative avant de procéder au docking. Deux modèles 3D de la P-gp de hamster chinois (N° accession : P21448) dont la taille de séquence protéique est de 1276 acides aminés ont été construits. Ces deux modèles représentent la P-gp de hamster dans les conformations *Open Inward-Facing* et *Outward-Facing*. Etant donné que la conformation *Closed Inward-Facing* représente un état catalytique, supposé intermédiaire et que la structure du seul template disponible dans la PDB est d'une résolution assez moyenne (5.5Å), nous avons concentré nos efforts de docking sur les modèles représentant les conformations *Open Inward-Facing* et *Outward-Facing*. Ce choix est justifié aussi parce que ces dernières conformations sont

supposées comme étant les plus informatives sur le cycle catalytique et l'évolution des sites de liaison transmembranaires de la P-gp (Elles sont supposées représenter les états de liaison et d'export des substrats).

Les étapes de modélisation, de sélection et d'affinement ont été réalisées de manière identique à celles qui avaient permis de construire les modèles 3D de la P-gp humaine présentés dans le chapitre précédent. La séquence protéique de la P-gp de hamster présente des identités de séquence de 29 %, 28% et 82% avec les séquences protéiques de St-MsbA, de Sav1866 et de la P-gp de souris, respectivement.

### III.2.2. Les ligands

L'autre partenaire du docking à considérer est le ligand. Ainsi, après avoir défini la protéine récepteur dans deux conformations différentes et distantes, nous avons choisi les ligands à arrimer sur la base des les molécules définissant les pharmacophores décrits dans [1]. Nous avons opté pour le docking de deux molécules représentant le pharmacophore n°2; la tentoxine et la vinblastine, et d'une molécule représentant le pharmacophore n°1, le vérapamil (voir Figure-D-2). Le choix de ces molécules se base sur leur grande représentativité de l'interaction supposée avec les deux pharmacophores et la poche de liaison imaginée pour la P-gp, ainsi que sur l'absence entre elles de toute ressemblance structurale. Ces molécules représentent également un grand intérêt pharmacologique et même thérapeutique :

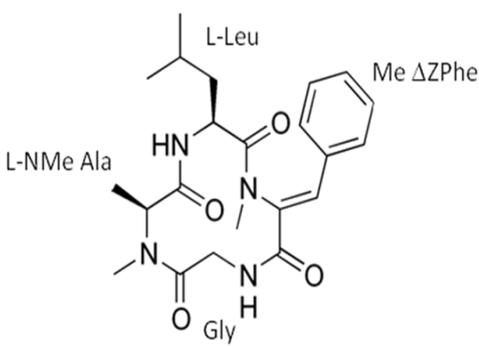
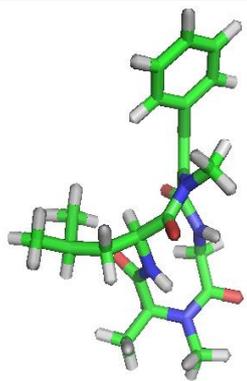
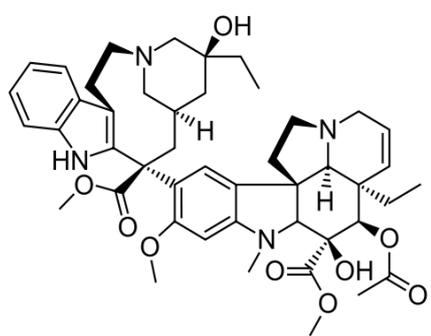
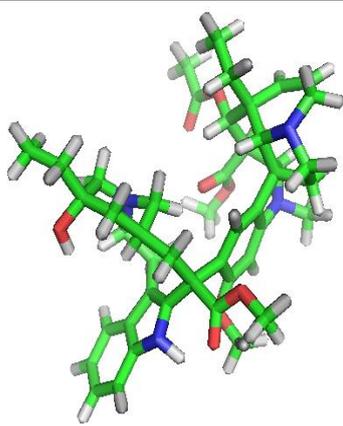
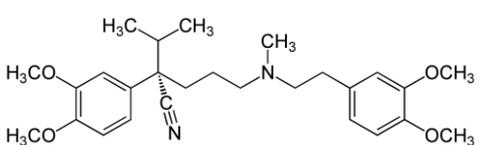
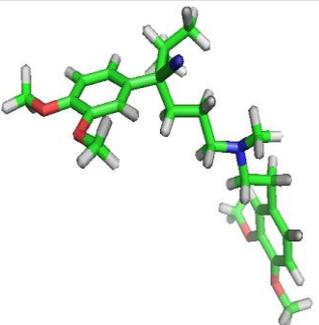
- **La tentoxine** : C'est un térapeptide cyclique naturel utilisé au laboratoire comme substrat modèle amphiphile reconnu par les enzymes de détoxication. Cette molécule a un effet phytotoxique, probablement dû à l'inhibition spécifique de l'ATP-synthase du chloroplaste. La tentoxine est un inhibiteur de la fonction de transport de la P-gp, mais un modulateur de son activité ATPasique [13].

- **La vinblastine** : Il s'agit d'un poison du fuseau mitotique. Elle est utilisée à des fins médicales, notamment dans le traitement du cancer du poumon, du cancer du sein et des lymphomes. La vinblastine est transportée par la P-gp, et module son activité ATPasique [14].

- **Le vérapamil** est un médicament qui agit comme inhibiteur des canaux calciques. Il est utilisé dans le traitement de l'hypertension, de l'angor (l'angine de poitrine), et de certains types d'arythmies. Le vérapamil est LE substrat « modèle » typique de transport de la P-gp,

reconnu par la communauté scientifique. Il inhibe la fonction de transport de la P-gp et augmente l'activité ATPasique basale de celle-ci (stimulateur de l'activité ATPasique) [14].

Nous avons repris les structures 3D de ces molécules telles qu'elles ont été utilisées pour l'étude de 2002. Ces structures sont présentées dans le tableau-D-1.

	Formule moléculaire	Structure 2D	Structure 3D
Tentoxine	$C_{22}H_{30}N_4O_4$		
Vinblastine	$C_{46}H_{58}N_4O_9$		
Vérapamil	$C_{27}H_{38}N_2O_4$		

**Tableau-D-1 : Données chimiques et structurales de substrats de la P-gp utilisés dans l'étude de docking.** *Le vérapamil est représenté dans sa forme énantiomérique S, la tentoxine dans sa conformation B « boat up » [15]. Les atomes d'oxygène sont présentés en rouge, l'azote en bleu, le carbone en vert, et l'hydrogène en blanc.*

Comme nous l'avons évoqué précédemment, le programme AutoDock dans sa version 4.2 a été utilisé pour réaliser les travaux de docking présentés. Les calculs AutoDock nécessitent en quatre étapes :

- 1) Préparation des fichiers de coordonnées : AutoDock 4.2 est paramétré pour utiliser des modèles de protéines et de ligands qui incluent les atomes hydrogènes polaires, mais pas les atomes liés aux atomes de carbone. Ainsi, des fichiers de coordonnées spécifiques, de format PDBQT, sont créés, incluant les charges atomiques partielles ainsi que les types d'atomes. Les fichiers PDBQT incluent également les degrés de liberté de torsion. Dans nos études de docking, seul le ligand est flexible et donc seul son fichier PDBQT contient les degrés de liberté de torsion. Le programme AutoDockTools [16] est utilisé pour générer ces fichiers.
- 2) Les calculs AutoGrid : AutoGrid est un programme qui pré-calculé les cartes du réseau des énergies d'interaction pour divers types d'atomes, comme les carbones aliphatiques, les carbones aromatiques... etc, avec la protéine récepteur. Dans les fichiers de paramètres grid/grille correspondant à chaque docking, nous avons limité la distance entre chaque point de la carte (grid spacing) entre 0.550 et 0.650 Å, permettant ainsi d'avoir une bonne précision de nos résultats par comparaison à la taille de l'espace 3D exploré (au niveau des domaines TMD). Dans la P-gp en effet, il n'y a pas de site actif bien défini, dans le sens où il n'est pas unique et il y a plusieurs zones d'interactions possibles. Le double pharmacophore s'étend potentiellement sur une grande zone, et sa position est mal définie, autour de nombreux hotspots possibles. De ce fait, la zone d'exploration fait au moins 40 à 50 Å de côté. En adoptant un grid spacing de l'ordre de 0.550 à 0.650 Å pour garder une bonne « résolution » de docking, chaque espace de recherche (parallélépipède) est maillé par environ 80 points dans dimension (x,y et z) (Figure-Ax-6).
- 3) Docking : Pour chaque ligand, nous avons réalisé 100 recherches conformationnelles par l'algorithme génétique Lamarckien proposé par AutoDock. Le programme analyse les différentes poses de docking proposées, et calcule un score énergétique pour chaque pose.

- 4) Analyse des résultats en utilisant AutoDockTools (ADT) : ADT inclut différentes méthodes d'analyses des résultats de docking, incluant des outils de « *clusterisation* » des ligands dans différentes poses par similarité conformationnelle, des outils de visualisation des conformations et des outils de visualisation des interactions entre les ligands et les protéines.

### **III.3. Résultats du docking**

Dans ce paragraphe, nous présenterons les résultats du docking des ligands décrits précédemment sur deux conformations modélisées de P-gp de hamster ainsi que dans les conformations équivalentes de la P-gp humaine. Il est à noter qu'étant donné la flexibilité du ligand, une énergie moyenne de liaison est calculée pour chaque cluster généré par AutoDock. Une ou plusieurs conformations du ligand peuvent définir des poses de docking quasi identiques aux scores énergétiques voisins.

En se basant sur le score énergétique généré par AutoDock, nous avons choisi les meilleures poses de docking via l'outil ADT. Chaque pose est ensuite analysée pour localiser le ligand vis-à-vis des feuillettes membranaires et déterminer les résidus de la P-gp interagissant avec les différents ligands arrimés.

#### **III.3.1. Résultats du docking sur la P-gp de hamster**

##### *III.3.1.1. La conformation Open Inward-Facing*

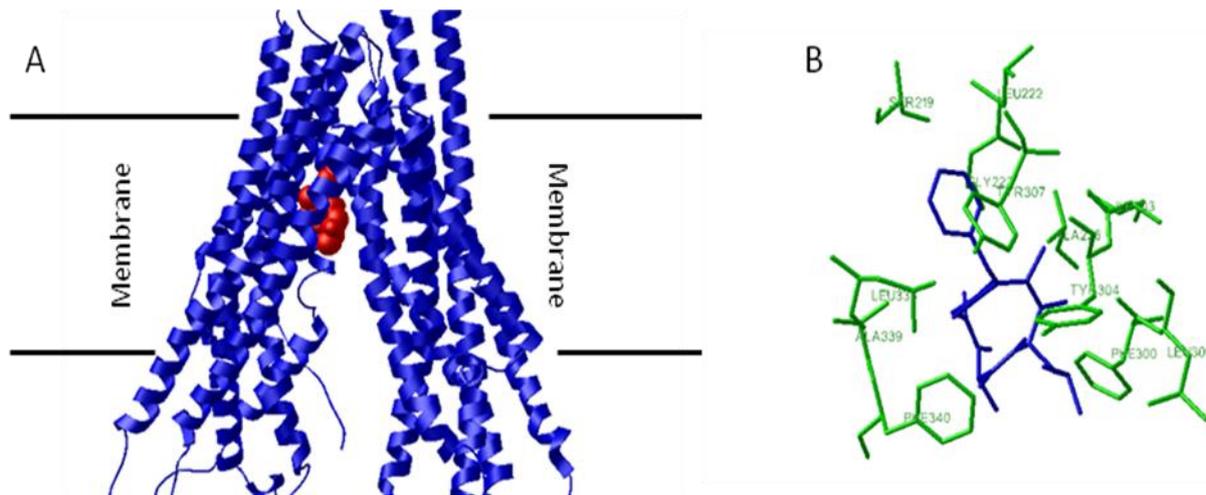
###### **1) La tentoxine**

Le docking de cette petite molécule a donné lieu à la détection d'un cluster de poses bien spécifique décrivant une énergie moyenne de liaison entre la tentoxine et la P-gp assez élevée en comparaison aux autres poses détectées par AutoDock (-8.70 Kcal/mol pour ce cluster de poses contre -7.84 Kcal/mol pour le second cluster) (voir Figure-Ax-7). Les résidus impliqués dans les interactions entre la tentoxine et la P-gp dans la meilleure pose de docking sont présentés dans le tableau-D-2.

Nous pouvons constater que l'énergie moyenne de liaison calculée par AutoDock résulte de trois poses semblables de la tentoxine. Les résidus d'interaction correspondant à chacune de ces trois poses sont localisés au niveau des hélices transmembranaires 4, 5 et 6 de la moitié N-terminale de la P-gp. Nous distinguons également que 7 résidus hotspots sont présents au niveau des trois conformations décrivant cette pose. Ces résidus partagent le point commun d'une localisation au niveau du feuillet exoplasmique de la bicouche lipidique (Figure-D-3).

Poses de docking	Pose 1	Pose 2	Pose 3
<b>Les résidus d'interaction</b>	<b>Ser 219 (TM4E)</b>	<b>Ser 219 (TM4E)</b>	<b>Ser 219 (TM4E)</b>
	Leu 222 (TM4E)	Ala 226 (TM4E)	Gly223 (TM4E)
	Gly223 (TM4E)	Phe 300 (TM5E)	Ala 226 (TM4E)
	Ala 226 (TM4E)	<b>Leu 301 (TM5E)</b>	Phe 300 (TM5E)
	Phe 300 (TM5E)	<b>Ile 303 (TM5E)</b>	<b>Leu 301 (TM5E)</b>
	<b>Leu 301 (TM5E)</b>	<b>Tyr 304 (TM5E)</b>	<b>Ile 303 (TM5E)</b>
	<b>Ile 303 (TM5E)</b>	Tyr 307 (TM5E)	<b>Tyr 304 (TM5E)</b>
	<b>Tyr 304 (TM5E)</b>	<b>Leu 336 (TM6E)</b>	Tyr 307 (TM5E)
	Tyr 307 (TM5E)	Ala 339 (TM6E)	<b>Leu 336 (TM6E)</b>
	<b>Leu 336 (TM6E)</b>	<b>Phe 340 (TM6E)</b>	<b>Ala 339 (TM6E)</b>
	<b>Ala 339 (TM6E)</b>		<b>Phe 340 (TM6E)</b>
<b>Phe 340 (TM6E)</b>			
<b>Energie moyenne de liaison (Kcal/mol)</b>	<b>- 8.70</b>		

**Tableau-D-2 : Les résidus d'interaction de la P-gp de hamster dans la conformation *Open Inward-Facing* avec la tentoxine.** Les résidus hotspots de la P-gp de hamster, homologues à ceux révélés chez la P-gp humaine, sont présentés en **gras**. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. **La lettre E** spécifie la localisation « Exoplasmique » du résidu.



**Figure-D-3 : Docking de la tentoxine sur la P-gp de hamster dans la conformation *Open Inward-Facing*.** A. Représentation de la pose 1 du docking de la tentoxine (en représentation CPK en rouge) au niveau de la P-gp de hamster (en représentation cartoon en bleu). Un positionnement approximatif de la membrane est décrit par des traits horizontaux. B. Représentation des résidus d'interaction de la tentoxine (en bleu) avec la P-gp (en vert).

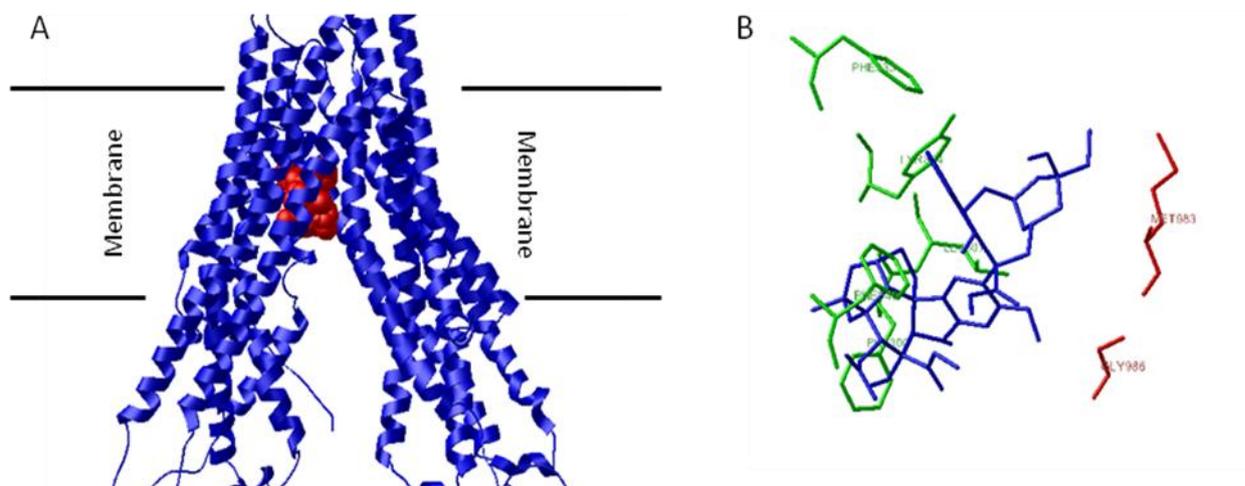
## 2) La vinblastine

Les résultats du docking de la vinblastine mettent en évidence deux clusters de poses aux valeurs énergétiques proches (Tableau-D-3). A partir de la liste des résidus d'interaction détectés, nous constatons que ces deux clusters correspondent à des poses très proches avec une majorité d'hélices et de résidus communs (TM5, 6 et 12). En comparaison avec la tentoxine, nous constatons que, pour la vinblastine, un certain nombre de résidus d'interaction avec la P-gp se trouvent dans la moitié C-terminale (TM7 et TM12), ce qui n'était pas du tout le cas pour la TTX. De manière remarquable, comme pour la tentoxine, la majorité des résidus d'interaction sont des résidus hotspots.

Du point de vue de la localisation membranaire, les résidus d'interaction P-gp de hamster/VBL sont tous localisés au niveau du feuillet exoplasmique de la bicouche lipidique (Figure-D-4). En comparant les résidus d'interaction des meilleures poses de docking de la tentoxine et de la vinblastine, nous constatons qu'il y a quatre résidus en commun : Phe 300, Leu 301, Tyr 304 et Phe 340, les trois derniers étant des hotspots. Ce résultat est en accord avec les conclusions émises par l'étude pharmacologique de 2002 qui concluait à l'existence d'une compétition et d'une exclusion mutuelle entre VBL et TTX avec des éléments communs de reconnaissance. Il est aussi intéressant de noter que les énergies de liaison correspondant aux deux ligands sont sensiblement proches.

Poses de docking	Pose 1	Pose 2	Pose 3
<b>Les résidus d'interaction</b>	Phe 300 (TM5E) <b>Leu 301 (TM5E)</b> Tyr 304 (TM5E) Phe 333 (TM6E) Phe 340 (TM6E) Met 983 (TM12E) Gly 986 (TM12E)	Phe 300 (TM5E) <b>Tyr 304 (TM5E)</b> <b>Phe 340 (TM6E)</b> Met 983 (TM12E) Gly 986 (TM12E) Ser 990 (TM12E)	Gly 297 (TM5E) Phe 300 (TM5E) <b>Leu 301 (TM5E)</b> <b>Tyr 304 (TM5E)</b> <b>Phe 340 (TM6E)</b> Gly 722 (TM7E) Phe 725 (TM7E) Met 983 (TM12E) Gly 986 (TM12E) Ser 990 (TM12E)
<b>Energie moyenne de liaison (Kcal/mol)</b>	<b>-8.90</b>		<b>-8.74</b>

**Tableau-D-3 : Les résidus d'interaction de la P-gp de hamster dans la conformation *Open Inward-Facing* avec la vinblastine dans les meilleures poses issues du protocole AutoDock.** Les résidus hotspots de la P-gp de hamster, homologues à ceux révélés chez la P-gp humaine, sont présentés en **gras**. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. **La lettre E** spécifie la localisation « Exoplasmique » du résidu.



**Figure-D-4 : Docking de la vinblastine sur la P-gp de hamster dans la conformation *Open Inward-Facing*.** A. Représentation de la pose 1 du docking de la vinblastine (en représentation CPK en rouge) au niveau de la P-gp de hamster (en représentation cartoon en bleu). Un positionnement approximatif de la membrane est décrit par des traits horizontaux. B. Représentation des résidus d'interaction de la P-gp (en vert pour la moitié N-terminale et en rouge pour la moitié C-terminale) avec la vinblastine (en bleu)

### 3) Le vérapamil

Le docking de ce substrat typique de la P-gp révèle deux poses de docking aux valeurs énergétiques très proches sont révélées par AutoDock (Tableau-D-4), avec la présence de quatre résidus communs aux deux poses : Met 68, Phe 71, Leu 329 et Phe 333.

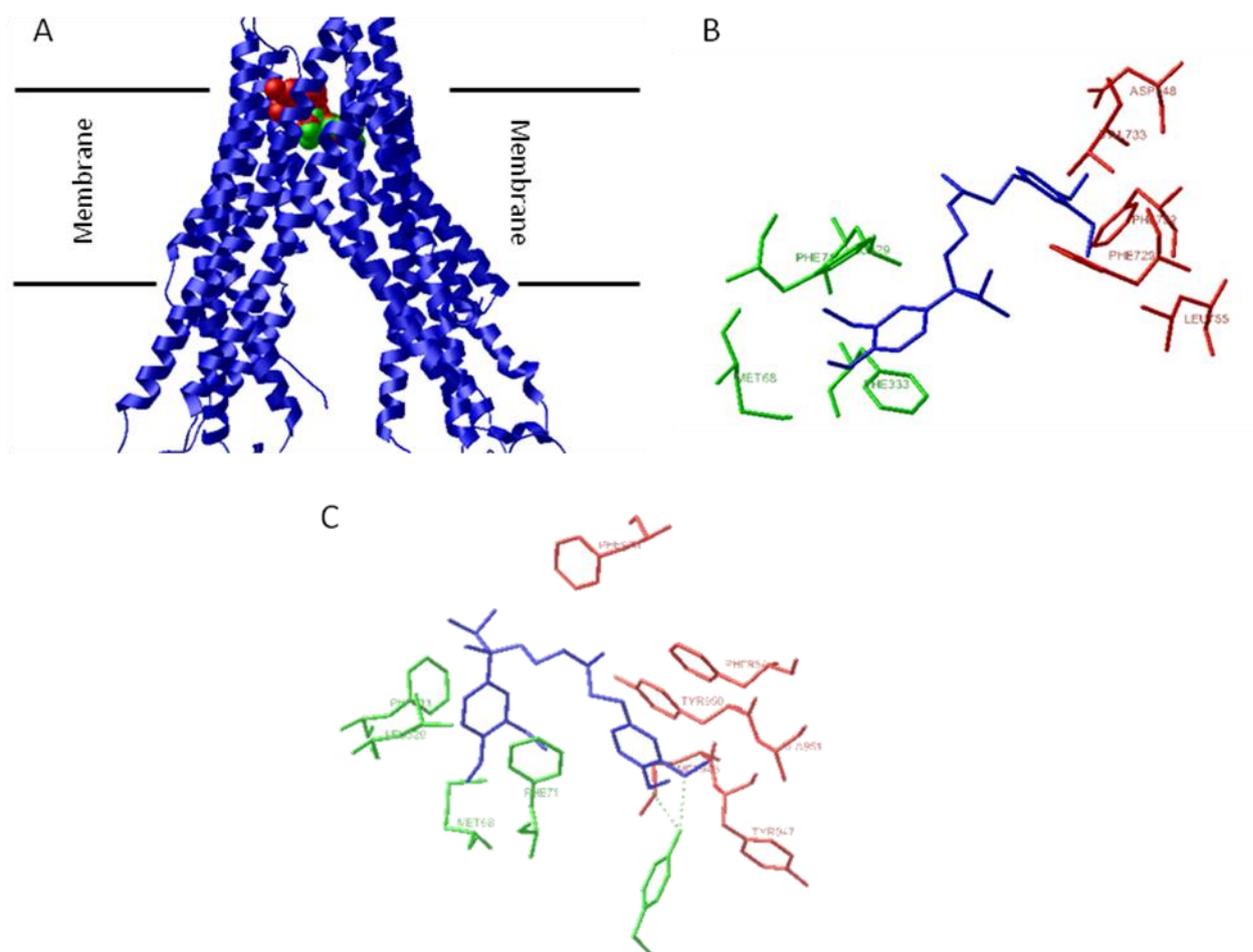
La première pose implique les résidus des hélices transmembranaires de la moitié N-terminale : TM 1 et 6, ainsi que des hélices transmembranaires de la moitié C-terminale : TM 7 et 8. La seconde pose implique, en plus des mêmes résidus de la moitié N-terminale, des résidus des hélices transmembranaires C-terminales TM11 et TM12.

Comme pour TTX et VBL, les résidus d'interaction révélés sont localisés au niveau du feuillet exoplasmique de la bicouche lipidique (Figure-D-5). Sept de ces résidus font partie des résidus hotspots.

Enfin, la comparaison des tableaux D2, D3 et D4 nous montre qu'il y a un seul résidu d'interaction commun entre VRP et VBL : le Phe 333. Aucun résidu de la P-gp interagissant avec VRP, n'interagit avec TTX. Nous sommes donc dans le cas de deux sites s'excluant mutuellement pour ces deux ligands. Jusqu'ici, sur la forme *Open Inward-Facing*, nous confirmons (de manière presque surprenante) les interactions mutuelles entre substrats déduites de l'étude *in vitro* de 2002 : VRP et TTX ne sont pas en compétition, alors que VBL est en compétition partielle avec VRP, et fait partie majoritairement du pharmacophore de la TTX (Voir Figure-D-2).

Poses de docking	Pose 1	Pose 2
Les résidus d'interaction	<b>Met 68 (TM1E)</b>	<b>Met 68 (TM1E)</b>
	Phe 71 (TM1E)	Phe 71 (TM1E)
	Leu 329 (TM6E)	Leu 329 (TM6E)
	<b>Phe 333 (TM6E)</b>	<b>Phe 333 (TM6E)</b>
	<b>Phe 729 (TM7E)</b>	Met 946 (TM11E)
	Val 733 (TM7E)	<b>Tyr 947 (TM11E)</b>
	Asp 748 (TM8E)	<b>Tyr 950 (TM11E)</b>
	Phe 752 (TM8E)	Ala 951 (TM11E)
	Leu 755 (TM8E)	<b>Phe 954 (TM11E)</b>
	<b>Phe 975 (TM12E)</b>	
<b>Energie moyenne de liaison (Kcal/mol)</b>	<b>-6.84</b>	<b>-6.73</b>

**Tableau-D-4: Les résidus d'interaction de la P-gp de hamster dans la conformation *Open Inward-Facing* avec le vérapamil dans les meilleures poses issues du protocole AutoDock. Les résidus hotspots de la P-gp de hamster, homologues à ceux révélés chez la P-gp humaine, sont présentés en gras. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. La lettre E spécifie la localisation « Exoplasmique » du résidu.**



**Figure-D-5 : Docking du vérapamil sur la P-gp de hamster dans la conformation *Open Inward-Facing*.** A. Représentation des deux meilleures poses de docking du vérapamil (en représentation CPK en rouge pour la première et en CPK en vert pour la seconde) au niveau de la P-gp (en représentation cartoon en bleu). Un positionnement approximatif de la membrane est décrit par des traits horizontaux. B. Représentation des résidus d'interaction de la P-gp (en vert pour la moitié N-terminale et en rouge pour la moitié C-terminale) avec le vérapamil (en bleu) dans la première meilleure pose (en rouge dans A). C. Même légende que B avec le vérapamil représenté dans sa seconde meilleure pose.

### III.3.1.2. Docking simultané de deux substrats : VRP et TTX sur la forme *Open Inward-Facing* de la P-gp de hamster

Ces premiers résultats de docking mono-substrat sont compatibles avec l'absence d'une compétition entre les deux ligands VRP et TTX. Nous avons alors pensé confirmer ces résultats en réalisant des expériences de docking simultané, qui consistent à arrimer un des deux ligands (VRP ou TTX) quand l'autre ligand est déjà arrimé dans sa meilleure pose précédemment révélée (« co-docking séquentiel »).

### 1) Docking du VRP avec la TTX déjà arrimée

Les résultats de ce docking (Tableau-D-5) sont parfaitement compatibles avec ceux révélés par le docking du VRP seul : Deux poses de docking sont possibles pour le vérapamil malgré l'arrimage préalable de la TTX. Ces poses présentent des localisations membranaires semblables à celles présentées précédemment:

- i) Les énergies de liaison correspondant à ces poses sont sensiblement proches de celles détectées précédemment.
- ii) Les résidus d'interaction détectés sont tous localisés au niveau du feuillet exoplasmique de la bicouche lipidique.
- iii) Les deux poses révélées impliquent une majorité des hélices transmembranaires précédemment révélées (les TM1, 2, 6, 11 et 12).
- iv) Des résidus hotspots font toujours partie des résidus d'interaction entre la P-gp et le VRP.

En conclusion, VRP interagit donc avec la P-gp de manière indépendante de l'arrimage préalable de la TTX. Une vue de dessus de la P-gp, liant simultanément VRP et TTX, est présentée Figure-D-6A.

Poses de docking	Pose 1	Pose 2
<b>Les résidus d'interaction</b>	<b>Met 68 (TM1E)</b> Phe 71 (TM1E) Tyr 114 (TM2E) Thr 330 (TM6E) <b>Phe 333 (TM6E)</b> <b>Tyr 950 (TM11E)</b> <b>Phe 954 (TM11E)</b>	Met 67 (TM1E) <b>Met 68 (TM1E)</b> Phe 71 (TM1E) Tyr 114 (TM2E) <b>Phe 333 (TM6E)</b> Met 946 (TM11E) <b>Ala 951 (TM11E)</b> <b>Phe 975 (TM12E)</b>
<b>Energie moyenne de liaison (Kcal/mol)</b>	<b>-7.12</b>	<b>-6.25</b>

**Tableau-D-5:** Les résidus d'interaction de la P-gp de hamster dans la conformation *Open Inward-Facing* avec le vérapamil après un arrimage préalable de la TTX. Les résidus hotspots de la P-gp de hamster, homologues à ceux révélés chez la P-gp humaine, sont présentés en **gras**. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. **La lettre E** spécifie la localisation « Exoplasmique » du résidu.

## 2) Docking de la TTX avec le VRP déjà arrimé

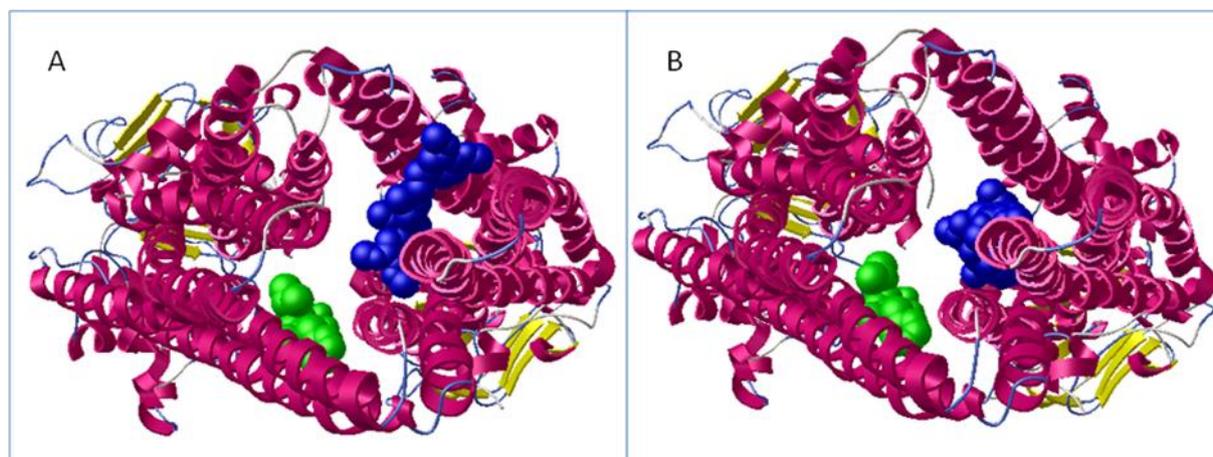
Dans ce paragraphe, nous présentons les résultats du docking inverse. Comme dans le cas du VRP, les résultats de ce docking (Tableau-D-6) sont également en accord avec ceux décrivant le docking de la TTX seule, à savoir:

- i) Le cluster des meilleures poses de ce docking se caractérise par un score énergétique identique à celui révélé précédemment.
- ii) Ce score reflète quatre conformations de la TTX avec des résidus d'interaction avec la P-gp quasi identiques.
- iii) Ces résidus d'interaction sont les mêmes que ceux révélés par le docking de la TTX seule, y compris les résidus hotspots.
- iv) La localisation membranaire est identique à celle présentée précédemment.
- v) La TTX interagit donc avec la P-gp de manière indépendante de l'arrimage préalable du VRP. Une vue de dessus de la P-gp, liant simultanément VRP et TTX, est présentée Figure-D-6B.

Nous avons donc, sur la base de notre modèle structural de la P-gp de hamster en conformation *Open Inward-Facing* que les résultats de docking simultané de la tentoxine et du vérapamil sont en accord avec les conclusions émises par l'étude pharmacologique de 2002 quant à l'existence de deux sites distincts nécessaire pour interpréter l'absence de compétition entre ces deux ligands pour moduler l'activité ATPasique de la P-gp (utilisée comme « sonde fonctionnelle d'interaction » des ligands pour leur liaison à la P-gp).

Poses de docking	Pose 1	Pose 2	Pose 3	Pose 4
<b>Les résidus d'interaction</b>	<b>Ser 219 (TM4E)</b>	<b>Ser 219 (TM4E)</b>	<b>Ser 219 (TM4E)</b>	<b>Ser 219 (TM4E)</b>
	Leu 222 (TM4E)	Ala 226 (TM4E)	Gly223 (TM4E)	Leu 222 (TM4E)
	Ala 226 (TM4E)	Phe 300 (TM5E)	Ala 226 (TM4E)	Gly223 (TM4E)
	Gly 227 (TM4E)	<b>Leu 301 (TM5E)</b>	Phe 300 (TM5E)	Ala 226 (TM4E)
	Phe 300 (TM5E)	<b>Ile 303 (TM5E)</b>	<b>Leu 301 (TM5E)</b>	Gly 227 (TM4E)
	<b>Leu 301 (TM5E)</b>	<b>Tyr 304 (TM5E)</b>	<b>Ile 303 (TM5E)</b>	Phe 300 (TM5E)
	<b>Ile 303 (TM5E)</b>	Tyr 307 (TM5E)	<b>Tyr 304 (TM5E)</b>	<b>Ile 303 (TM5E)</b>
	<b>Tyr 304 (TM5E)</b>	<b>Leu 336 (TM6E)</b>	<b>Leu 336 (TM6E)</b>	<b>Tyr 304 (TM5E)</b>
	Tyr 307 (TM5E)	Ala 339 (TM6E)	Ala 339 (TM6E)	Tyr 307 (TM5E)
	<b>Leu 336 (TM6E)</b>	<b>Phe 340 (TM6E)</b>	<b>Phe 340 (TM6E)</b>	<b>Leu 336 (TM6E)</b>
	Ala 339 (TM6E)			Ala 339 (TM6E)
	<b>Phe 340 (TM6E)</b>			<b>Phe 340 (TM6E)</b>
<b>Energie moyenne de liaison (Kcal/mol)</b>	<b>-8.70</b>			

**Tableau-D-6: Les résidus d'interaction de la P-gp hamster dans la conformation *Open Inward-Facing* avec la tentoxine avec un arrimage préalable du vérapamil. Les résidus hotspots de la P-gp de hamster, homologues à ceux révélés chez la P-gp humaine, sont présentés en *gras*. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. La lettre E spécifie la localisation « Exoplasmique » du résidu.**

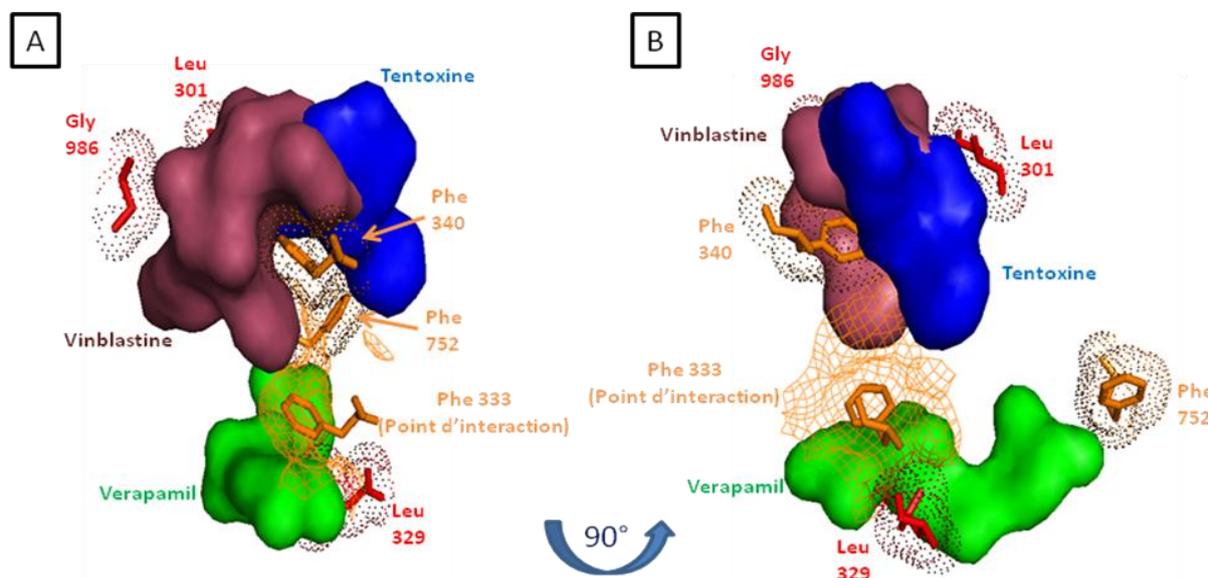


**Figure-D-6 : Docking simultané du vérapamil et de la tentoxine sur la P-gp de hamster dans la conformation *Open Inward-Facing*.. A. Vue de dessus de la P-gp de hamster dans la conformation *Open-Inward* présentant la meilleure pose du docking du VRP (en représentation CPK en bleu) avec arrimage préalable de la tentoxine (en représentation CPK en vert). B. Vue de dessus de la P-gp dans une conformation *Open-Inward* présentant la meilleure pose du docking de la tentoxine (en représentation CPK en vert) avec arrimage préalable du VRP.**

### III.3.1.3. Analyse des interactions mutuelles entre VRP, TTX et VBL et de leurs sous-sites respectifs

Les résultats des diverses expériences de docking, présentées ci-dessus, mettent en évidence des relations d'interaction entre les trois substrats de la P-gp étudiés (Figure-D-7) :

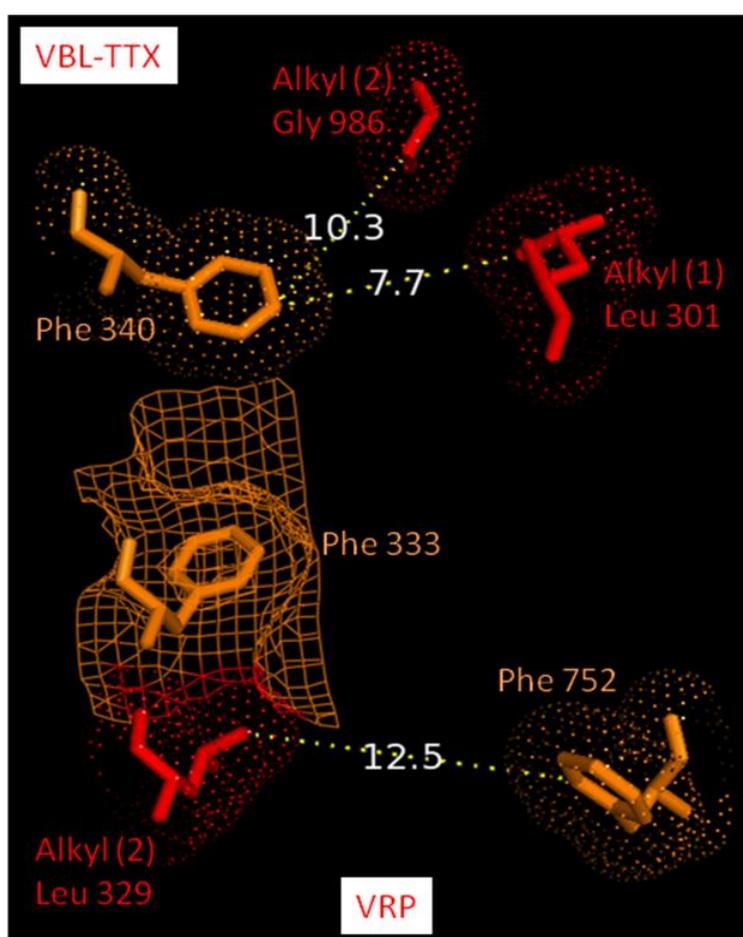
- i) Avec quatre résidus communs d'interaction, VBL et TTX partagent un même sous-site de liaison. Ces deux substrats de la P-gp seraient donc en compétition dans leur liaison à la P-gp par appartenance à une même « classe de compétition », c'est-à-dire présentant des homologies structurales avec des éléments communs de reconnaissance.
- ii) TTX, VBL et VRP se lient à la P-gp dans sa conformation *Open Inward-Facing* au niveau du feuillet exoplasmique de la bicouche lipidique.
- iii) VRP possède un sous-site de liaison indépendant de celui qui lie VBL et TTX.
- iv) VBL et VRP seraient également en compétition, mais cette fois par exclusion stérique, à cause de leur liaison à un même acide aminé aromatique (Phe 333). Ce résidu pourrait représenter le point d'interaction déduit de l'étude pharmacophorique de 2002.



**Figure-D-7 : Cartographie des interactions entre VBL, VRP et TTX.** VRP, VBL et TTX sont respectivement présentés en vert, rouge clair et bleu en mode surface de Connolly. Le point d'interaction (Phe 333) entre VBL et VRP est présenté en orange en représentation mesh et licorice. Les acides aminés, pouvant fournir un groupe alkyle sont présentés en rouge en représentation dots, alors que les acides aminés pouvant fournir un groupe aromatique sont présentés en orange en représentation dots. Pour une meilleure visibilité, la cartographie a été présentée sous deux angles de vue différents (B : même représentation que A mais avec une rotation de 90° par rapport un axe vertical)

Les datas de docking permettent de préciser l'orientation spatiale relative des deux sous-sites, que l'étude enzymologique ne pouvait évidemment pas prévoir. Il faut ainsi noter que le double pharmacophore de 2002 a été publié avec une marge d'erreur symbolisée par un axe de rotation et un axe de « bending ». Notre bipharmaconophore « structural » reste dans les limites proposées en 2002.

Pour parachever la comparaison avec l'étude de 2002, nous avons recherché les paramètres d'orientation spatiale de nos ligands en compétition dans le modèle structural. Ils sont donnés à titre indicatif, sachant que la protéine récepteur a été considérée comme rigide dans les calculs AutoDock, sans avoir la possibilité d'adapter l'orientation des chaînes latérales au cours du docking. Nous avons étudié la distribution des surfaces hydrophobes (groupes alkyles et aromatiques) pour la comparer avec celle prédite par Garrigues *et al.* (Figure-D-8).



**Figure-D-8 : Distribution spatiale des groupes alkyles et aromatiques de la P-gp en contact avec les différents substrats.** Les distances entre les différents groupes ont été calculées grâce au programme de visualisation Pymol.

- On trouve bien 2 alkyles et un aromatique dans le site VBL/TTX, et un alkyle et un aromatique dans le site VRP, et les distances calculées entre les différents résidus correspondent aux estimations faites par Garrigues et al.

### III.3.1.4. La conformation *Outward-Facing*

Pour l'étude de cette conformation, nous retenons une seule pose de docking pour chaque ligand, celle du meilleur score, à l'exception du VRP pour lequel deux poses étaient favorables d'énergie équivalente. Les résultats de l'analyse des résidus d'interaction entre la P-gp et chaque ligand sont ainsi regroupés dans un même tableau (Tableau-D-7).

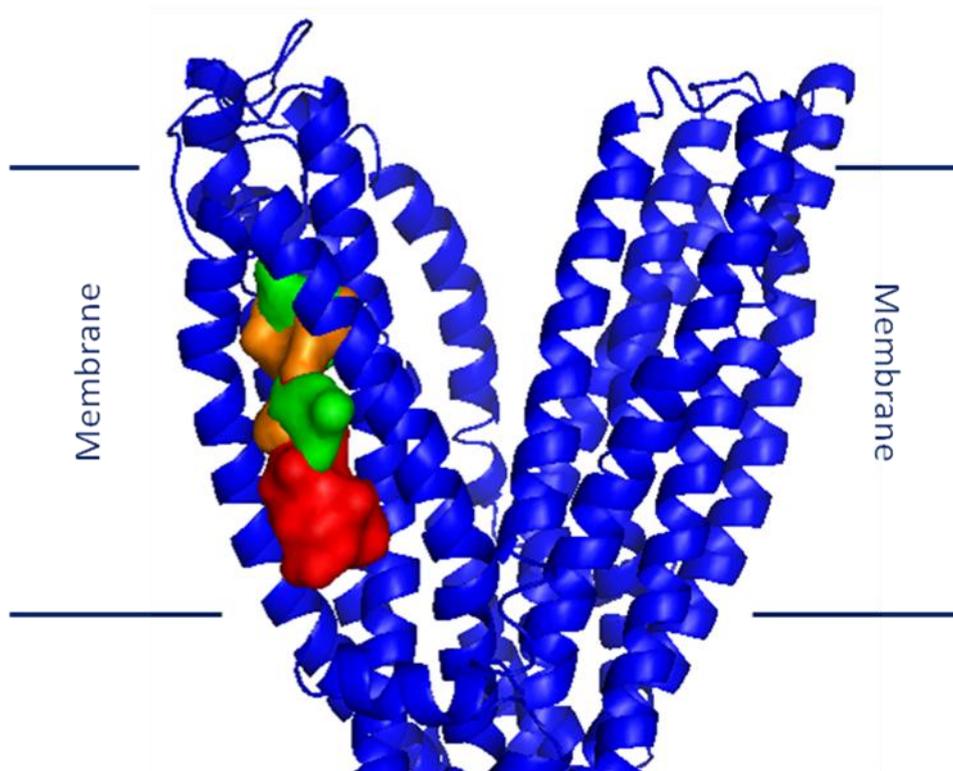
Ligand	TTX	VBL	VRP	
Energie moyenne de liaison (Kcal/mol)	-9.14	-8.08	-7.55	-7.39
Conformations	Conformation	Conformation	Conformation 1	Conformation 2
Les résidus d'interaction	Gln 133 (TM2C) <b>Leu 840(TM9E)</b> Ile 844 (TM9) <b>Val 862 (TM10E)</b> <b>Ile 865 (TM10E)</b> Ala 866 (TM10) Leu 868 (TM10C) Ala 977 (TM12E) <b>Ile 978 (TM12E)</b> <b>Gly 981 (TM12E)</b>	Ile 833 (TM9C) Ile 837 (TM9) <b>Leu 840 (TM9E)</b> <b>Val 862 (TM10E)</b> <b>Ile 865 (TM10E)</b> <b>Gly 869 (TM10C)</b> Val 870 (TM10C) Met 873 (TM10C) <b>Thr 942 (TM11C)</b> <b>Val 985 (TM12E)</b> <b>Val 988 (TM12E)</b> Ser 989 (TM12) Phe 991 (TM12C) Ala 992 (TM12C) Tyr 995 (TM12C)	Gln 133 (TM2C) Ile 837 (TM9) <b>Leu 840 (TM9E)</b> Gly 841 (TM9E) Gly 843 (TM9E) Ser 847 (TM9E) Leu 854 (TM10) Leu 858 (TM10E) Ser 847 (TM9E) Leu 854 (TM10) Leu 858 (TM10E) <b>Val 862 (TM10E)</b> <b>Ile 865 (TM10E)</b> Met 946 (TM11E) <b>Val 862 (TM10E)</b> <b>Ile 865 (TM10E)</b> Met 946 (TM11E) <b>Ile 978 (TM12E)</b> <b>Gly 981 (TM12E)</b> <b>Val 985 (TM12E)</b>	<b>Leu 840 (TM9E)</b> Gly 843 (TM9E) Ser 847 (TM9E) Leu 854 (TM10) Leu 858 (TM10E) <b>Val 862 (TM10E)</b> <b>Ile 865 (TM10E)</b> Met 946 (TM11E) Ala 977(TM12E) <b>Ile 978 (TM12E)</b> <b>Gly 981 (TM12E)</b> <b>Val 985 (TM12E)</b> <b>Val 988 (TM12E)</b>

**Tableau-D-7: Les résidus d'interaction de la P-gp de hamster dans la conformation *Outward-Facing* avec la tentoxine, la vinblastine et le vérapamil.** Sont présentés en *gras* les résidus hotspots détectés par alignement de séquences entre la P-gp humaine et la P-gp de hamster. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. **La lettre E** spécifie la localisation membranaire « Exoplasmique » du résidu. **La lettre C** spécifie la localisation membranaire « Cytoplasmique » du résidu. Les résidus communs aux différentes poses de docking sont présentés en rouge. Les résidus, pour lesquels une localisation membranaire n'a pas été spécifiée, sont ceux localisés au cœur de la membrane,

L'analyse comparée de ces données de fixation révèle que :

- i) Même dans la conformation *Outward-Facing*, des résidus hotspots font partie des résidus d'interaction de la P-gp avec les trois ligands étudiés. Ces résidus sont donc concernés par l'ensemble du cycle de translocation du substrat.
- ii) Une seule pose de docking est retenue pour VBL, TTX. Dans le cas du VRP, deux poses sont détectées comme cela a été pour le docking dans la conformation *Open Inward-Facing*. Mais nous pouvons constater à partir des résidus de la P-gp avec lesquels le VRP interagit que ce sont des poses quasi identiques dans leur localisation et leur orientation.
- iii) Dans cette conformation, les hélices transmembranaires impliquées dans la liaison aux substrats sont des hélices de la moitié C-terminale de la P-gp : TM9, 10, 11 et 12 révélant des sous-sites de liaisons différents de ceux détectés dans la conformation *Open Inward-Facing*.
- iv) Ces sous-sites de liaison se répartissent sur toute la bicouche lipidique du feuillet exoplasmique, y compris le feuillet cytoplasmique (Figure-D-9).

La comparaison des résidus impliqués en commun avec la conformation Open-Inward facing montre que pour TTX les sites sont nettement différents, alors que pour VBL et VRP il y a quelques résidus en commun (983-990 et 946-975-978 respectivement). Cela indique qu'il y aurait possiblement deux types de mécanismes de « déstabilisation relargante » en fonction de la nature du ligand transporté, soit par changement de site, soit par scindage du site, se traduisant dans les deux cas par une diminution de l'affinité, caractéristique d'un transport actif.



**Figure-D-9 : Docking des substrats de la P-gp dans sa conformation *Outward-Facing*.** Représentation des meilleures poses de docking du vérapamil (en représentation surface en vert), de la tentoxine (en représentation surface en orange) et de la vinblastine (en représentation surface en rouge) au niveau de la P-gp (en représentation cartoon en bleu). Un positionnement approximatif de la membrane est décrit par des traits horizontaux.

### III.3.2. Résultats du docking sur la P-gp humaine

Au vu des résultats intéressants obtenus à l'issu des travaux de docking réalisés sur des modèles 3D de la P-gp de hamster dans deux conformations catalytiques, nous avons décidé de réaliser le docking des mêmes substrats (VBL, VRP et TTX) au niveau des modèles 3D de la P-gp humaine (pour laquelle les hotspots ont été déterminés) dans les mêmes conformations catalytiques (*Open Inward-Facing* et *Outward-Facing*), dans le but de tester les caractéristiques qui restent vraies.

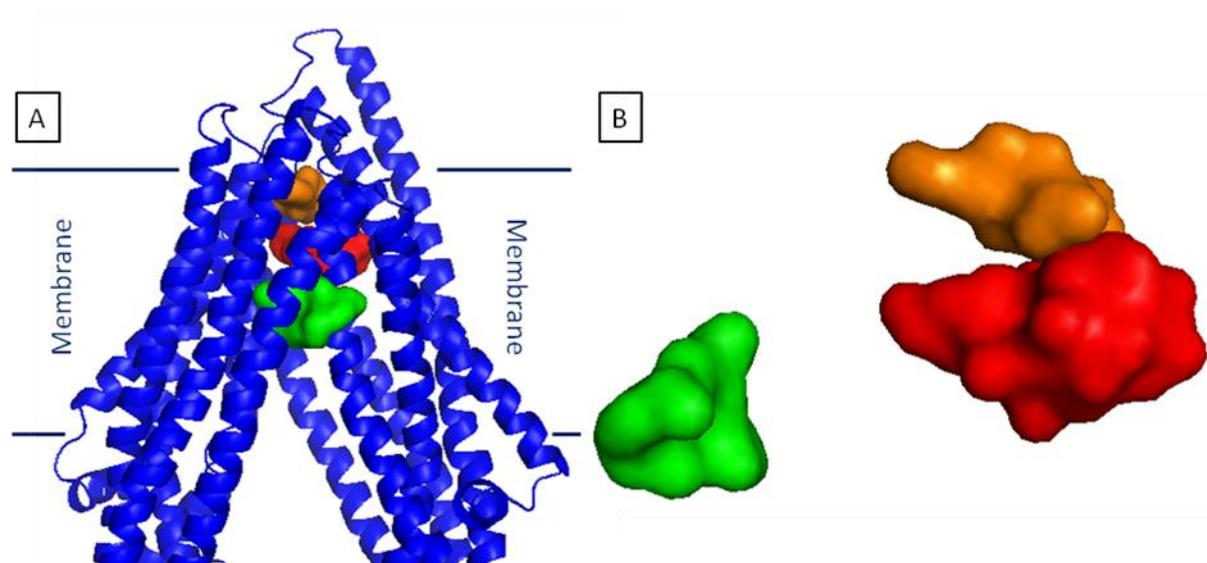
#### III.3.2.1. La conformation *Open Inward-Facing*

Les résultats des différents docking (Tableau-D-8), réalisés sur la P-gp humaine dans sa conformation catalytique ouverte vers le milieu intracellulaire, décrivent:

- Une localisation membranaire exoplasmique de TTX, VBL et VRP, semblable à celle révélée précédemment sur la P-gp de hamster (Figure-D-10).
- Pour la TTX, pas de résidus d'interaction communs avec la P-gp de hamster. Ces résidus font majoritairement partie des hélices transmembranaires TM1 et TM7.
- Un résidu hotspot fait partie des résidus d'interaction de la P-gp humaine avec TTX : Phe 975.
- Pour la VBL, un seul résidu d'interaction en commun avec la P-gp de hamster qui est le résidu hotspot Phe 336 (Phe 333 pour la P-gp de hamster). Les hélices transmembranaires impliquées par les divers résidus d'interaction sont TM1, 2, 6, 7, 11 et 12.
- Cinq résidus hotspots font partie des résidus d'interaction de la P-gp humaine avec VBL : Phe 336, Phe 732, Tyr 953, Leu 975 et Phe 978.
- Deux résidus de la P-gp humaine interagissent à la fois avec TTX et VBL : Phe 72 et Leu 975. Ceci traduit le maintien d'une compétition entre les deux substrats dans leur liaison à la protéine chez l'Homme.
- Les énergies de liaisons correspondant aux meilleures poses de docking de TTX et VBL sont assez proches, comme cela a été le cas précédemment pour leur docking dans la P-gp de hamster.
- Pour le VRP, les résidus d'interaction impliqués ne sont pas homologues à ceux détectés précédemment. Ils appartiennent aux trois dernières hélices transmembranaires de la moitié N-terminale de la P-gp : TM4, TM5 et TM6.
- Cinq résidus de la P-gp humaine font partie des résidus d'interaction avec VRP : Ser 222, Ile 306, Gly 341, Ala 342, Phe 343. Tous sont des résidus hotspots.
- Malgré l'absence d'un point d'interaction distinctif comme cela a été le cas pour la P-gp de hamster, VBL et VRP demeurent spatialement voisins. La grande poche de liaison décrite précédemment pour la P-gp de hamster dans la conformation Open Inward-Facing semble exister pour la P-gp humaine aussi.

Ligand	TTX	VBL	VRP
Energie moyenne de liaison (Kcal/mol)	-7.18	-7.53	-7.75
Poses de docking	Pose	Pose	Pose
Les résidus d'interaction	Phe 72 (TM1E) Thr 76 (TM1E) Phe 79 (TM1E) Trp 315 (TM5E) Ile 736 (TM7E) Phe 739 (TM7E) Thr 740 (TM7E) <b>Leu 975 (TM12E)</b>	Phe 72 (TM1E) Met 75 (TM1E) Tyr 117 (TM2E) <b>Phe 336 (TM6E)</b> <b>Phe 732 (TM7E)</b> <b>Tyr 953 (TM11E)</b> Ala 954 (TM11E) Phe 957 (TM11E) <b>Leu 975 (TM12E)</b> <b>Phe 978 (TM12E)</b>	<b>Ser 222 (TM4E)</b> Pro 223 (TM4E) Leu 225 (TM4E) Gly 226 (TM4E) Ala 229 (TM4E) Ala 230 (TM4E) Phe 303 (TM5E) <b>Ile 306 (TM5E)</b> Tyr 310 (TM5E) <b>Gly 341 (TM6E)</b> <b>Ala 342 (TM6E)</b> <b>Phe 343 (TM6E)</b>

**Tableau-D-8: Les résidus d'interaction de la P-gp humaine avec la tentoxine, la vinblastine et le vérapamil. . Les résidus hotspots de la P-gp humaine, sont présentés en gras. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. La lettre E spécifie la localisation « Exoplasmique » du résidu.**



**Figure-D-10 : Docking des substrats de la P-gp humaine dans sa conformation *Open Inward-Facing*. A. Représentation des meilleures poses de docking du vérapamil, de la tentoxine et de la vinblastine au niveau de la P-gp. Le code couleur ainsi que la représentation choisies pour la protéine et ses substrats sont les mêmes que ceux utilisés dans la Figure-D-9. Un positionnement approximatif de la membrane est décrit par des traits horizontaux. B. Représentation spatiale des trois substrats arrimés.**

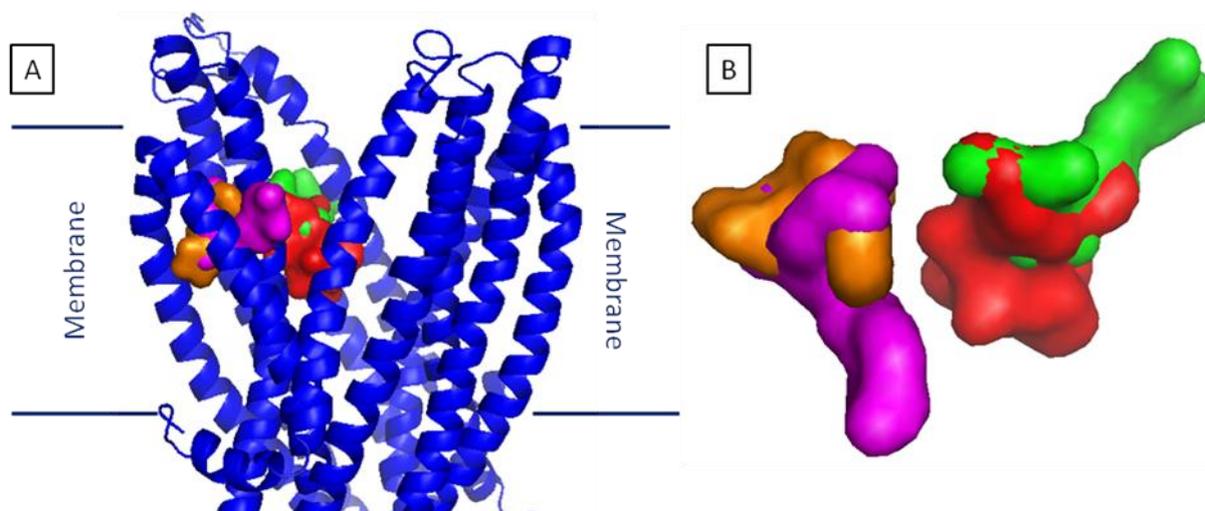
### III.3.2.2. La conformation *Outward-Facing*

Les résultats des calculs de docking réalisés sur la P-gp humaine dans la conformation *Outward-Facing* ouverte vers le milieu extracellulaire (Tableau-D-9), décrivent :

- Une pose de docking distinctive pour TTX et VBL contre deux poses révélées à l'issue du docking de VRP (Figure-D-11), comme pour la forme hamster.
- Des résidus hotspots font partie des résidus d'interaction de la P-gp humaine dans sa conformation *Outward-Facing* avec ses substrats TTX, VBL et VRP.
- La pose de docking de TTX dans la P-gp humaine est quasi identique à celle révélée précédemment pour le docking de ce substrat dans la P-gp de hamster. Elle implique, ainsi, les hélices transmembranaires TM9, 10, et 12 de la moitié C-terminale.
- Comme ça a été le cas avec la P-gp de hamster, les résidus d'interaction de la P-gp humaine, dans sa conformation *Outward-Facing*, avec TTX reflète une localisation membranaire exoplasmique de ce substrat.
- La pose de docking de VBL dans la P-gp humaine est proche de celle révélée précédemment pour le docking de ce substrat dans la P-gp de hamster. La localisation de substrat est cependant exoplasmique, avec deux résidus d'interaction en commun avec TTX : Met 949 et Ala 985.
- Les résidus d'interaction de la P-gp humaine avec VBL impliquent les hélices transmembranaires : TM1, 2, 11 et 12.
- VRP présente deux poses de docking « préférentielles », présentant chacune une grande identité avec les poses révélées pour TTX et VBL. Ces deux poses reflètent donc une localisation membranaire exoplasmique de VRP.
- La première pose, proche de celle de VBL, impliquent majoritairement les hélices transmembranaires de la moitié N-terminale : TM1 et TM2, ainsi que quelques résidus des hélices transmembranaires de la moitié C-terminale: TM11, TM12.
- La seconde pose est quasi identique à celle de TTX, avec 9 résidus d'interaction en commun. Elle est aussi quasi identique à la pose révélée par le docking de VRP dans la P-gp de hamster.

Ligand	TTX	VBL	VRP	
Energie moyenne de liaison (Kcal/mol)	-8.69	-7.60	-6.91	-6.89
Conformations	Conformation	Conformation	Conformation 1	Conformation 2
Les résidus d'interaction	<b>Leu 843 (TM9E)</b>		Met 69 (TM1E)	Asn 842 (TM9E)
	Gly 846 (TM9E)	Glu 74 (TM1E)	Leu 70 (TM1E)	<b>Leu 843 (TM9E)</b>
	Ile 847 (TM9E)	Tyr 117 (TM2E)	Phe 72 (TM1E)	Gly 846 (TM9E)
	Ser 850 (TM9E)	Val 125 (TM2E)	Gly 73 (TM1E)	Ile 847 (TM9E)
	Leu 861 (TM10E)	<b>Phe 942 (TM11C)</b>	Thr 76 (TM1E)	Ile 849 (TM9E)
	Val 865 (TM10E)	<b>Gln 946 (TM11E)</b>	Tyr 117 (TM2E)	Ser 850 (TM9E)
	<b>Ile 868 (TM10)</b>	<b>Met 949 (TM11E)</b>	Gly 120 (TM2E)	Leu 861 (TM10E)
	<b>Met 949 (TM11E)</b>	<b>Tyr 950 (TM11E)</b>	Gly 124 (TM2E)	Val 865 (TM10E)
	Ala 980 (TM12E)	<b>Tyr 953 (TM11E)</b>	Val 125 (TM2E)	<b>Val 981 (TM12E)</b>
	<b>Val 981 (TM12E)</b>	<b>Ala 985 (TM12E)</b>	<b>Met 949 (TM11E)</b>	<b>Phe 983 (TM12E)</b>
	<b>Gly 984 (TM12E)</b>	Gly 989 (TM12E)	<b>Tyr 950 (TM11E)</b>	<b>Gly 984 (TM12E)</b>
	<b>Ala 985 (TM12E)</b>	Gln 990 (TM12E)	<b>Tyr 953 (TM11E)</b>	Ala 987 (TM12E)
	Val 988 (TM12E)		<b>Ala 985 (TM12E)</b>	Val 988 (TM12E)

**Tableau-D-9 : Les résidus d'interaction de la P-gp humaine avec la tentoxine, la vinblastine et le vérapamil. .** Les résidus hotspots sont présentés en **gras**. Le numéro de l'hélice transmembranaire concernée par l'interaction est présenté entre parenthèses. Les résidus d'interaction en commun entre les différentes poses sont présentés en rouge, ceux en commun entre les poses de VRP et TTX sont présentés en noir alors que ceux en commun entre VRP et VBL sont présentés en vert. **La lettre E** spécifie la localisation « Exoplasmique » du résidu. La lettre C spécifie la localisation « cytoplasmique » du résidu.



**Figure-D-11 : Docking des substrats de la P-gp humaine dans sa conformation Outward-Facing.** A. Représentation des meilleures poses de docking du vérapamil, de la tentoxine et de la vinblastine au niveau de la P-gp. Le code couleur ainsi que la représentation choisis pour la protéine et ses substrats sont les mêmes que ceux utilisés dans la Figure-D-9. Le vérapamil, arrimé dans sa seconde pose, est représenté en magenta. Un positionnement approximatif de la membrane est décrit par des

traits horizontaux. B. Représentation spatiale des trois substrats arrimés avec le vérapamil présenté dans ses deux poses préférentielles.

#### **IV. Conclusion**

Nos travaux de docking réalisés sur deux modèles de la P-gp de hamster décrivant les conformations catalytiques (*Open Inward-Facing* et *Outward-Facing*) sont en accord avec les observations et les hypothèses émises par l'étude pharmacologique de 2002.

Nous pouvons constater qu'il y a beaucoup moins de différences entre les formes hamster et humaine pour la conformation *Outward-Facing* que pour la conformation *Open Inward-Facing*. Ceci peut suggérer un mécanisme général d'expulsion des ligands, alors que pour leur liaison, l'important est le caractère multispécifique réalisé par le grand nombre de sites possibles obtenu en combinant les différents éléments de reconnaissance dans une grande poche, conduisant à divers types de relations mutuelles entre les ligands transportés.

L'importance des résidus hotspots est confirmée par ces travaux de docking. Ces résidus sont concernés par l'ensemble du cycle de translocation du substrat.

## V. Bibliographie

1. Garrigues, A., et al., *Characterization of two pharmacophores on the multidrug transporter P-glycoprotein*. Mol Pharmacol, 2002. **62**(6): p. 1288-98.
2. Brooijmans, N. and I.D. Kuntz, *Molecular recognition and docking algorithms*. Annu Rev Biophys Biomol Struct, 2003. **32**: p. 335-73.
3. Leach, A.R. and S.R. Kilvington, *Automated molecular design: a new fragment-joining algorithm*. J Comput Aided Mol Des, 1994. **8**(3): p. 283-98.
4. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J Comput Aided Mol Des, 2001. **15**(5): p. 411-28.
5. Kramer, B., M. Rarey, and T. Lengauer, *Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking*. Proteins, 1999. **37**(2): p. 228-41.
6. Jain, A.N., *Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine*. J Med Chem, 2003. **46**(4): p. 499-511.
7. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. J Med Chem, 2004. **47**(7): p. 1739-49.
8. Halgren, T.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening*. J Med Chem, 2004. **47**(7): p. 1750-9.
9. Venkatachalam, C.M., et al., *LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites*. J Mol Graph Model, 2003. **21**(4): p. 289-307.
10. Trosset, J.Y. and H.A. Scheraga, *Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines*. Proc Natl Acad Sci U S A, 1998. **95**(14): p. 8011-5.
11. Goodsell, D.S., G.M. Morris, and A.J. Olson, *Automated docking of flexible ligands: applications of AutoDock*. J Mol Recognit, 1996. **9**(1): p. 1-5.
12. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. J Comput Chem, 2009. **30**(16): p. 2785-91.
13. Santolini, J., et al., *Kinetic analysis of tentoxin binding to chloroplast F1-ATPase. A model for the overactivation process*. J Biol Chem, 1999. **274**(2): p. 849-58.
14. Garrigos, M., L.M. Mir, and S. Orłowski, *Competitive and non-competitive inhibition of the multidrug-resistance-associated P-glycoprotein ATPase--further experimental evidence for a multisite model*. Eur J Biochem, 1997. **244**(2): p. 664-73.
15. Loiseau, N., et al., *Predicting the conformational states of cyclic tetrapeptides*. Biopolymers, 2003. **69**(3): p. 363-85.
16. Sanner, M.F., *Python: a programming language for software integration and development*. J Mol Graph Model, 1999. **17**(1): p. 57-61.

# Bilan général et perspectives

# Bilan général et perspectives

---

## I. Bilan général

Comme cela était indiqué à la fin du chapitre A, le but général de ce travail de thèse était de mieux comprendre la multispécificité de transport qui caractérise la P-gp, une protéine phare de la superfamille des transporteurs ABC. Pour atteindre ce but, nous avons utilisé des outils de bioinformatique structurale, modélisation comparative et docking *in silico*, ainsi qu'une approche de génomique comparative qui a été jusqu'à présent peu utilisée dans l'étude des protéines ABC.

Nous nous étions fixé quatre objectifs : (i) rechercher des informations sur la fonction et les mécanismes de reconnaissance de substrats au travers d'une étude de phylogénie des transporteurs ABC humains, (ii) réaliser des modèles 3D de la P-gp humaine, la décrivant dans des états catalytiques différents, sur la base des rares données cristallographiques disponibles dans les ABC, (iii) décrire le site actif, cœur de la reconnaissance multispécifique, et étudier la localisation des résidus impliqués dans le changement de résistance chez la P-gp ; puis positionner ces éléments structuraux (site de reconnaissance, résidus hotspots, et voies d'accès) dans l'environnement membranaire (position dans la bicouche lipidique) et (iv) étudier par le biais de docking *in-silico* la localisation des sous-sites de liaisons de substrats dans la P-gp de deux espèces animales, hamster et humaine, pour réévaluer et mettre à jour le modèle bipharmacophorique établi en 2002.

### **Etude phylogénétique**

Dans ce cadre, j'ai réalisé deux analyses phylogénétiques complémentaires :

- La première analyse a porté sur les isoformes membranaires des transporteurs ABC humains, excluant les membres des sous-familles ABCE et ABCF, pour ne pas diluer l'information d'évolution contenue dans les domaines membranaires. Cette analyse a mis en évidence la forte probabilité que les transporteurs ABC humains soient issus d'un même ancêtre commun en ayant évolué en deux grandes branches. La première branche rassemble les membres des sous-familles ABCA et ABCG, ceux qui ne

possèdent qu'une seule signature distinctive, celle qui est contenue dans toutes les protéines ABC : motifs Walker A, Walker B et signature « LSGGQ ». La seconde branche rassemble les membres des sous-familles ABCB, ABCC et ABCD, des protéines possédant une signature supplémentaire contenue au niveau des domaines TMD. Notre analyse a aussi suggéré que des phénomènes de duplication interne ont marqué l'évolution dans les sous-familles des transporteurs ABC entiers.

- Dans la seconde analyse, j'ai ré-intégré deux séquences protéiques représentant les seuls ABC bactériennes dont les structures sont disponibles dans la PDB : Sav1866 et MsbA. J'ai ainsi mis en évidence un lien évolutif entre ces transporteurs ABC bactériens et la P-gp. Ce lien a révélé que la phylogénie peut être mise au service d'un travail de modélisation comparative. On peut parler ici d'une approche de phylogénie structurale.

### **Etudes de modélisation comparative**

Dans le cadre de ce volet de mon travail doctoral, j'ai pu construire trois modèles 3D de la P-gp humaine la représentant dans trois états catalytiques différents : *Open Inward-Facing*, *Closed Inward-Facing* et *Outward-Facing*. Ces modèles ont été construits pour donner une vision moléculaire du cycle de transport de la P-gp. Grâce à un placement tri-dimensionnel dans une bicouche lipidique, j'ai pu révéler la localisation strictement membranaire des résidus hotspots responsables du changement de résistance chez la P-gp dans les trois états catalytiques putatifs que nous avons décrits, alors que cette information n'était pas disponible dans la littérature. L'utilisation de plusieurs conformations du cycle catalytique nous a permis de donner la première vision moléculaire du devenir de ces sites « hotspots » au cours du cycle catalytique, caractérisé par une très importante trans-conformation. Notre publication dans *J. of Biological Chemistry* est la première à offrir une vision structurale complète du site actif de la P-glycoprotéine réintégrée dans une membrane, et de son évolution morphologique au cours du cycle.

En accord avec les résultats phylogénétiques, un rapprochement structural et fonctionnel a pu être établi entre la P-gp et une protéine de la famille ABCC : SUR1. Des régions homologues et bien spécifiques ont ainsi pu être mises en valeur chez les deux protéines dans leurs structures modélisées les représentant dans l'état catalytique : *Outward-Facing*.

A ce stade de notre étude, il est difficile de discriminer lequel des deux modèles alternatifs de transport, « floppase » ou « aspirateur hydrophobe », peut décrire au mieux le fonctionnement de la P-gp. Nous avons néanmoins pu distinguer, via une étude des canaux d'accès, l'implication des résidus hotspots (majoritairement de la classe exoplasmique) dans tous les tunnels révélés par l'étude, quelle que soit la conformation de la P-gp considérée. Pour faire court, les tunnels et la poche sont tapissés par les résidus hotspots. Il y a donc adéquation satisfaisante, voire remarquable, entre nos modèles structuraux dans les trois états décrits, et les nombreuses données de la littérature sur les mutants fonctionnels de la P-glycoprotéine, symbolisés par les résidus hotspots.

### **Etudes de docking**

Pour compléter cette étude, il était intéressant de revenir à l'interprétation des données *in vitro* obtenues au laboratoire et rassemblées dans la publication à *Molecular Pharmacology* en 2002. Pour cela, il a fallu d'abord recommencer le travail de construction des modèles de la P-glycoprotéine de *Cricetulus griseus* (hamster chinois), selon le protocole publié dans notre article, pour mettre en relation les données fonctionnelles et d'interaction entre substrats obtenues *in vitro*.

Les travaux de docking que j'ai réalisés ensuite sur les modèles 3D de la P-gp de hamster ont porté sur les deux conformations catalytiques jugées les plus fiables (*Open Inward-Facing* et *Outward-Facing*). Nous avons été surpris de voir à quel point les positions de docking obtenues étaient en accord avec les observations et hypothèses émises par l'étude pharmacologique de 2002. Ainsi, les docking de trois substrats représentatifs des pharmacophores analysés dans cette étude (la tentoxine, la vinblastine et le vérapamil), révèlent des sous-sites de liaisons spatialement proches, décrivant une grande poche de liaison, qui a été souvent évoquée dans la littérature pour décrire sa large multispécificité enzymatique, mais jamais décrite sur le plan moléculaire. Même les interactions mutuelles révélées par l'étude *in vitro* sont, de manière surprenante, confirmées par les résultats du docking semi-flexible.

La confrontation de ces résultats de docking avec ceux obtenus sur les modèles 3D de la P-gp humaine, nous a permis de mettre en évidence un possible mécanisme général d'expulsion des ligands pouvant être adopté par la P-gp indépendamment de l'espèce. La grande poche de liaison de reconnaissance de substrats existe dans les deux espèces étudiées (hamster et homme), avec des variations spécifiques des résidus tapissant les voies d'accès ou

la poche multispécifique, ce qui implique également que les interactions mutuelles entre substrats peuvent varier d'une espèce à une autre.

Enfin, à travers ces travaux de docking, l'importance des résidus hotspots, sur l'ensemble du cycle de translocation de substrat, a été une nouvelle fois confirmée.

## II. Perspectives

Les résultats décrits dans ce manuscrit représentent une avancée encourageante dans le cadre de l'étude d'une protéine d'un grand intérêt en chimiothérapie. Nous n'avons bien sûr pas résolu toutes les questions posées sur les mécanismes de reconnaissance et de transport de la P-gp mais un certain nombre d'hypothèses ont pu être confirmées et de nouvelles ont pu être émises. La suite de ce travail pourrait porter sur :

- Le docking des autres substrats du pharmacophore I : la ciclosporine, l'actinomycine D, la bromocriptine et la pristinamycine IA dans le but de confirmer les hypothèses émises par l'étude pharmacophorique de 2002.
- Un criblage virtuel de petites molécules avec comme but l'étude d'autres éventuels sous-sites de liaison, probables, étant donné la taille de la poche de liaison.
- Une étude de dynamique moléculaire, en solvant et membrane explicites, pour étudier plus finement les étapes constituant le cycle de translocation de la P-gp, voire simuler les transitions structurales du transporteur associées aux différents états fonctionnels de la P-gp dans son contexte physiologique. D'autres méthodes (dynamique essentielle, modes normaux, dynamique hors équilibre sous contrainte de coordonnée réactionnelle, ..) permettraient sans doute de mener ces études théoriques sans être confronté à la limite de puissance calcul.
- la région linker dans les protéines ABC, dont le rôle est mal connu. Cette question est généralement peu abordée car la région est absente des structures résolues par cristallographie des Rayons X, et pourtant elle doit significativement contribuer à la mécanique globale du transporteur. Est ce qu'un enroulement superhélicoïdal peut être détecté dans d'autres protéines ABC comme cela a été le cas pour la P-gp et SUR1 ? Cette région pourrait être caractérisée par une étude conjointe *in vitro* (mutants de l'heptade répétée) et *in silico*.

# Annexes

# Figure-Ax-1

Protein: P-glycoprotein

Length: 1280

N-terminus: IN

Number of transmembrane helices: 12

Transmembrane helices: 49-73 117-141 190-209 214-233 299-323 328-352 710-733 754-778 833-852 857-874 935-957 974-996

Total entropy of the model: 17.0185

Entropy of the best path: 17.0243

The best path:

```
seq MDLEGDRNGG AKKKNFFKLN NKSEKDKKEK KPTVSVFMSF RYSNWLDKLY 50
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIiiiiiii iiiiiiiiiHH

seq MVVGTLA AII HGAGLPLMML VFGEMTDIFA NAGNLEDLMS NITNRS DIND 100
pred HHHHHHHHHH HHHHHHHHHH HHHoooooooo oooooooooO OOOOOOOOOO

seq TGFFMNLEED MTRYAYYYSG IGAGVLVAAY IQVSFWCLAA GRQIHKIRKQ 150
pred Oooooooooo ooooooHHHH HHHHHHHHHH HHHHHHHHHH Hiiiiiiiiii

seq FFHAIMRQEI GWFDVHDVGE LNTRLTDDVS KINEGIGDKI GMFFQSMATF 200
pred iiiiiiIIII IIIIIIIIII IIIIIiiiiii iiiiiiiiiIH HHHHHHHHHH

seq FTGFIVGFTR GWKLTLVILA ISPVLGLSAA VWAKILSSFT DKELLAYAKA 250
pred HHHHHHHHHH oooHHHHHHH HHHHHHHHHH HHHiiiiiii iiiiiiiiiII

seq GAVAEVLA A IRTVIAFGGQ KKELE RYNKN LEEAKRIGIK KAITANISIG 300
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIiiiiiii iiiiiiiiiHH

seq AAFLLIYASY ALAFWYGTTL VLSGEYSIGQ VLTVFFSVLI GAFSVGQASP 350
pred HHHHHHHHHH HHHHHHHHHH HHHoooooooo HHHHHHHHHH HHHHHHHHHH

seq SIEAFANARG AAYEIFKIID NKPSIDSYSK SGHKPDNIKG NLEFRNVHFS 400
pred HHiiiiiiii iiiiiiiiii IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq YPSRKEVKIL KGLNLKVQSG QTVALVGN SG CGKSTTVQLM QRLYDPTEGM 450
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq VSDGQDIRT INVRFLREII GVVSQEPVLF ATTIAENIRY GRENVMTDEI 500
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq EKAVKEANAY DFIMKLPKHF DTLVGERGAQ LSGGQKQRIA IARALVRNPK 550
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq ILLLDEATSA LDTESEAVVQ VALDKARKGR TTIVIAHRLS TVRNADV IAG 600
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq FDDGVIVEKG NHDELMKEKG IYFKLVTMQT AGNEVELENA ADESKSEIDA 650
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq LEMSSNDSRS SLIRKRSTRR SVRGSQAQDR KLSTKEALDE SIPPVSEFWRI 700
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIiiiiiii

seq MKLNLTEWPY FVVGVFCAII NGGLQPAFAI IFSKIIGVFT RIDDPETKRQ 750
pred iiiiiiiiiH HHHHHHHHHH HHHHHHHHHH HHHoooooooo oooooooooo
```

```

seq  NSNLFSLFL  ALGIISFITF  FLQGFTFGKA  GEILTKRLRY  MVFRSMLRQD  800
pred  oooHHHHHHH  HHHHHHHHHH  HHHHHHHHHH  iiiiiiiiii  iiiIIIIIII

seq  VSWFDDPKNT  TGALTTRLAN  DAAQVKGAIG  SRLAVITQNI  ANLGTGIIIS  850
pred  IIIIIIIIII  IIIIIIIIII  iiiiiiiiii  iiHHHHHHHH  HHHHHHHHHH

seq  FIYGWQLTLL  LLAIVPIIAI  AGVVEMKMLS  GQALKDKKEL  EGSGKIATEA  900
pred  HHooooHHHH  HHHHHHHHHH  HHHHiiiiii  iiiiiiiiii  IIIIIIIIII

seq  IENFRTVVSL  TQEQKFEHMY  AQSLQVPYRN  SLRKAHIFGI  TFSFTQAMMY  950
pred  IIIIIIIIII  IIIIIIIIII  iiiiiiiiii  iiiHHHHHHH  HHHHHHHHHH

seq  FSYAGCFRFG  AYLVAHKLMS  FEDVLLVFSA  VVFGAMAVGQ  VSSFAPDYAK  1000
pred  HHHHHHHooo  oooooooooo  ooHHHHHHHH  HHHHHHHHHH  HHHHHHiiii

seq  AKISAAHIIM  IIEKTPLIDS  YSTEGLMPNT  LEGNVTFGEV  VFNYPTRPDI  1050
pred  iiiiiiiiii  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

seq  PVLQGLSLEV  KKGQTLALVG  SSGCGKSTVV  QLLERFYDPL  AGKVLLDGKE  1100
pred  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

seq  IKRLNVQWLR  AHLGIVSQEP  ILFDCSIAEN  IAYGDNSRVV  SQEEIVRAAK  1150
pred  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

seq  EANIHFIES  LPNKYSTKVG  DKGTQLSGGQ  KQRIAIARAL  VRQPHILLLD  1200
pred  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

seq  EATSALDTES  EKVVQEALDK  AREGRTCIVI  AHRLSTIQNA  DLIVVFQNGR  1250
pred  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

seq  VKEHGTHQQL  LAQKGIYFSM  VSVQAGTKRQ  1280
pred  IIIIIIIIII  IIIIIIIIII  IIIIIIIIII

```

Protein: SUR1

Length: 1581

N-terminus: OUT

Number of transmembrane helices: 17

Transmembrane helices: 37-54 75-94 107-124 135-152 161-185 300-324 351-368

399-423 450-473 536-560 575-594 1010-1034 1061-1085 1104-1121 1148-1172

1253-1270 1279-1298

Total entropy of the model: 17.0268

Entropy of the best path: 17.0333

The best path:

```

seq  MPLAFCGSEN  HSAAYRVDQG  VLNNGCFVDA  LNVVPHVFL  FITFPILFIG  50
pred  OOOOOOOOOO  OOOOOOOOOO  Oooooooooo  ooooooHHHH  HHHHHHHHHH

seq  WGSQSSKVHI  HHSTWLHFP  HNLRWILTFM  LLFVLVCEIA  EGILSDGVTE  100
pred  HHHHiiiiii  iiiiiiiiii  iiiHHHHHHH  HHHHHHHHHH  HHHHoooooo

seq  SHHLHLYMPA  GMAFMAAVTS  VVYHNIETS  NFPKLLIALL  VYWTLAFITK  150
pred  ooooooHHHH  HHHHHHHHHH  HHHHiiiiii  iiiHHHHHHH  HHHHHHHHHH

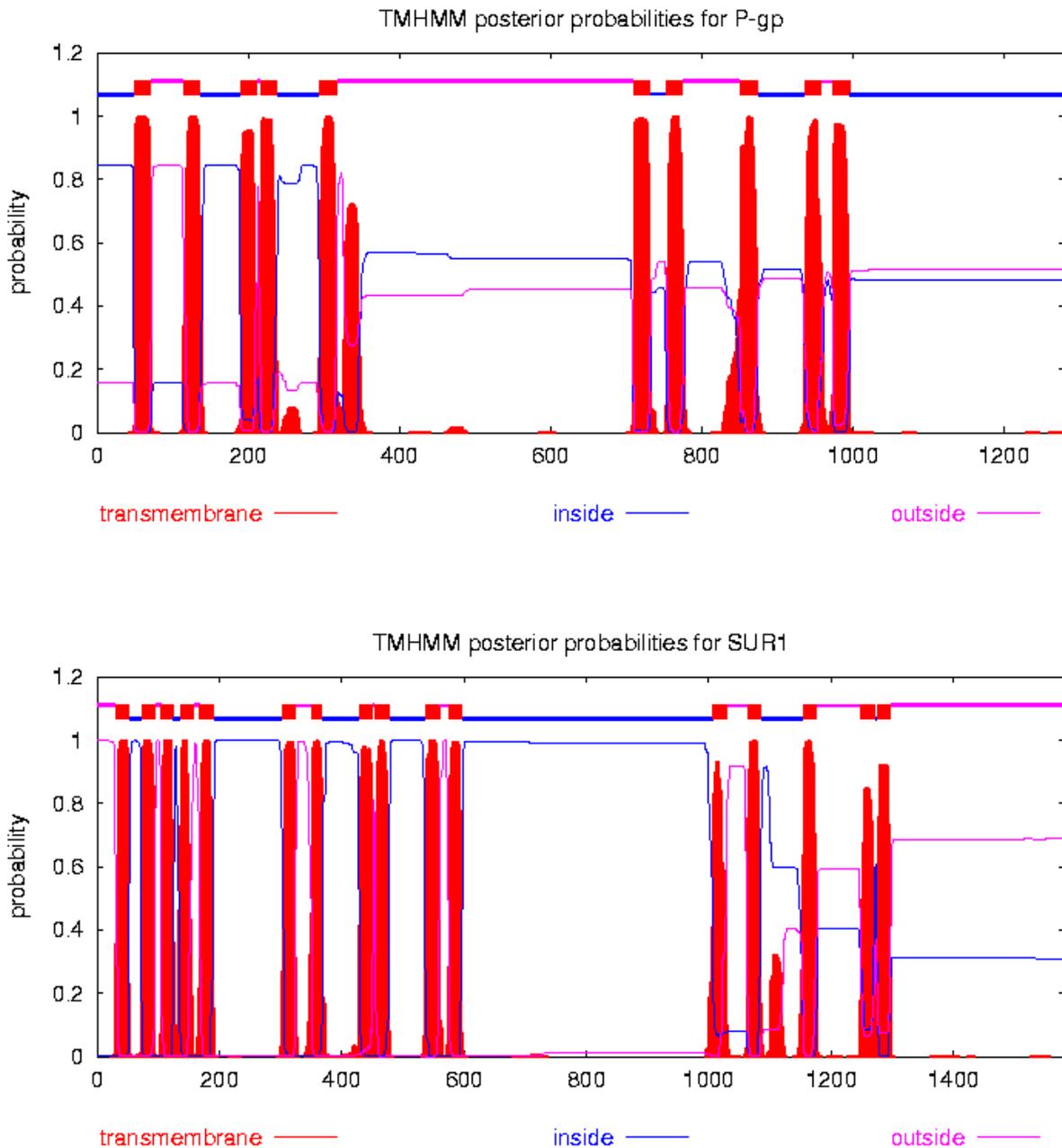
seq  TIKFVKFLDH  AIGFSQLRFC  LTGLLVILYG  MLLLVEVNVI  RVRRYIFFKT  200
pred  HHoooooo  HHHHHHHHHH  HHHHHHHHHH  HHHHHiiiiii  iiiiiiiiii

```

seq	PREVKPPEDL	QDLGVRFLQP	FVNLLSKGTY	WWMNAFIKTA	HKKPIDLRAI	250
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	GKLP IAMRAL	TNYQRLCEAF	DAQVRKDIQG	TQGARAIWQA	LSHAFGRRLV	300
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIiiiiii	iiiiiiiiiiH	
seq	LSSTFRILAD	LLGFAGPLCI	FGIVDHLGKE	NDVFQPKTQF	LGVYFVSSQE	350
pred	HHHHHHHHHH	HHHHHHHHHH	HHHHoooooooo	oooooooooooo	oooooooooooo	
seq	FLANAYVLAV	LLFLALLLQR	TFLQASYYYVA	IETGINLRGA	IQTKIYNKIM	400
pred	HHHHHHHHHH	HHHHHHHHHi	iiiiiiiiiii	iiiiiiiiiii	iiiiiiiiiiH	
seq	HLSTSNLSMG	EMTAGQICNL	VAIDTNQLMW	FFF LCPNLWA	MPVQIIVGVI	450
pred	HHHHHHHHHH	HHHHHHHHHH	HHHoooooooo	oooooooooooo	ooooooooooooH	
seq	LLYYILGVSA	LIGAAVIILL	APVQYFVATK	LSQAQRSTLE	YSNERLKQTN	500
pred	HHHHHHHHHH	HHHHHHHHHH	HHHiiiiiii	iiiiiiiiiii	IIIIIIIIII	
seq	EMLRGIKLLK	LYAWENIFRT	RVETTRRKEM	TSLRAFAIYT	SISIFMNTAI	550
pred	IIIIIIIIII	IIIIIIIIII	iiiiiiiiiii	iiiiiiHHHHH	HHHHHHHHHH	
seq	PIAAVLITFV	GHVSFFKEAD	FSPSVAFASL	SLFHILVTPL	FLLSSVVRST	600
pred	HHHHHHHHHH	oooooooooooo	ooooHHHHHHH	HHHHHHHHHH	HHHHiiiiiii	
seq	VKALVSVQKL	SEFLSSAEIR	EEQCAPHEPT	PQGPASKYQA	VPLRVVNRKR	650
pred	iiiiiiiiiii	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	PAREDCRGLT	GPLQSLVPSA	DGDADNCCVQ	IMGGYFTWTP	DGIPTLSNIT	700
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	IRIPRGQLTM	IVGQVCGKKS	SLLLAALGEM	QKVGSAVFW	SLPDSEIGED	750
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	PSPERETATD	LDIRKRG PVA	YASQKPWLLN	ATVEENIIFE	SPFNKQRYKM	800
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	VIEACSLQPD	IDILPHGDQT	QIGERGINLS	GGQRQRISVA	RALYQHANVV	850
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	FLDDPFSALD	IHLSDHLMQA	GILELLRDDK	RTVVVLVTHKL	QYLPHADWII	900
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	AMKDGTIQRE	GTLKDFORSE	CQLFEHWKTL	MNRQDQELEK	ETVTERKATE	950
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	
seq	PPQGLSRAMS	SRDGLLQDEE	EEEEEEAAESE	EDDNLSSMLH	QRAEIPWRAC	1000
pred	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIIIIIII	IIIIiiiiiii	
seq	AKYLSSAGIL	LLSLLVFSQL	LKHMVLVAID	YWLAKWTDSA	LTLTPAARNC	1050
pred	iiiiiiiiiiH	HHHHHHHHHH	HHHHHHHHHH	HHHHoooooooo	oooooooooooo	
seq	SLSQECTLDQ	TVYAMVFTVL	CSLGIVLCLV	TSVTVEWTGL	KVAKRLHRSL	1100
pred	oooooooooooo	HHHHHHHHHH	HHHHHHHHHH	HHHHiiiiiii	iiiiiiiiiii	
seq	LNRIILAPMR	FFETTPLGSI	LNRFSSDCNT	IDQHIPSTLE	CLSRSTLLCV	1150
pred	iiiHHHHHHH	HHHHHHHHHH	Hoooooooooooo	oooooooooooo	ooooooooHHH	
seq	SALAVISYVT	PVFLVALLPL	AIVCYFIQKY	FRVASRDLQQ	LDDTTQLPLL	1200
pred	HHHHHHHHHH	HHHHHHHHHH	HHiiiiiii	iiiiiiiiiii	IIIIIIIIII	
seq	SHFAETVEGL	TTIRAFRYEA	RFQQKLELYT	DSNNIASLFL	TAANRWLEVR	1250



## Figure-Ax-2



**Figure-Ax-2 : Les profils de prédiction des topologies membranaires de la P-gp et de SUR1 issus de TMHMM. Les hélices transmembranaires sont présentées en rouge, les régions intracellulaires en bleu et les régions extracellulaires en rose. 10 hélices transmembranaires sont prédites avec une probabilité  $> 0.8$  dans le cas de la P-gp. 16 hélices transmembranaires sont prédites avec une probabilité  $> 0.8$  dans le cas de SUR1**





Pred: HHHHCCCCCCCCEEEECCCCCCCCCHHHHHHCCEEECCCCCCCCCHHHHHCCCCCCC  
 AA: QLLERFYDPLAGKVLLDGGKEIKRLNVQWLRHLGIVSQEPILFDCSIAENIAYGDNRSRVV  
                   1090                  1100                  1110                  1120                  1130  
 1140

Conf: 99999999998040588750998555654788878892588899999984199722211  
 Pred: CHHHHHHHHHHCCCHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCEEEEC  
 AA: SQEEIVRAAKEANIHFIEISLPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLD  
                   1150                  1160                  1170                  1180                  1190  
 1200

Conf: 666777733399999999998169909985012000016968999859869670481688  
 Pred: CCCCCCHHHHHHHHHHHHHHCCCEEEEECCCHHHCCCEEEEEECCEEEEECCCHHHH  
 AA: EATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQQL  
                   1210                  1220                  1230                  1240                  1250  
 1260

Conf: 62799139999886410179  
 Pred: HHCCCCHHHHHHHHHHCCCC  
 AA: LAQKGIYFSMVSQAGTKRQ  
                   1270                  1280

**Figure-Ax-3 : Fichier de sortie du programme PSIPRED représentant le résultat de prédiction des structures secondaires de la P-gp.** *Conf* : le degré de confiance d'exactitude de la prédiction dans une échelle allant de 0 à 9. *Pred* : la structure secondaire prédite pour chaque acide amine avec C pour Coil, H pour Helix, E pour Extended. AA pour la séquence. La région-linker est représentée en jaune

## Figure-Ax-4

```
>P1;3B60A
structureX:3B60A:10:A:582:::::
-----WQTFRRLWPTIAPFKA
GLIVAGIALIILNAASDTFMLSLLKPLDDGFG-----KTDRS
VLLWMPPLVVIGLMILRGITSYISSYICISWVSGKVMTMRRRLFHMMGMPVAFFD--KQS
TGTLLSRITYDSEQVASSSSGALITVVREGASIIGLFIMMFYYSWQLSIIILVVLAPIVSI
AIRVVSKRFRSISKNMQNTMGQVTTSAEQMLKGHKEVLIFGGQEVETKRFDKVSNMRLQ
GMKMSVASSISDPIIQLIASLALAFVLYAASFPSVMDSLTAGTITVVFSSMIALMRPLKS
LTNVNAQFQRGMAACQTLFAILDSEQEKDE---GKRVIDRATGDLEFRNVFTFTYPGR-EV
PALRNINLKIIPAGKTVALVGRSGSGKSTIASLITRFYDIDEGHILMDGHDREYTLASLR
NQVALVSQNVHLFNDTVANNIAYARTEEYSREQ-IEEAARMAYAMDFINKMDNGLDTIIG
ENGVLLSGGQQRARIAARALLRDSPIILILDEATSALDTESERAIQAALDELQKNRTSLVI
AHLRSTIEQADEIVVVEDGIIVERGTHSELLAQHGVIYAQLHKMQFG----*
>P1;2hydA
structureX:2hydA:1:A:578:::::
-----MIKRYLQFVKPYKY
RIFATIIIVGIIKFGIPMLIPLLIKYAIDGVINHA-----LTTDEKVHH
LTIAIGIALFIFVIVRPPIEFIRQYLAQWTSNKILYDIRKKLYNHLQALSARFYA--NNQ
VGQVISRVINDVEQTKDFILTGLMNIWLDCITIIIALSIMFFLDVKLTLAALFIFPFYIL
TVYVFFGRLRKLTRERSQALAEVQGFHERVQGISVVKSF AIEDNEAKNFDDKNTNFLTR
ALKHTRWNAYSFAAINTVTDIGPIIIVIGVAYLAISGSITVGTAAAFVGYLELLFGPLRR
LVASFTTLTQSFASMDRVFQLIDEDYDIKNG-VGAQPIEIKQGRIDIDHVSFQYNDN-EA
PILKDINLSIEKGETVAFVGMSSGGKSTLINLIPRFYDVTSGQILIDGHNIKDFLTGSLR
NQIGLVQQDNILFSDTVKENILLG-RPTATDEE-VVEAAKMANAHDFIMNLPQGYDTEVG
ERGVKLSGGQKQRLSIARIFLNNPPIILILDEATSALDLESESI IQEALDVLSKDRTTLIV
AHLRSTITHADKIVVIENGHIVETGTHRELIQKQAYEHLYSIQNL-----*
>P1;1P-gp_Hum
sequence:1P-gp_Hum:43: :630:::::
-----SNWLDKLYMVVGTAA
IIHGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRS DINDT---GFFMNLEEDMTR--
----YAYYYSGIGAGVLVAAYIQVSWFCLAAGRQIHKIRKQFFHAIMRQEIGWFD--VHD
VGELNTRLTDDVSKINEGIGDKIGMFFQSMATFFTGFIVGFTRGWKLTTLVILAI SPVLGL
SAAVWAKILSSFTDKELLAYAKAGAVAEVLA AIRTVIAFGGQKKE LERYNKNLEEAKRI
GIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTVFFSVLIGAFSVGQ
ASPSIEAFANARGAAYEIFKIIDNKPSIDSYKSGHKPDNIKGNLEFRNVHFSYPSRKEV
KILKGLNLKVQSGQTVLVGNSSGCGKSTTVQLMQRLYDPT EGMVSVDDGQDIRTINVRFLR
EIIIGVVSQEPVLFATTIAENIRYGREN--VTMDEIEKAVKEANAYDFIMKLP HKFDTLVG
ERGAQLSGGQKQRIARIAARALVRNPKILLLDEATSALDTESEAVVQVALDKARKGRTTIVI
AHLRSTVRNADVIAGFDDGVIVEKGNHDELMKEGIYFKLVMTMQT-----*
```

**Figure-Ax-4: Exemple d'un alignement format .ali soumis à Modeller.** Alignement de la moitié N-terminale de la P-gp humaine à reconstruire contre deux séquences templates des monomères de St\_MsbA et de Sav1866

## Figure-Ax-5

```
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

env.io.atom_files_directory = ['.', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hetatm = True

# Build 10 models, and assess with both DOPE and GA341
a = loopmodel(env, alnfile = '1P-gp_Hum.ali', knowns = ('3B60A', '2hydA'),
              sequence = '1P-gp_Hum', assess_methods=(assess.DOPE, assess.GA341))
a.starting_model= 1
a.ending_model  = 10
a.make()

# Get a list of all successfully built models from a.outputs
ok_models = filter(lambda x: x['failure'] is None, a.outputs)

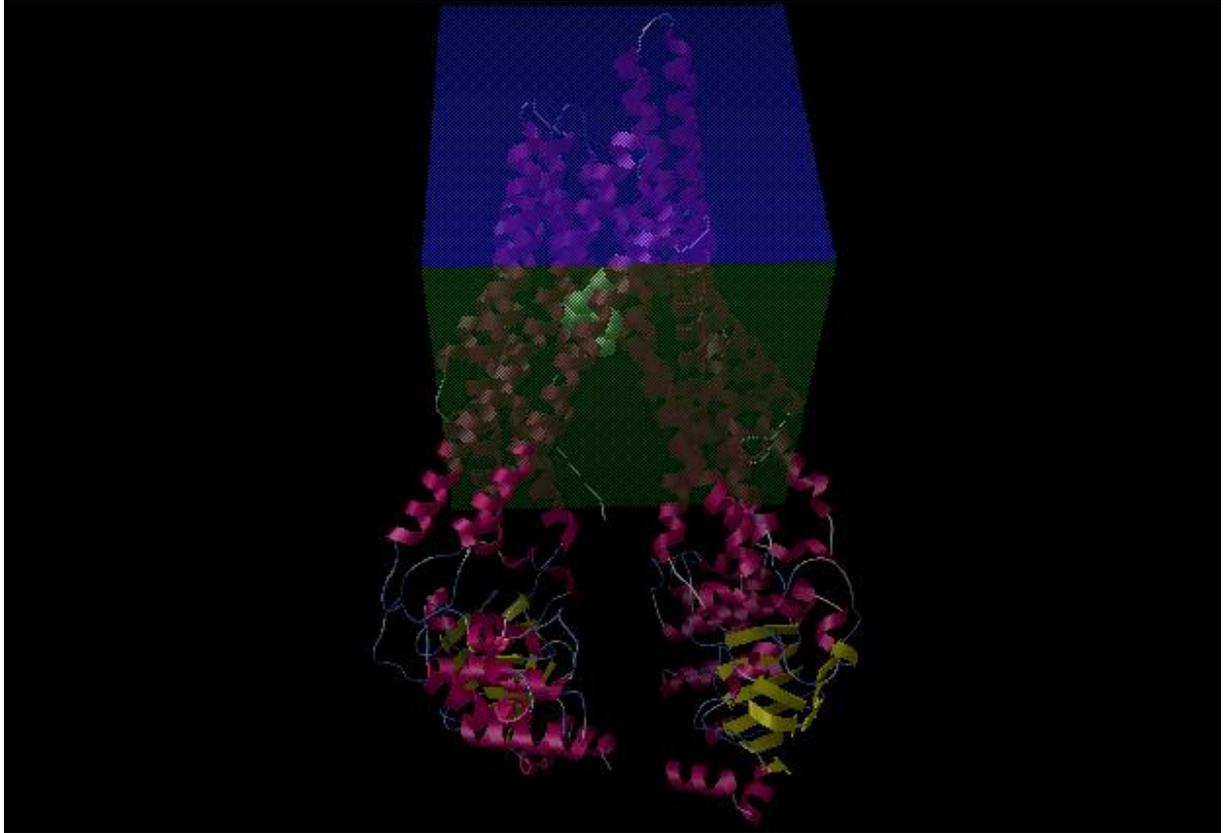
# Thorough MD optimization:
a.md_level = refine.slow

# Rank the models by DOPE score
key = 'DOPE score'
ok_models.sort(lambda a,b: cmp(a[key], b[key]))

# Get top model
m = ok_models[0]
print "Top model: %s (DOPE score %.3f)" % (m['name'], m[key])
```

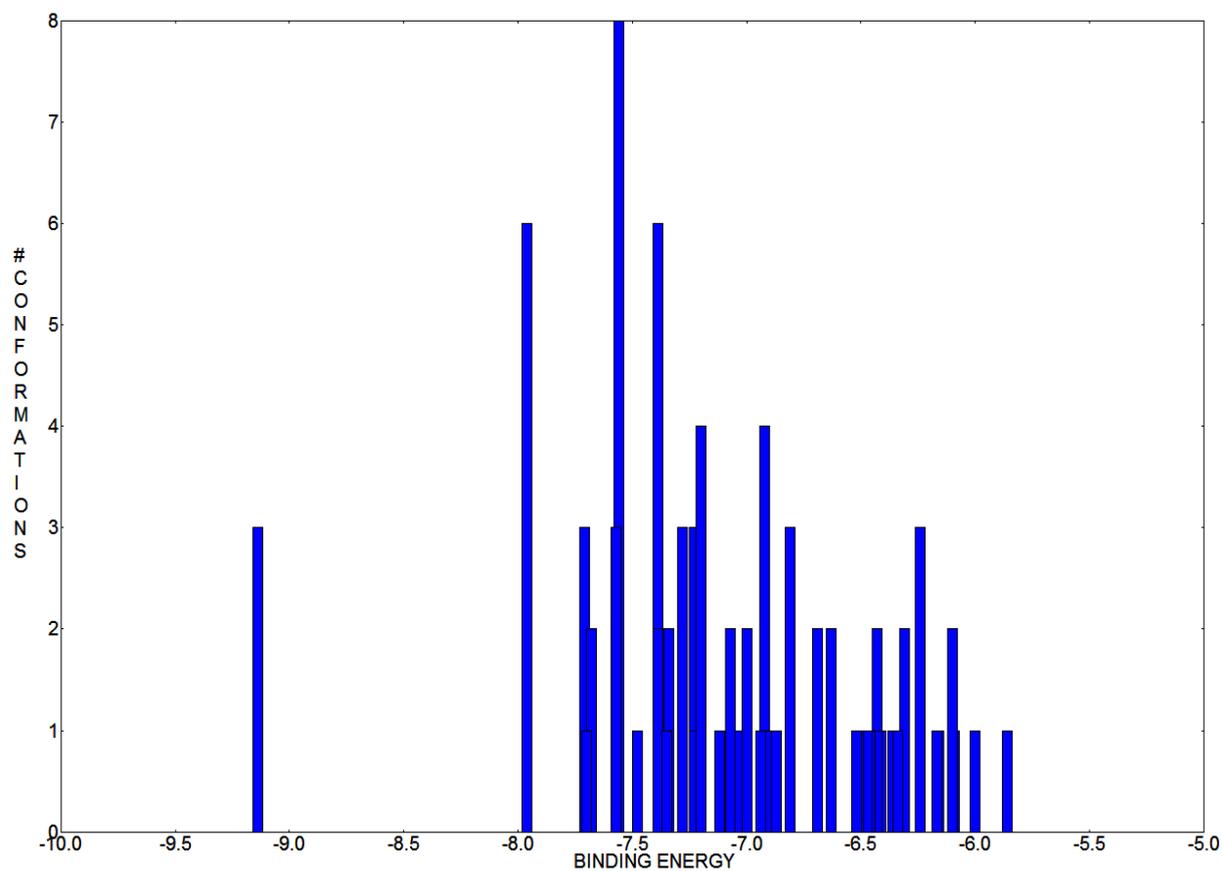
Figure-Ax-5 : Exemple de script modeller ayant servi à la génération des modèles 3D

## Figure-Ax-6



**Figure-Ax-6** : Exemple type d'une grille d'exploration d'Autodock « couvrant » un modèle de la P-gp de hamster dans la conformation *Open Inward-Facing*. Grille en forme de parallélépipède générée par AutoGrid et représentant l'espace de recherche exploré pendant l'étape de docking.

## Figure-Ax-7



**Figure-Ax-7 :** Histogramme du docking de la tentoxine sur la P-gp de hamster dans sa conformation *Open Inward-Facing*. Le meilleur cluster des poses de docking de la tentoxine correspond à un cluster de trois poses de docking quasi identiques. Il faut noter que les énergies de liaison indiquées sur l'axe des abscisses sont celles de la pose au plus faible score énergétique composant le cluster.

Sub-family	Member	Proteins (a.a)	ABC transporter_2 (a.a)		Acc. Nb (SwissProt)	Synonyms (Swissprot//TCDB)	TCDB Classification
ABCA	ABCA1	2261	N	899-1131	O95477	ABC-1, Cholesterol efflux regulatory protein, Cholesterol/phospholipid flippase	3.A.1.211.1
			C	1912-2144			
	ABCA2	2435	N	990-1221	Q9BZC7	ABC2, Multidrug resistance pump	3.A.1.211.3
			C	2050-2285			
	ABCA3	1704	N	530-763	Q99758	ABC-C	3.A.1.211.1
			C	1381-1614			
	ABCA4	2273	N	929-1160	P78363	RIM ABC transporter, RIM protein, RmP, Stargardt disease protein, The retinal-specific ABC transporter	3.A.1.211.2
			C	1938-2170			
	ABCA5	1642	N	478-713	Q8WWZ7	ATP-binding cassette sub-family A member 5	3.A.1.211.1
			C	1290-1533			
	ABCA6	1617	N	478-713	Q8N139	ATP-binding cassette sub-family A member 6	3.A.1.211.1
			C	1288-1513			
	ABCA7	2146	N	807-1038	Q8IZY2	ATP-binding cassette sub-family A member 7 (Macrophage ABC transporter) (Autoantigen SS-N) (ABCA-SSN)	3.A.1.211.1
			C	1793-2025			
	ABCA9	1524	N	481-716	Q8IUA7	None	3.A.1.211.3
			C	1288-1521			
	ABCA10	1543	N	391-626	Q8WWZ4	ATP-binding cassette sub-family A member 10	3.A.1.211.1
			C	1206-1440			
	ABCA12	2595	N	1346-1577	Q86UK0	ATP-binding cassette 12	3.A.1.211.1
			C	2254-2489			
	ABCA13	5058	N	3842-4074	Q86UQ4	ATP-binding cassette sub-family A member 13	Not present
			C	4718-4956			
ABCB	ABCB1	1280	N	392-628	P08183	Multidrug resistance protein 1 (MDR1), P-glycoprotein 1, Broad specificity multidrug resistance (MDR) efflux pump	3.A.1.201.1
			C	1035-1273			
	ABCB2	808		563-802	Q03518	APT1, Peptide transporter TAP1, Peptide transporter PSF1, Peptide supply factor 1, PSF-1, MHC heterodimeric peptide exporter	3.A.1.209.1
	ABCB3	686		468-686	Q03519	APT2, Peptide transporter TAP2, Peptide transporter PSF2, Peptide supply factor 2, PSF-2, MHC heterodimeric peptide exporter	3.A.1.209.1
	ABCB4	1286	N	394-630	P21439	Multidrug resistance protein 3 (MDR3), P-glycoprotein 3, Short chain fatty acid phosphatidylcholine translocase	3.A.1.201.3

			C	1034-1279			
	ABCB5	812	N	2-177	Q2M3G0	ATP-binding cassette sub-family B member 5	
			C	570-808			
	ABCB6	842		590-824	Q9NP58	Mitochondrial ABC transporter 3, ABC transporter umat, Iron transporter	3.A.1.210.6
	ABCB7	752		472-706	O75027	ABC transporter 7 protein (mitochondrial), ABC7 iron transporter, (X-linked sideroblastis anemia protein)	3.A.1.210.4
	ABCB8	735		472-709	Q9NUT2	Mitochondrial ATP-binding cassette 1, M-ABC1	3.A.1.201.1
	ABCB9	766		504-740	Q9NP78	TAP-like protein, TAPL, hABCB9	3.A.1.209.1
	ABCB10	738		492-731	Q9NRK6	Mitochondrial ATP-binding cassette 2, M-ABC2	3.A.1.212.1
	ABCB11	1321	N	420-656	O95342	Bile salt export pump, BSEP or SPGP	3.A.1.201.2
			C	1078-1316			
<b>ABCC</b>	ABCC1	1531	N	644-868	P33527	MRP1, Leukotriene C(4) transporter, LTC4 transporter, Drug resistance pump	3.A.1.208.8
			C	1293-1527			
	ABCC2	1545	N	637-861	Q92887	Canalicular multispecific organic anion transporter 1, MRP2, Canalicular multidrug resistance protein, Hepatic canalicular conjugate exporter	3.A.1.208.2
			C	1300-1534			
	ABCC3	1527	N	629-851	O15438	Canalicular multispecific organic anion transporter 2, MRP3, Multi-specific organic anion transporter-D, MOAT-D	3.A.1.208.9
			C	1291-1523			
	ABCC4	1325	N	410-633	O15439	MRP4, Multi-specific organic anion transporter-B, MOAT-B	3.A.1.208.7
			C	1041-1274			
	ABCC5	1437	N	562-783	O15440	MRP5, Multi-specific organic anion transporter-C, MOAT-C, pABC11, SMRP	3.A.1.208.13
			C	1193-1427			
	ABCC6	1503	N	629-853	O95255	MRP6, Anthracycline resistance-associated protein, Multi-specific organic anion transporter-E, MOAT-E, Multidrug (anthracycline) resistance organic anion efflux pump, The pseudoxanthoma elasticum disease protein)	3.A.1.208.10
			C	1265-1499			
	ABCC7	1480	N	423-646	P13569	Cystic fibrosis transmembrane conductance regulator, CFTR, cAMP-dependent chloride channel	3.A.1.202.1
			C	1210-1443			
	ABCC8	1581	N	679-929	Q09428	Sulfonylurea receptor 1, SUR1	3.A.1.208.4
			C	1344-1578			
	ABCC9	1549	N	672-912	O60706	Sulfonylurea receptor 2, SUR2	3.A.1.208.4
			C	1312-1546			
	ABCC10	1492	N	598-824	Q5T3U5	MRP10	3.A.1.208.9
			C	1246-1479			
	ABCC11	1382	N	508-732	Q96J66	MRP8, Cyclic nucleotide (cAMP and cGMP) efflux pump	3.A.1.208.13
			C	1141-1375			

	ABCC12	1359	N	467-701	Q96J65	Multidrug resistance-associated protein 9	3.A.1.208.13
			C	1120-1354			
<b>ABCD</b>	ABCD1	745		474-700	P33897	Adrenoleukodystrophy protein, ALDP, Long chain fatty acid (LCFA) transporter	3.A.1.203.3
	ABCD2	740		478-704	Q9UBJ2	Adrenoleukodystrophy-related protein., hALDR, Adrenoleukodystrophy-like 1	3.A.1.203.3
	ABCD3	659		440-659	P28288	70 kDa peroxisomal membrane protein, PMP70, Peroxisomal long chain fatty acyl (LCFA) transporter	3.A.1.203.1
	ABCD4	606		389-603	O14678	Peroxisomal membrane protein 69, PMP69, Peroxisomal membrane protein 1-like, PXMP1-L, P70R	3.A.1.203.1
<b>ABCG</b>	ABCG1	678	N	77-317	P45844	White protein homolog, ATP-binding cassette transporter 8	3.A.1.204.1
			C				
	ABCG2	655	N	37-286	Q9UNQ0	Placenta-specific ABC transporter, Breast cancer resistance protein, Mitoxantrone resistance-associated protein, CDw338 antigen	3.A.1.204.2
			C				
	ABCG4	646	N	61-301	Q9H172	WHITE2	3.A.1.204.1
			C				
	ABCG5	651	N	52-293	Q9H222	Sterolin-1	3.A.1.204.5
			C				
	ABCG8	673	N	47-313	Q9H221	Sterolin-2	3.A.1.204.5
			C				
<b>ABC_Bac</b>							
	Sav1866	578		340-575	Q99T13	Putative multidrug export ATP-binding/permease protein SAV1866	3A.1.106.2
	MsbA	582		342-578	P63359	Lipid A export ATP-binding/permease protein MsbA	3.A.1.106.1

**Tableau-Ax-1: Transporteurs ABC humains et bactériens.** Avec leurs sequences d'acides amines sélectionnés pour les analyses phylogénétiques, leurs codes d'accèsions Swiss-Prot, leurs alias ainsi que leurs numéros de classification TCDB. Pour les transporteurs entiers : N pour N-terminal et C pour C-terminal. ABC\_Bac : Les transporteurs ABC bactériens ayant intégré l'étude phylogénétique.

## Tableau-Ax-2

	TMD1/TMD	TMD2	NBD1/NBD	NBD2
P-gp_Hum	51-357	711-1000	392-628	1035-1273
P-gp_Mouse	50-356	709-998	391-628	1033-1271
SUR1	299-602	1012-1306	679-929	1344-1578
Sav1866	16-306	-	340-575	-
St-MsbA	27-310	-	342-578	-
Ec-MsbA	27-310	-	342-578	-
Vc-MsbA	27-310	-	342-578	-

**Tableau-Ax-2 : Les résidus représentant les domaines TMD et NBD sélectionnés pour les calculs des scores d'alignements présentés dans le Tableau-C-1.** *Les TMD et les NBD des monomères bactériens sont respectivement présentés dans les mêmes colonnes que les TMD1, NBD1 dans protéines ABC entières. P-gp\_Hum : la P-gp humaine, Pgp\_Mouse : la P-gp de souris, Sa : Staphylococcus aureus, Ec : Escherichia coli, Vc : Vibrio cholerae, St : Salmonella typhimurium*

## Tableau-Ax-3

	P-gp_Hum	SUR1	Sav1866	Ec-MsbA	Vc-MsbA	St-MsbA	P-gp_Mouse
P-gp_Hum	100						
SUR1	15	100					
Sav1866	<b>28</b>	<b>20</b>	100				
Ec-MsbA	<b>31</b>	21	32	100			
Vc-MsbA	<b>28</b>	22	33	67	100		
St-MsbA	<b>31</b>	<b>22</b>	33	96	67	100	
Pgp_Mouse	<b>87</b>	<b>16</b>	28	29	28	29	100

**Tableau-Ax-3 : Les scores d'alignement des séquences protéiques entières de protéines ABC.** Les scores ont été obtenus grâce au programme d'alignement multiple ClustalW [10]. Sont présentés en rouge, les scores d'alignement des domaines représentatifs des templates et ceux représentatifs des modèles à construire. P-gp\_Hum : la P-gp humaine, Pgp\_Mouse : la P-gp de souris, Sa : Staphylococcus aureus, Ec : Escherichia coli, Vc : Vibrio cholerae, St : Salmonella typhimurium