DISSERTATION

Titel der Dissertation

Molecular Basis of Drug-P-glycoprotein Interaction

Verfasserin

Dipl.-Ing. (FH) Freya Klepsch

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien, 2012

Acknowledgments

First of all I want to thank my supervisor, Prof. Gerhard F. Ecker, who gave me this great possibility to do a PhD in his group at the University of Vienna. Gerhard, I really appreciate all your support and encouraging words during the last years. Personally, as well as scientifically, I hold you in very high esteem. In my feeling, we had a very good connection, which is reflected in the quality of this work. You have been a great supervisor.

Furthermore, I want to express my gratitude to all the collaborators, I had the possibilty to work with. I was lucky to cooperate with Prof. Peter Chiba from the Medical University of Vienna, whose knowledge about ABC transporters, and P-gp in particular, impressed me over and over again.
Additionally, I want to thank Dr. Thomas Stockner for his support regarding the molecular dynamics simulation experiments.
Very special thanks go to Prof. Daan Geerke from the VU University, Amsterdam, who is the most considerate and helpful person I have ever met. Thank you for being so patient when introducing me into molecular dynamics simulations, and sorry for switching to GROMACS when I returned to Vienna.

I also want to thank all my colleagues I met during the last four years. Lars, Daniela, René and Andi, you have been with me from the start and I extremely appreciate all your help.
Thank you Lars for teaching me the fine art of svl and for answering all the questions that bothered me, whether or not related to science. Since you are one of the most intelligent people I met, working with you was extremely valuable for me.
Thank you Daniela for your patience and helpfulness. From the beginning on I could always count on you and even now, at the end of my PhD, I thank you for all your help regarding LaTeX. Thank you René for being you. You are one of the most unique persons I met. And although you seemed to be stressed out almost all the time, you could never refuse a request from

Contents

Chapter 1

Introduction

According to the *International Agency for Research on Cancer* there were 12.7 million new cancer cases in 2008 worldwide.[1] Although early diagnosis and improved therapies lead to higher survival rates of affected people, there were still 7.6 million deaths according to cancer the same year. This prompted the World Health Organization (WHO) to consider cancer as one of the ten leading causes of death worldwide.[2] Especially lung cancer shows an extremely low survival rate, accounting for 18.2% of all cancer deaths. A major reason for the failure of anti-cancer therapies is the development of multidrug resistance (MDR) . Already in 1976, Juliano and Ling could link the occurrence of MDR to the expression of a membrane protein, P-glycoprotein (P-gp, ABCB1)[3] .

P-gp, expressed by the *mdr1* gene, belongs to the ATP binding cassette (ABC) superfamily, whose members are mainly involved in the active translocation of various substances. Depending on the family member, transported substrates range from amino acids, lipids to hydrophobic or charged small molecules. The transport profile of P-gp and a number of other ABC transporters is highly polyspecific, including a vast variety of chemically diverse substrates. These so called multidrug transporters exhibit essential detoxifying and barrier functions.[4] Even so, disease related overexpression results in the extrusion of therapeutic drugs and as a consequence triggers MDR.[5,6]

Considering the acquisition of transporter related MDR, there are several strategies to address this problem:[7]

- Down-regulation of the transporter expression

- Prevention of disease-related up-regulation of transporter expression

- Transporter inhibition

- By-passing the transporter

Besides the genetically and cell-signaling based attempts to avoid the occurrence of MDR by interfering with transporter-expression, inhibiting or by-passing the protein are approaches that are investigating direct interaction with the efflux pump P-gp.[8,9] These approaches can be picked up by *in silico* methods and help overcoming MDR by either screening for new P-gp inhibitors, that would be able to stop the transporter to export therapeutics, or by early detection of P-gp substrates and subsequent out-designing of substrate properties. A detailed description how *in silico* methods can contribute to this problem is described in the two reviews in Sections 2.1.1 and 2.1.2.

It is striking, that although P-gp is already known for more than 30 years, there is still no P-gp inhibitor that entered the market. This can be probably explained by the polyspecific ligand recognition pattern and the lack of high resolution structural information that are responsible for the limited information available on the molecular basis of ligand/transporter interaction.

Thus, the aim of this thesis was to analyze and understand the intermolecular interaction between P-gp and its inhibitors by performing structure-based *in silico* methods. The obtained information should be furthermore used for the identification of new P-gp inhibitors and the classification of existing drugs or candidates.

Although the high flexibility and polyspecificity of P-gp make the use of structure-based approaches rather difficult, the application of such is essen-

tial to understand the mechanism of binding of a drug to its target. Generally, working with membrane proteins bears a lot of challenges, ranging from the lack of high-resolution structures to not fully understood transport cycles (outlined in the book chapter "Molecular Modeling and Simulation of Membrane Transport Proteins" included in Section 2.2.1). That is why the crystal structure of mouse P-gp,[10] resolved in 2009, raised a lot of hope and expectations for the application of structure-based design on this protein. To which extent those could be met is described in the review, presented in section 2.1.3. However, the review also points out the limitations of the X-ray structure, thus recommending careful validation when performing structure-based methods. This implements integrating as much external information as possible to validate the outcome of such studies.

Chapter 3 describes two examples, that show how the combination of docking and information-based pose selection can be used to explain inhibitor-protein interaction and finally discover new P-gp inhibitors.

At first, the publication by Jabeen et al., (Section 3.1), presents the successful application of a docking protocol, that implements agglomerative hierarchical clustering in the pose selection process to investigate the stereoselective effect of benzopyrano oxazines.

Similarly to benzopyrano oxazines, also propafenone and its derivatives are known for their P-gp modulating activity and are the focus of the paper "Exhaustive Sampling of Docking Poses Reveals Binding Hypotheses for Propafenone Type Inhibitors of P-Glycoprotein" (Section 3.2) and Section 3.3. The core of the study is the in-depth docking workflow that stands out for its thorough pose evaluation protocol, integrating the information of ligand-based studies about propafenone derivatives.[11-15]

Besides screening for new P-gp inhibitors the identification of such is of high interest for drug discovery. As compounds interfering with the efflux pump's function might be involved in unwanted drug-drug interactions, the Food and Drug Administration (FDA) commits new drug candidates to be screened for P-gp interaction.[16] Thus, the manuscript presented in Section 3.4 describes the development of supervised machine-learning classification models and their performance in comparison with docking as

a representative for a structure-based classification approach.

In brief, this thesis guides you through the challenging task of understanding the binding of P-gp inhibitors to the protein and using this information for finding new P-gp modulators.

Bibliography

[1] International Agency for Research on Cancer. World cancer report 2008. http://www.iarc.fr/en/publications/pdfs-online/wcr/.

[2] American Society of Cancer (2011). Global cancer facts and figures. http://www.cancer.org/Research/Cancer FactsFigures/GlobalCancerFactsFigures/index.

[3] Juliano RL, Ling V (1976) A surface glycoprotein modulating drug permeability in chinese hamster ovary cell mutants. Biochimica et biophysica acta 455: 152–62.

[4] Borst P, Elferink RO (2002) Mammalian abc transporters in health and disease. Annu Rev Biochem 71: 537–92.

[5] Persidis A (1999) Cancer multidrug resistance. Nat Biotechnol 17: 94-5.

[6] Gottesman MM, Ambudkar SV (2001) Overview: Abc transporters and human disease. J Bioenerg Biomembr 33: 453–8.

[7] Potschka H (2011) Role of cns efflux drug transporters in antiepileptic drug delivery: Overcoming cns efflux drug transport. Advanced drug delivery reviews .

[8] Szakács G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM (2006) Targeting multidrug resistance in cancer. Nat Rev Drug Discov 5: 219–34.

[9] Nobili S, Landini I, Mazzei T, Mini E (2011) Overcoming tumor multidrug resistance using drugs able to evade p-glycoprotein or to exploit its expression. Medicinal research reviews .

[10] Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of p-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–22.

[11] Ecker G, Chiba P, Hitzler M, Schmid D, Visser K, et al. (1996) Structure-activity relationship studies on benzofuran analogs of propafenone-type modulators of tumor cell multidrug resistance. J Med Chem 39: 4767–74.

[12] Chiba P, Ecker G, Schmid D, Drach J, Tell B, et al. (1996) Structural requirements for activity of propafenone-type modulators in p-glycoprotein-mediated multidrug resistance. Mol Pharmacol 49: 1122–30.

[13] Chiba P, Hitzler M, Richter E, Huber M, Tmej C, et al. (1997) Studies on propafenone-type modulators of multidrug resistance iii: Variations on the nitrogen. Quant Struct-Act Rel 16: 361–366.

[14] Ecker G, Huber M, Schmid D, Chiba P (1999) The importance of a nitrogen atom in modulators of multidrug resistance. Mol Pharmacol 56: 791–6.

[15] Ecker GF, Stockner T, Chiba P (2008) Computational models for prediction of interactions with abc-transporters. Drug Discov Today 13: 311–7.

[16] Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (2012). Drug interaction studies - study design, data analysis, implications for dosing, and labeling recommendations. http://www.fda.gov/downloads/Drugs/GuidanceCompliance RegulatoryInformation/Guidances/UCM292362.pdf.

Chapter 2

Background

2.1 *In silico* studies of P-glycoprotein

Contribution of the thesis author

Pharmacoinformatic Approaches to Design Natural Product Type Ligands of ABC- Transporters F. Klepsch, I. Jabeen, P. Chiba and G.F. Ecker

F. Klepsch gathered literature about the topic of this review and wrote the parts *Structure-based studies*, *Homology models*, *Binding sites* and *Ligand docking*.

Using Structural and Mechanistic Information to Design Novel Inhibitors/ Substrates of P-Glycoprotein F. Klepsch, T. Stockner, T. Erker, M. Müller, P. Chiba and G.F. Ecker

F. Klepsch gathered literature about the topic of this review and wrote the parts *Protein homology models of P-gp*, *Binding sites* and *Ligand docking*.

Impact of the Recent Mouse P-Glycoprotein Structure for Structure-Based Ligand Design F. Klepsch and G.F. Ecker

F. Klepsch gathered literature about the topic of this review and wrote the parts *ABC transporter structures*, *Homology models for structure-based design* and *Docking studies*.

Pharmacoinformatic Approaches to Design Natural Product Type Ligands of ABC-Transporters

F. Klepsch[1], I. Jabeen[1], P. Chiba[2] and G. F. Ecker[1,*]

[1]*University of Vienna, Department of Medicinal Chemistry, Althanstrasse 14, 1090 Vienna, Austria,* [2]*Medical University of Vienna, Institute of Medical Chemistry, Währinger Straße 10, 1090 Vienna, Austria*

Abstract: ABC-transporter have been recognized as being responsible for multiple drug resistance in tumor therapy, for decreased brain uptake and low oral bioavailability of drug candidates, and for drug-drug interactions and drug induced cholestasis. P-glycoprotein (ABCB1), the paradigm protein in the field, is mainly effluxing natural product toxins and shows very broad substrate specificity. Within this article we will highlight SAR and QSAR approaches for designing natural product type inhibitors of ABCB1 and related proteins as well as in silico strategies to predict ABCB1 substrates and inhibitors in order to design out undesirable drug/protein interaction.

Keywords: Natural products, ABC transporter, P-glycoprotein, *in silico* methods.

INTRODUCTION

More than 30 years ago P-glycoprotein (P-gp, ABCB1), the paradigm ABC-transporter, has been discovered as being responsible for decreased accumulation of natural product toxins in tumor cells. [1] It soon became evident that P-gp has a remarkably broad substrate pattern transporting numerous structurally and functionally diverse natural products across cell membranes. The multispecific nature of this drug efflux transporter and its potential role in clinical drug resistance raised high expectations and initiated development of inhibitors that would re-establish sensitivity to standard therapeutic regimens [2]. However, since the identification of the P-gp inhibitory potential of verapamil [3] almost 3 decades have passed and still no P-gp inhibitor entered the market. Furthermore, since the discovery of P-gp in 1976 [4], additional 47 human ABC-transporters have been identified of which several have been related to either human disease or drug resistance [5]. Within the past decade considerable progress has been made in unravelling the physiological function of P-gp and other ABC-transporters. Results clearly demonstrated the multiple involvement of several members of the ABC-transporter family in drug-uptake, -disposition and –elimination [6] rendering them antitargets rather than classical targets suited for drug therapy. Within this article we will highlight ligand- and structure-based approaches targeting P-gp and some of its homologues by natural products and related compounds. In addition, we will also summarise recent attempts for predicting P-gp substrates, a topic which is becoming more and more important in the ABC-transporter field.

LIGAND BASED APPROACHES

P-glycoprotein and its congeners are membrane-spanning proteins and thus until very recently only little structural information was available. Therefore, in lead optimization programs, mainly ligand-based approaches have been pursued. These include QSAR studies on structurally homologous series of compounds, such as verapamil analogues, triazines, acridonecarboxamides, phenothiazines, thioxanthenes, flavones, dihydropyridines, propafenones and cyclosporine derivatives [7, 8]. These studies pinpoint the importance of H-bond acceptors and their strength, of the distance between aromatic moieties and H-bond acceptors as well as the influence of global physicochemical parameters, such as lipophilicity and molar refractivity. In the quest for designing more

potent inhibitors of ABC-transporter with high selectivity, also natural products served as basic scaffolds for lead optimization programs. In the following section we will highlight selected studies dealing with flavonoids, steroids and sesquiter-penes.

Flavonoids

Flavonoids represent a major class of natural compounds widely present in foods and herbal products (Fig. (**1**)). They have been shown to block both the breast cancer resistance protein (BCRP, ABCG2) [9, 10] and P-glycoprotein (P-gp) [11]. In order to develop more potent inhibitors of ABCG2, a set of flavonoids covering five flavonoid subclasses (flavones, isoflavones, chalcones, flavonols and flavanones) (Fig. (**2**)), were selected for quantitative structure activity (QSAR) relationship studies [9].
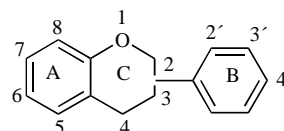
Fig. (1). Basic Structure of flavonoids (taken from [9]).

Systematic structure activity relationship studies showed that the presence of a 2, 3-double bond in ring C, ring B attached at position 2, hydroxylation at position 5, lack of hydroxylation at position 3 and hydrophobic substituents at positions 6, 7, 8 or 4´, are the structural requirements for potent flavonoid- type BCRP inhibitors. Remarkably, although both ABCB1 and ABCG2 are polyspecific in ligand recognition, flavonoids show a different SAR pattern for the two transporters. A notable difference is that 3-hydroxylation was shown to increase flavonoid–P-gp interaction, whereas O-methylation of this hydroxyl group markedly decreased the interaction. Furthermore, hydroxylation at position 7 did not alter flavonoid–Pgp interaction [12], but moderately increased the flavonoid–BCRP interaction. Also in the series of propafenone-type inhibitors, subtle differences in ABCB1 and ABCG2 inhibitory activity could be observed within the same chemical scaffold [13]. In a study on tariquidar analogs, Wiese and co-workers performed Free- Wilson [14] analyses to identify the structural elements which significantly influence the inhibitory effect on ABCB1 and ABCG2 [15]. It was shown that methoxy groups in positions 6 and 7 of the tetrahydroisoquinolinylamide substructure contribute statistically significant to ABCB1 inhibition. In contrast, the elimination of methoxy groups in positions 6 and 7 of the tetrahydroisoquinoline substructure strengthened the interaction with ABCG2. Moreover, it

*Address correspondence to this author at the Department of Medicinal Chemistry, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; Tel: +43-1-4277-55110; Fax: +43-1-4277-9551; E-mail: Gerhard.f.ecker@univie.ac.at
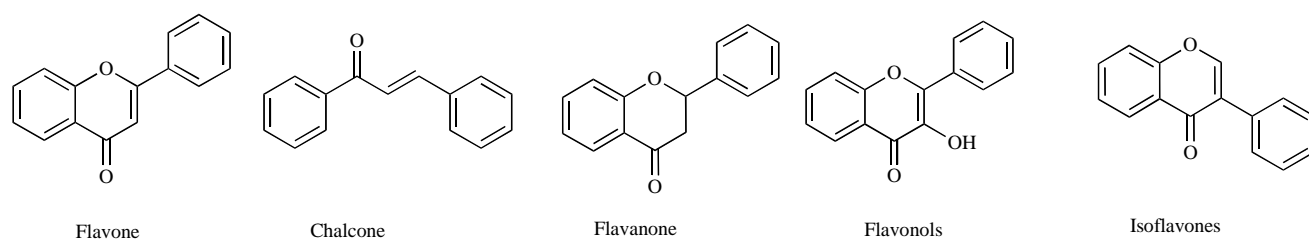
Fig. (2). Basic structures of five flavonoid subclasses (flavones, isoflavones, chalcones, flavanols and flavanones) used for QSAR study (taken from [9])

was demonstrated that the introduction of an electrophilic substituent, such as a nitro group, increases ABCG2 inhibitory potency relative to that for ABCB1.

However, in contrast to propafenones, flavonoids are supposed to interact with the nucleotide binding domain of the transporter. Thus, these differences in the SAR pattern may reflect the distinct structural requirements for binding to the NBDs of ABCG2 and ABCB1. Based on the QSAR model derived, logP makes a positive contribution to the ABCG2 inhibition activity. These findings were considered useful for developing potent flavonoid type inhibitors of ABCG2 (e. g. 7, 8-benzoflavone) with potential clinical applicability [9].

STEROIDS

Steroids have been shown in numerous experiments to exhibit typical properties of MDR-reversing agents [16, 17]. Steroids are perfectly suited for 3D-QSAR studies such as CoMFA and CoMSIA, as they are rather rigid and small differences in structure give rise to considerable changes in biological activity. [18] Remarkably, in the class of steroids CoMSIA models were built for distinguishing which characteristic features are important for a steroid to be a substrate or an inhibitor of ABCB1. [19] Twenty steroids were selected from the literature [20] and divided into two groups: the substrate group contain 13 compounds, while the inhibitor group comprised all 20 compounds (Table **1**). The overall chemical structures are shown in Fig. (**3**).
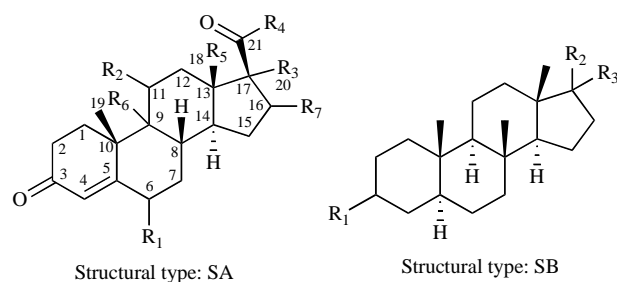
Fig. (3). Template structures of two different types of steroidal Compounds (taken from [19]).

The authors conclude that the requirement for strong hydrophobicity is more essential for inhibitors than for substrates. Another major difference is that for steroid substrates bulky substitutions surrounding C-6 are not well tolerated, whereas electronegative charged groups in position C-11 β are favorable. Moreover, for steroid inhibitors bulky groups around C-3 decrease the activity, while there is no specific requirement at C-3 for steroid substrates. Any substituents around C-17α and C-21 β favor inhibitory potency, but disfavor or have little impact on substrates properties (Fig. (**4**) and (**5**)).

Table 1.   Steroidal Data Set Used in 3D-QSAR Analysis

No	Steroid Compound	Structural Type	Substrate (S)/ Inhibitor (I)
1	Cortisol	SA	S + I
2	17 α-Hydroxyprogesterone	SA	S + I
3	Progesterone	SA	S + I
4	Corticosterone	SA	S + I
5	11-Deoxycortisol	SA	S + I
6	Medroxyprogesterone Acetate	SA	S + I
7	Aldosterone	SA	S + I
8	Dexamethasone¶	SA	S + I
9	Dehydroepiandrosterone†	SB	S + I
10	Pregnenolone†	SB	S + I
11	Testosterone‡	SB	S + I
12	Androstenedione‡	SB	S + I
13	Dihydrotestosterone	SB	S + I
14	Deoxycorticosterone	SA	I Only
15	Medroxyprogesterone	SA	I Only
16	16 α-Methylprogestrone	SA	I Only
17	17 α- Hydroxypregnenolone†	SB	I Only
18	Androsterone	SB	I Only
19	Pregnanedione	SB	I Only
20	6, 16-α-Methylpregnenolone†	SB	I Only

1, 2-Double bond, †5, 6-Double bond, ‡4, 5-Double bond

SESQUITERPENES

Sesquiterpenes have been isolated from the extracts of the *Celastraceae* family and have been used for centuries in traditional medicine. Furthermore, they have shown clinical potential as anticancer drugs [21]. In a comprehensive study, 76 Dihydro-*β*-agarofuran derivatives were used to inhibit P-gp-mediated daunorubicin (DNR) efflux from intact cells [22] (Fig. (**6**)).

Structure-activity relationship studies [22] of compounds varied at the A-ring of sesquiterpenes suggest that an ester group at position C-2 seems essential for the inhibition of ABCB1.
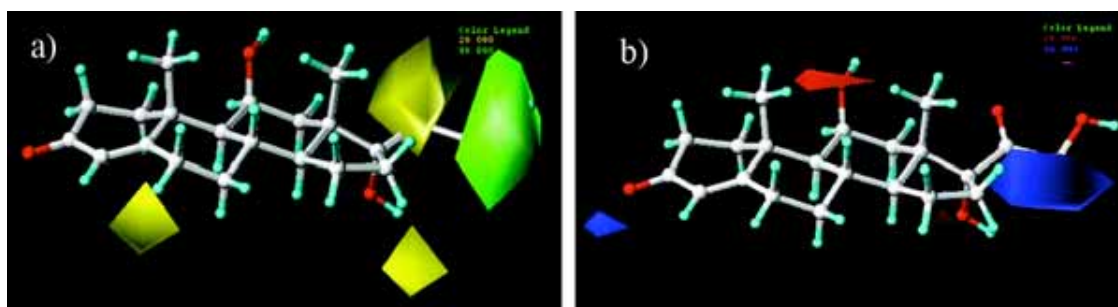
Fig. (4). (a) shows steric contour maps of steroid substrates, the green contours suggest that the larger substituent around the C-21α position is sterically favorable while substitutions at C-6, C-17α and C-21β positions are sterically unfavorable. (b) shows the electrostatic contour maps of steroid substrates, red and blue contours describe the electrostatic regions, which are favorable and unfavorable to a negative charge, respectively. A negatively charged substituent at C-11β and electrostatic groups around C-3, C-17α and C-21 are favorable for interaction between steroid substrates and Pgp (taken from [19]).
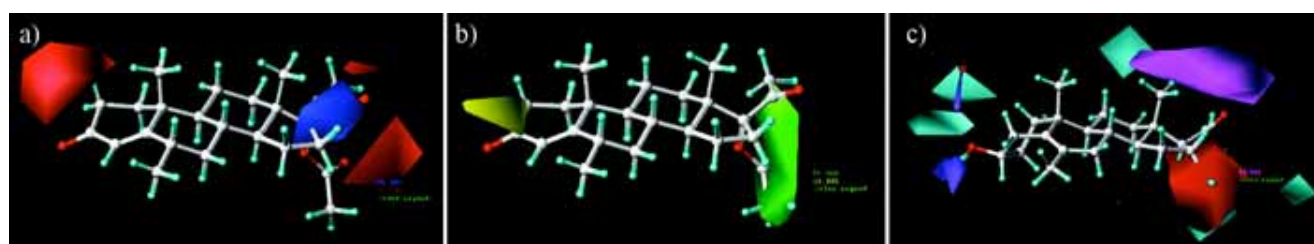
Fig. (5). (a) shows electrostatic contour maps of steroid inhibitors, negative charge favored red regions were found near C-3, C-17 and C-21 positions while positive charge favored or negative charge unfavored blue region is found around C-16β position. (b) shows steric contour plots of steroid inhibitors. Bulky groups in the vicinity of C3 are not tolerated, whereas a bulky substituent like –C (O) CH3 around C-21 may greatly enhance the binding affinity to P-gp. (c) Representation of H-bond donor and acceptor contour maps of steroid inhibitors. The cyan and purple contours indicate regions, where an H-bond donor group increases or decreases activity, respectively. The magenta and red contours indicate regions, in which an H-bond acceptor group increases or decreases activity, respectively. Small purple contours around C-3 suggest that a hydrogen-bond acceptor such as a carbonyl group may increase the inhibitory effect. Large cyan contours around the first hexagonal ring (constituted by C-1–C-6 with the exception of C-3), and the C-21a positions reveal that hydrogen-bond donors such as a methyl or hydroxyl group may enhance the inhibitory potency. Red and magenta contours around C-17 and C-21 indicate that these regions are very sensitive to hydrogen-bond donor or acceptor strength with respect to interaction with P-gp (taken from [19]).
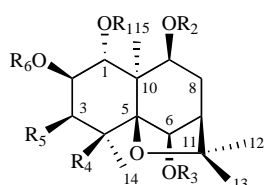
R1 = Ac, Bz, MeBut, Cin
R2 = Fu, Bz, Cin
R3 = Fu, Ac, H, TFAc, MeO-Bz, NO_2-Bz, Bz
R4 = OH, H,
R5 = OAc, H, ONic, OBz
R6 = Fu, Bz, H

Fig. (6). Common Scaffold of sesquiterpenes assayed for the inhibition of the human P-gp.

Sesquiterpenes with the OAc substituent at position C-3 were found to be more potent than the compounds with a hydroxyl or hydrogen group at the same position. It seems that the presence of an H-bond acceptor at C-3 is important for activity.

CoMSIA and CoMFA studies demonstrated that the carbonyl groups at the C-2, C-3, and C-8 position, act as acceptors for H-bond donors in the binding site (Fig. (**7**)). In addition, the models also point towards the importance of a bulky hydrophobic substituent at the C-2β position (depicted as a green sphere) and a
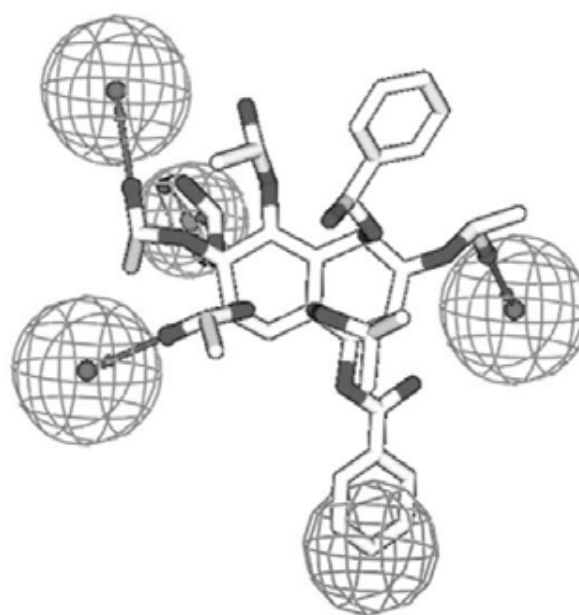
Fig. (7). Summary of the most prominent structural elements of ligands that are important for high P-gp activity obtained by 3D-QSAR/CoMFA (taken from [22]).

hydrophobic substituent at the C-6 position (depicted as a blue sphere). In general, the important features rendering sesquiterpenes highly active are the overall esterification level of the compounds, the presence of at least two aromatic-ester moieties (such as a benzoate-nicotinate or benzoate-benzoate), and the size of the molecule. Tetra- or penta-substituted sesquiterpenes show the highest potency, whereas additional ester moieties in the molecule lead to inactive compounds.

STRUCTURE-BASED STUDIES

The general architecture of ABC transporters are more or less the same throughout this superfamily (Fig. (**8**)). Two transmembrane (TM) and two nucleotide binding (NB) domains are necessary to yield a functional efflux pump which can export its substrates. Since the NB domains harbor the hallmark ABC motifs they are highly conserved among all ABC transporters. Much less sequence identity can be found in the two transmembrane domains (TMD) which are generally responsible for drug binding and therefore the reason for diverse substrate/inhibitor profiles of representatives of this protein family. The structures of majorly prokaryotic ABC transporters were recently reviewed by Rees et al. [23], so we will concentrate on the three main human ABC transporters that are involved in multidrug resistance, ABCB1, ABCC1 and ABCG2. In the case of ABCB1 and ABCC1 all four domains are fused into a single polypeptide chain with the first TMD containing the N-terminus and the second NBD representing the C-terminus of the proteins. By contrast ABCG2 is a half transporter which has to homodimerize to be functional [24]. In addition, an inverse topology with respect to ABCB1 and ABCC1 can be observed, indicating that the NBD lies N-terminal of the TMD [25]. The hallmark of the ABCC1 transporter is a third TMD at the N-terminus referred to as TMD0.
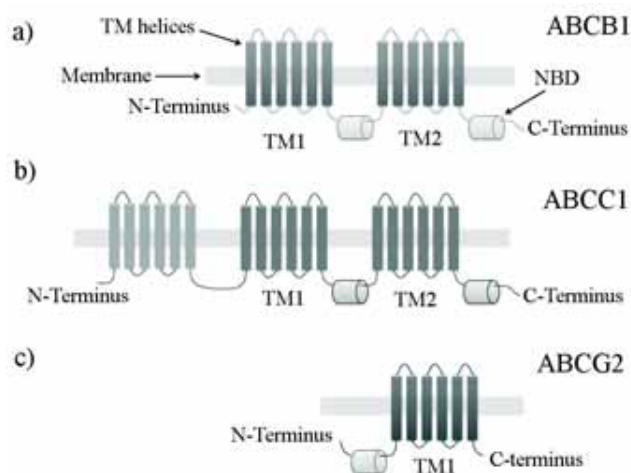
Fig. (8). Comparison of different domain architecture of the ABC transporters ABCB1, ABCC1 and ABCG2.

ABC efflux pumps are flexible proteins that in association with drug binding and subsequent ATP hydrolysis undergo conformational changes. ABCB1 adopts at least three different states following ATP-binding and subsequent hydrolysis (reviewed in [26]). The apo or "open-inward" conformation is considered the ground state. In this conformation the protein shows an inverted "V" open towards the cytosolic environment of the cell. Substrates are considered to bind to this state with higher affinity. The second conformation that can be captured by ABCB1 is the nucleotide-bound form which is open to the extracellular space. After hydrolysis of two ATP molecules ABCB1 returns to the initial state (Fig. (**9**)).
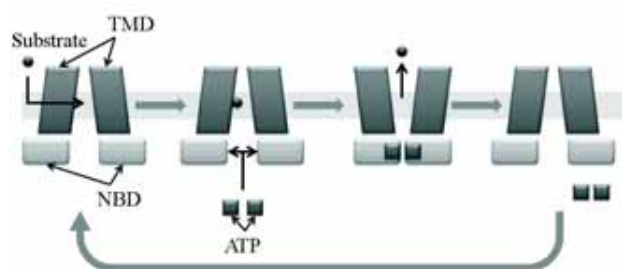
Fig. (9). Schematic illustration of the catalytic cycle of ABC transporters on the basis of ABCB1. The two different conformations are depicted before and after drug binding.

HOMOLOGY MODELS

The fact that ABC transporters are embedded in the membrane complicates the crystallization process of such proteins. Therefore, protein homology modeling based on templates of bacterial homologues representing different catalytic states, was the method of choice for structure-based studies. Table **2** gives an overview of current available homology models of selected ABC transporters. Due to its high resolution the crystal structure of the *Staphylococcus aureus* transporter SAV1866 (PDB code: 2HYD, resolution: 3. 00 Å) [27] in the ADP bound "outward-facing" form often served as modeling template [28-33]. The same transporter crystallized in the AMP. PNP bound state[34] also served as modeling template [28]. Several high resolution structures of different catalytic states of ABC-proteins were also obtained with the bacterial transporter MsbA [35] as template. This information gave new insights into the transport cycle and the associated conformational change of ABC proteins (Table **3**).

Since March 2009 the first X-ray structure of a eukaryotic ABC efflux pump, ABCB1 (mouse) is available [36] (PDB code: 3G5U, resolution: 3. 8 Å). With 87 % sequence identity to human ABCB1 and moderate resolution (3. 80) it serves as a good template for homology modeling [37]. Additionally the structure was published together with two co-crystallised enantiomeric cyclic peptide inhibitors (CPPIs; QZ59-RRR and QZ59-SSS) (Fig. (**10**)). This new information sheds light on possible ligand binding areas of ABCB1.
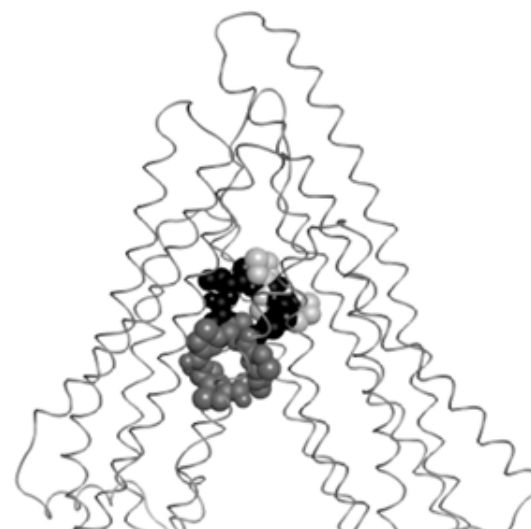
Fig. (10). Cocrystallized ABCB1 with cyclic P-gp inhibitors (CPPIs) QZ59-RRR (black) QZ59-SSS (dark and light grey).

Table 2. **Homology Models of the ABC Transporters ABCB1, ABCC1 and ABCG2**

ABC Transporter	Template	Sequence Identity / Homology	Catalytic State	Reference
ABCB1	Mouse ABCB1	87 % / 93 %	Apo	[36]
ABCB1	SAV1866	34 % / 52 %	ADP-bound	[28-30, 33, 37]
ABCB1	MsbA	37 % / 57 %	AMP-PNP	[28]
ABCB1	MalK	31 % / 50 %	Apo	[33]
ABCC1	SAV1866	28 % / 49 %	ADP	[31]
ABCG2	SAV1866	27 % / 49 %	ADP	[32]

Table 3. **Structures of Whole ABC Transporters that are Available Until Now**

ABC Transporter	Organsim	Catalytic State	Resolution [Å]	PDB Code	References
ABCB1	Mouse	Apo	3. 80	3G5U	[36]
ABCB1	Mouse	Apo	4. 40	3G60	[36]
ABCB1	Mouse	Apo	4. 35	3G61	[36]
ABCB1	Hamster	Apo	~20	-	[68]
ABCB1	Hamster	AMP-PNP	~20	-	[68]
ABCB1	Hamster	ATP	8	-	[69]
ABCC1	Human	ATP	~22	-	[70]
ABCG2	Insect	ATP	~18	-	[71]
SAV1866	Staphylococcus aureus	ADP	3. 00	2HYD	[27]
SAV1866	Staphylococcus aureus	AMP-PNP	3. 40	2ONJ	[34]
MsbA	Escherichia coli	Apo	5. 30	3B5W	[35]
MsbA	Vibrio cholerae	Apo	5. 50	3B5X	[35]
MsbA	Salmonella typhimurium	AMP-PNP	4. 50	3B5Y	[35]
MsbA	Salmonella typhimurium	ADP-OV	4. 20	3B5Z	[35]
MsbA	Salmonella typhimurium	AMP-PNP	3. 70	3B60	[35]

BINDING SITES

It was shown that a functional unit of ABC-transporters has to consist of two TM and two NB domains. Only with this architecture a functional transporter can be obtained. Nevertheless, mutational studies showed that ABC transporters consisting of just two TMD regions without NBDs were able to bind ligands [38]. This led to the assumption that drug binding occurs in the TMD region.

Numerous experimental studies were performed trying to determine the different drug binding sites of P-glycoprotein, comprising among others cysteine and arginine scanning and photoaffinity labeling (reviewed in [26, 39, 40]). The overall assumption in this case is that P-glycoprotein possesses a huge binding pocket with at least four distinct binding sites, with TM 6 as main interaction helix. Well characterized are the binding sites of Rhodamine and Hoechst 33342, the so called R- and the H-site [41, 42]. Additionally, there is evidence for an allosteric regulatory site as well as a region where progesterone and prazosin may bind [43, 44]. These conclusions go hand in hand with the previously mentioned co-crystal structure of ABCB1 together with isomeric CPPIs [36]. The structure shows a huge binding pocket where the rather large cyclopeptides bind on different sites with partially overlapping interacting amino acid residues. Some of these residues are identical with the ones that are involved in rhodamine or verapamil binding [45, 46]. These data are also consistent with drug binding studies with the ABC transporter ABCG2. Also for ABCG2 at least four different binding sites, one H-site, a prazosin area and probably two different R-sites on each monomer have been postulated. [47]. The involvement of both monomers in rhodamine 123 binding can also be observed with ABCC1 where TMD1 and TMD2 are interacting [48].

Nature derived substrates, especially cytotoxins, are supposed to bind to a certain region in the binding pocket of the transmembrane domains of ABC transporters. However, large compounds with a steroidal architecture tend to bind to the ATP-binding site in the NBD region of the protein. As competitors of

ATP they are also able to inhibit the function of the MDR transporter.

LIGAND DOCKING

The computational method of ligand docking is a good way to validate experimentally derived binding pockets or even to propose new areas of binding. Several docking studies of natural compounds have been performed. Recently published docking results show quinazolinones binding at the same site like the CPPIs [37]. The docking poses are in accordance with pharmacophore modeling, which suggests a hydrogen bond between the ligand and the amino acid residue Tyr307 (TM5). In addition, protein-ligand interaction fingerprints (PLIF) were calculated, resulting in the residues Phe336 (TM6), Tyr953 (TM11) and Phe957 (TM11) performing contact interactions (Fig. (**11**)). The binding pocket was described as highly hydrophobic which excludes ionic interactions with tertiary amines. Therefore it was suggested that such interactions can be built after the conformational change of the protein and thus has to be validated with an outward facing model.

Similar results were also obtained in our group when performing docking studies with a homology model of ABCB1 and propafenone derivatives. Our results also showed interactions with the transmembrane helices mentioned above. This confirms the assumption of a large binding pocket and indicates overlapping quinazolinone and propafenone binding sites. In Fig. (**12**) an overview of interactions of drugs with certain TM helices is depicted. As can be noticed, TM 6 plays a crucial role in ligand binding.

The assumption that certain ABC transporter inhibitors of natural origin compete with ATP at the NBDs could also be confirmed by docking [49]. A screening of 122 compounds against the three MDR related proteins ABCB1, ABCC2 and ABCG2, revealed that several compounds showed multi-specificity. Since the highest sequence identity among these proteins can be found in the NBDs these compounds were docked into the crystal structure of the NBD1 of ABCC1 [50]. The results showed that the most hydrophilic natural products quercetin and sylimarin together with the potent compound MK571 were able to bind to the structure with
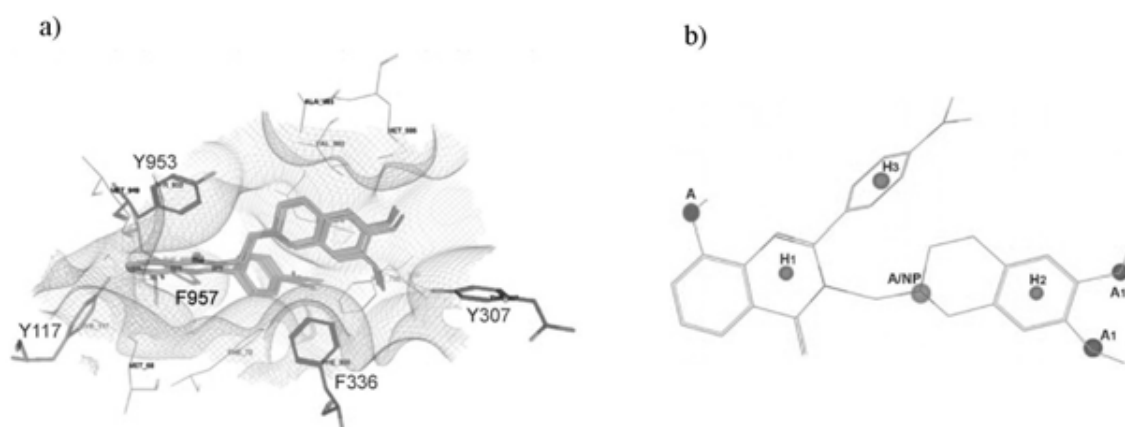

Fig. (11). a) Docking poses of quinazolinones in an ABCB1 homology model, b) Pharmacophore model (taken from [37]).
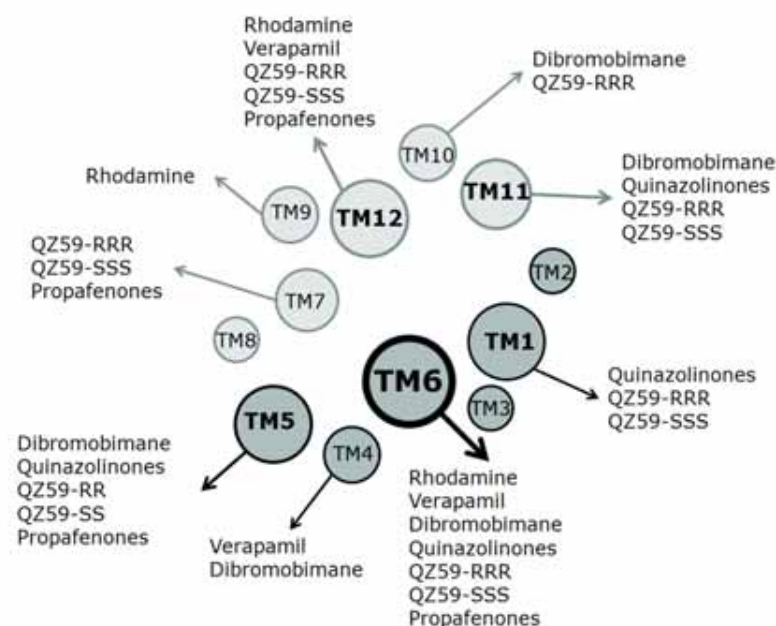
Fig. (12). Transmembrane (TM) helix interactions with investigated ABCB1 ligands. The circle size depends on the level of interaction.

high scores. More lipophilic inhibitors were not able to provide reasonable scoring values. Regarding the docking poses obtained it is noteworthy to mention that the negatively charged MK571 extends into the catalytic site and its aromatic rings are placed similar to the adenosine base ring of ATP. By contrast the poses of the lipophilic inhibitors showed no interaction with the catalytic site (Fig. (**13**)).

Also steroids and flavonoids were examined with respect to their binding affinity to the ATP binding site [51]. In this study docking of eleven different steroids, one flavonoid, ATP and MANT-ATP into ABCB1 and ABCG2 was performed. The results, which were rather the same for both transporters, suggest overlapping steroid and ATP binding sites near the P-loop of the nucleotide binding domains (Fig. (**14**) and (**15**)). The P-loop (or Walker A) is one of the three characteristic motifs of the NBDs of ABC transporters (Walker A, Walker B and signature motif C) and interacts with the phosphates of the nucleotides. The flavone kaempferide showed amino acid residue interactions similar to ATP. On the other hand the hydrophobic steroid RU-486 bound to a different area than the other steroids and ATP, but overlapped with the kaempferide and the MANT-ATP binding site. RU-486 and MANT-ATP share a highly hydrophobic moiety and both bind

within the hydrophobic cleft around I1050 (Fig. (**14c**)). Additionally the binding free energy of the complexes was calculated. According to this study the steroids investigated bind with the same affinity as ATP, which renders them potential competitors of ATP (Table **4**).

Similar findings were published in a docking study that concentrated on flavonoids, including flavones, flavonols, flavanones and chalcones [52] (Fig. (**16**)). Calculated binding free energies were compared to experimentally derived K_d-values and a good correlation could be obtained. This study also showed that flavonoids preferably bind to the P-loop of the NBD, especially interacting with residues L1076 and S1077. In addition, the B-ring of flavonoids was supposed to build hydrophobic interactions with Y1044, which originally interacts with the adenosine base of ATP [52]. Comparing the different flavonoid derivatives showed that the additional hydroxyl-group at position 3, which is the only difference between flavonols and flavones, decrease the predicted docking energy because an additional hydrogen bond could be formed. Additional hydrophobic substituents added to flavones and flavonols at positions 6 or 8 also had a positive effect on binding. Chalcones, which show higher flexibility due to the open C-ring structure, also showed reduced docking energy. Especially with
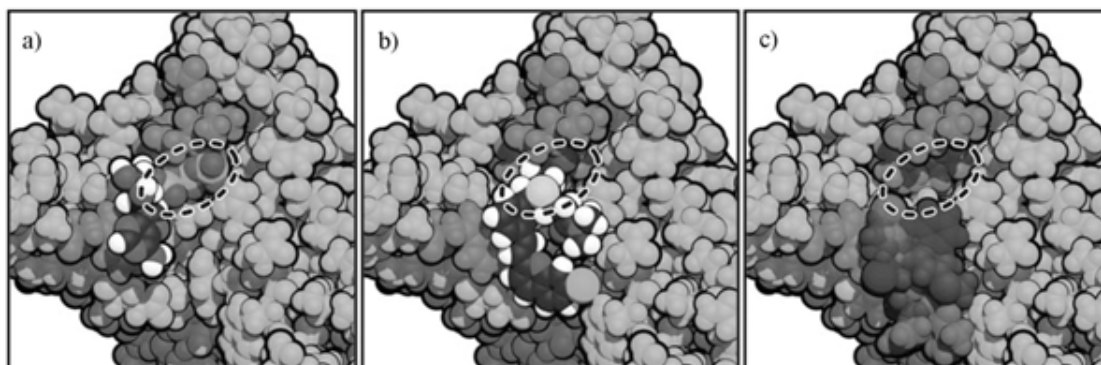
Fig. (13). a) MRP NBD1 cocrystallized with ATP. b) MRP NBD1 with MK-571. c) MRP with lipophilic inhibitors (taken from [50]).

Table 4.   Amino Acid Interactions Observed With Docking Studies of Steroids and Flavonoids Into the NBD

Compound	ABCB2 NBD2 Hydrophobic Interaction	Hydrophilic Interaction	References
Steroids	Y1044, I1050, V1052, G1075, S1077	L1076, R1047, Q1085, P1051	[51]
RU-486	I1050, P1051, V1052, Q1054, N1248	None predicted	[51]
Kaempferide	Y1044, V1052, G1073, G1075	R1047, S1077, T1078	[51]
ATP	Y1044, G1073	G1073, C1074, G1075, S1077, T1078	[51]
MANT-ATP	Y1044, I1050, V1052	G1073, C1074, G1075, L1076, S1077, T1078, Y1087	[51]
Flavones	G1070-T1077	L1076	[52]
substituted Flavones	Y1044, V1052, G1072-T1075, G1073, G1075,	S1072, L1076, S1077, T1078	[52]
Flavonols	G1071-T1076, E1201, D1200, H1232	S1071, C1074, G1075, S1077	[52]
Chalcones	Y1044, V1052, G1071-T1078,	S1077, T1078	[52]
Substituted Chalcones	Y1044, P1048, I1050-V1052, H1232, Q1247-E1249	S1072, G1073, C1074, G1075, L1076, S1077	[52]

Fig. (15). Docking poses of steroids in a homology models of ABCB1 NBD2 (taken from [51]).

of the protein with quite low sequence similarity. However, this trend will probably change due to the recent publication of the structure of mouse P-gp.

IMPORTANCE OF ABC-TRANSPORTER FOR ADMET

With our increasing knowledge on the physiological role of ABC transporter it became evident that there are several distinct transporters which are responsible for severe side effects of drugs and for drug/drug interactions. In these cases the focus shifts from the design of inhibitors to the design of "non-ligands". Thus, the major challenge is to establish models for prediction of substrate properties with the ultimate goal to avoid interaction with these proteins.

ABCB1 is constitutively expressed at several diffusion barriers, such as the blood-brain barrier, the kidney, the liver and the intestine. At the latter it plays an important role in limiting the intestinal absorption of a wide variety of orally administered drugs. One paradigm example is the quinidine-digoxin interaction, where the P-gp inhibitor quinidine increases the digoxin absorption rate by about 30%. But it is not only drug/drug interaction playing a role, there is also proven evidence for drug/nutrient interaction [53]. These include mainly flavonoids found in fruit juices, vegetables, flowers and tea. Especially grapefruit juice has been shown to interfere with plasma levels of colchicines [54], paracetamol [55], and cyclosporine [56].

Thus, the importance of drug transporters for uptake and disposition is now widely accepted and Benet and co-workers proposed a biopharmaceutics Classification System (BCS) which allows prediction of *in vivo* pharmacokinetic performance of drug candidates based on measurements of their permeability and solubility [57]. Subsequently, this classification system was modified in order to allow prediction of overall drug disposition, including routes of drug elimination and the effects of efflux and absorptive transporters on oral drug absorption [58]. The overall message is that compounds with low water solubility being substrates of P-glycoprotein bear the inherent risk of low bioavailability.

Also at the blood-brain barrier (BBB) the important role of ABCB1 and ABCG2 is increasingly recognised. *In vitro* studies demonstrated that the uptake of vincristine was reduced in primary cultured bovine capillary endothelial cells expressing P-gp at the luminal side and that this decreased accumulation was due to active efflux. Steady state uptake was significantly increased in the presence of the P-gp blocking agent verapamil [59]. Additionally, *mdr1a* double knock out mice show hypersensitivity to a range of

Fig. (14). Docking poses of MANT-ATP, ATP and RU-486 into the homology models of ABCB1 NBD2 and ABCG2 NBD. (taken from [51]).

substituted chalcone derivatives, such as O-*n*-C$_{10}$H$_{21}$ chalcone, very low docking energy values were predicted.

Until now the number of docking studies into ABC transporters is still low. As outlined above, most docking studies are restricted to the nucleotide binding domain. This can be explained by the lack of crystal structures of the transmembrane domain, which is the part

Fig. (16). Observed interaction of ABCB1 NBD2 with chrysin (flavones) (a), galangin (flavonol) (b) and O-*n*-C$_{10}$H$_{21}$-chalcone (c) [52].

drugs known to be transported by P-gp [60]. Undoubtedly, selected ABC-transporter are an important impediment for the entry of hydrophobic drugs into the brain.

PREDICTING SUBSTRATE PROPERTIES FOR ABCB1

As already outlined above, ABCB1 is constitutively expressed in several organs, such as kidney, liver, intestine and also at the blood brain barrier (BBB). P-gp substrates therefore show poor oral absorption, enhanced renal and biliary excretion and usually do not enter the brain [61]. Furthermore, they are likely to be affected by the MDR phenotype and are thus not suitable as anticancer agents. This spurred the development of medium- and high-throughput systems addressing the P-gp substrate properties of compounds of interest.

However, data sets for in silico classification studies are rather small and sometimes also inconsistent [62]. Recently the group of Gottesman published a comprehensive study analysing data f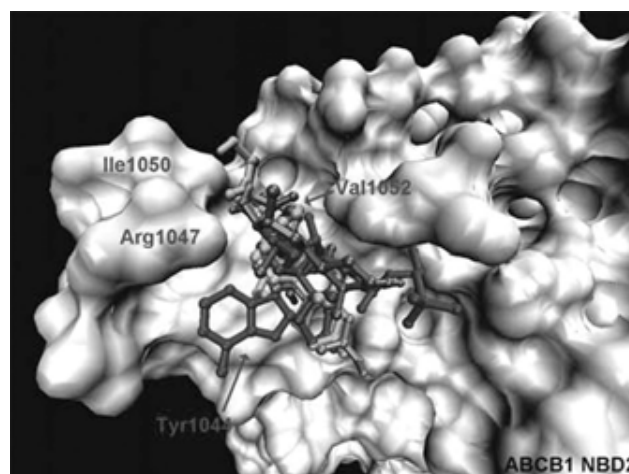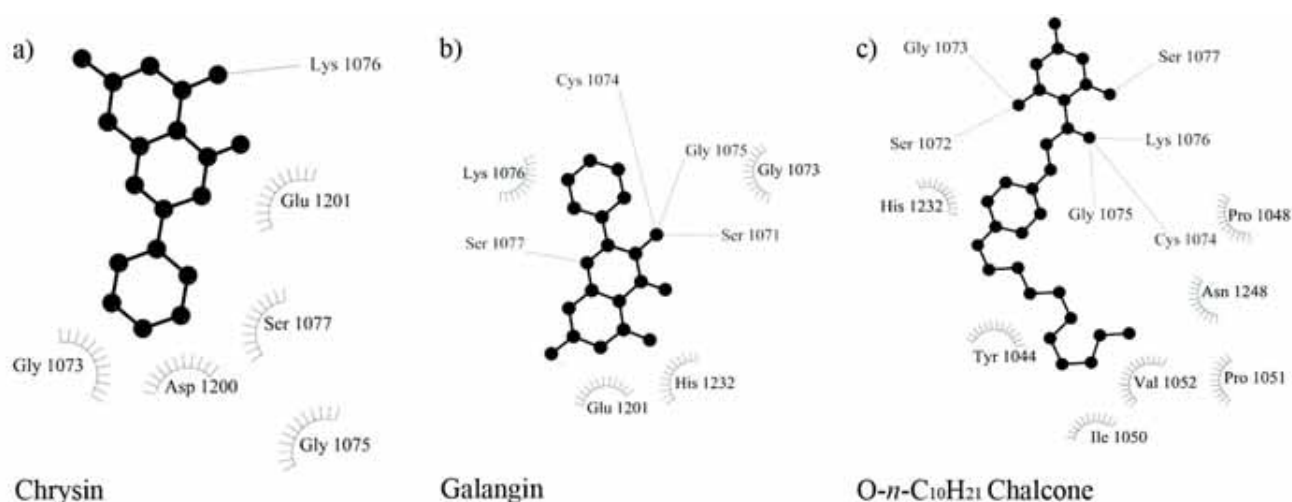rom the NCI60 screen [63], which comprises mostly natural product toxins. m-RNA levels of all 48 human ABC-transporter in 60 human tumour cell lines of the NCI60 anticancer drug screening panel were evaluated and correlated with cellular toxicity values of 1400 selected compounds. An inverse correlation between transporter mRNA levels and compound toxicity indicates that a compound is a substrate for the respective transporter. Undoubtedly, this is by far the largest consistent data set available by now. It is almost exclusively built of natural products, and studies from our group indicate that it might be successfully used as basis for P-gp substrate prediction models.

Based both on this data set as well as on a set of 259 compounds compiled from the literature we explored the performance of several classification methods combined with different descriptor sets. These include simple ADME-type descriptors (such as logP, number of rotable bonds, number of H-bond donors and acceptors), VSA descriptors as described by Labute [64] and 2D auto-correlation vectors. The latter have already been successfully applied for prediction of P-gp inhibitors [65]. When comparing binary QSAR and support vector machines, the latter gave more robust models with total accuracies in the range of 80%. Generally, the prediction of non-substrates performs better than those for substrates [66]. However, more detailed studies are necessary to fully explore the potential and limits of this data set. If successful, this approach might be useful for in silico screening of natural product libraries in order to identify hitherto unknown drug/nutrient

interactions at P-gp and related ABC-transporter involved in ADMET.

OUTLOOK

Although P-glycoprotein and its prominent role in tumour multidrug resistance is known since 1976, up to now no P-gp inhibitor has reached the market. Thus, there is still need for development of new, specific P-gp inhibitors. As P-gp is mainly addressing natural product toxins as substrates, compounds from natural origin are versatile starting points for design of new ligands. Due to the polyspecificity of the protein, complex methods such as self organising maps or random forest classification might pave the way for successful in silico screening approaches, targeted at natural compound libraries. However, within the past decade the focus of interest shifted towards the role of ABC-transporters for ADMET and drug/drug interactions. Several pharmaceutical companies established high throughput screening systems for measuring P-gp substrate properties of their compound libraries and *in silico* methods have been developed which reach classification accuracies in the range of 80%. In this case the most comprehensive data set available up to now uses data from the NCI60 screening library, which is mostly composed of natural product related toxins. Finally, the publication of the structure of mouse P-glycoprotein will aid in the understanding of the molecular principles underlying the ligand-polyspecificity of these transporters and pave the way for structure-based drug design approaches.

REFERENCES

[1]   Gottesman MM, Ling V. The molecular basis of multidrug resistance in cancer: the early years of P-glycoprotein research. FEBS Lett 2006; 580: 998-1009.

[2]   Szakacs G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM. Targeting multidrug resistance in cancer. Nat Rev Drug Discov 2006; 5: 219-34.

[3]   Tsuruo T, Iida H, Tsukagoshi S, Sakurai Y. Overcoming of vincristine resistance in P388 leukemia *in vivo* and *in vitro* through enhanced cytotoxicity of vincristine and vinblastine by verapamil. Cancer Res 1981; 41: 1967-72.

[4] Juliano RL, Ling V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 1976; 455: 152-62.

[5] Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. J Lipid Res 2001; 42: 1007-17.

[6] Szakacs G, Varadi A, Ozvegy-Laczka C, Sarkadi B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). Drug Discov Today 2008; 13: 379-93.

[7] Raub TJ. P-glycoprotein recognition of substrates and circumvention through rational drug design. Mol Pharm 2006; 3: 3-25.

[8] Pleban K, Ecker GF. Inhibitors of p-glycoprotein--lead identification and optimisation. Mini Rev Med Chem 2005; 5: 153-63.

[9] Zhang S, Yang X, Coburn RA, Morris ME. Structure activity relationships and quantitative structure activity relationships for the flavonoid-mediated inhibition of breast cancer resistance protein. Biochem Pharmacol 2005; 70: 627-39.

[10] Zhang S, Yang X, Morris ME. Flavonoids are inhibitors of breast cancer resistance protein (ABCG2)-mediated transport. Mol Pharmacol 2004; 65: 1208-16.

[11] Bansal T, Jaggi M, Khar RK, Talegaonkar S. Emerging significance of flavonoids as P-glycoprotein inhibitors in cancer chemotherapy. J Pharm Pharm Sci 2009; 12: 46-78.

[12] Conseil G, Baubichon-Cortay H, Dayan G, Jault JM, Barron D, Di Pietro A. Flavonoids: a class of modulators with bifunctional interactions at vicinal ATP- and steroid-binding sites on mouse P-glycoprotein. Proc Natl Acad Sci USA 1998; 95: 9831-6.

[13] Cramer J, Kopp S, Bates SE, Chiba P, Ecker GF. Multispecificity of drug transporters: probing inhibitor selectivity for the human drug efflux transporters ABCB1 and ABCG2. ChemMedChem 2007; 2: 1783-8.

[14] Free SM, Jr., Wilson JW. A Mathematical contribution to structure-activity studies. J Med Chem 1964; 7: 395-9.

[15] Pick A, Muller H, Wiese M. Structure-activity relationships of new inhibitors of breast cancer resistance protein (ABCG2). Bioorg Med Chem 2008; 16: 8224-36.

[16] Ueda K, Saeki T, Hirai M, Tanigawara Y, Tanaka K, Okamura M, *et al.* Human P-glycoprotein as a multi-drug transporter analyzed by using transepithelial transport system. Jpn J Physiol 1994; 44 Suppl 2: S67-71.

[17] Yang CP, DePinho SG, Greenberger LM, Arceci RJ, Horwitz SB. Progesterone interacts with P-glycoprotein in multidrug-resistant cells and in the endometrium of gravid uterus. J Biol Chem 1989; 264: 782-8.

[18] Hamilton KO, Yazdanian MA, Audus KL. Modulation of P-glycoprotein activity in Calu-3 cells using steroids and beta-ligands. Int J Pharm 2001; 228: 171-9.

[19] Li Y, Wang YH, Yang L, Zhang SW, Liu CH, Yang SL. Comparison of steroid substrates and inhibitors of P-glycoprotein by 3D-QSAR analysis. J Mol Structure 2005; 733: 111-8.

[20] Barnes KM, Dickstein B, Cutler GB, Jr. , Fojo T, Bates SE. Steroid treatment, accumulation, and antagonism of P-glycoprotein in multidrug-resistant cells. Biochemistry 1996; 35: 4820-7.

[21] Muñoz-Martínez F, Reyes CP, Pérez-Lomas AL, Jiménez IA, Gamarro F, Castanys S. Insights into the molecular mechanism of action of Celastraceae sesquiterpenes as specific, non-transported inhibitors of human P-glycoprotein. Biochimica et Biophysica Acta (BBA) - Biomembranes 2006; 1758: 98-110.

[22] Reyes CP, Munoz-Martinez F, Torrecillas IR, Mendoza CR, Gamarro F, Bazzocchi IL, *et al.* Biological evaluation, structure-activity relationships, and three-dimensional quantitative structure-activity relationship studies of dihydro-beta-agarofuran sesquiter-penes as modulators of P-glycoprotein-dependent multidrug resistance. J Med Chem 2007; 50: 4808-17.

[23] Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. Nat Rev Mol Cell Biol 2009; 10: 218-27.

[24] Kage K, Tsukahara S, Sugiyama T, Asada S, Ishikawa E, Tsuruo T, *et al.* Dominant-negative inhibition of breast cancer resistance protein as drug efflux pump through the inhibition of S-S dependent homodimerization. Int J Cancer 2002; 97: 626-30.

[25] Allikmets R, Schriml LM, Hutchinson A, Romano-Spica V, Dean M. A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance. Cancer Res 1998; 58: 5337-9.

[26] Seeger MA, van Veen HW. Molecular basis of multidrug transport by ABC transporters. Biochim Biophys Acta 2009; 1794: 725-37.

[27] Dawson RJ, Locher KP. Structure of a bacterial multidrug ABC transporter. Nature 2006; 443: 180-5.

[28] Becker JP, Depret G, Van Bambeke F, Tulkens PM, Prevost M. Molecular models of human P-glycoprotein in two different catalytic states. BMC Struct Biol 2009; 9: 3.

[29] Globisch C, Pajeva IK, Wiese M. Identification of putative binding sites of P-glycoprotein based on its homology model. Chem Med Chem 2008; 3: 280-95.

[30] Stockner T, de Vries SJ, Bonvin AM, Ecker GF, Chiba P. Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. FEBS J 2009; 276: 964-72.

[31] DeGorter MK, Conseil G, Deeley RG, Campbell RL, Cole SP. Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1). Biochem Biophys Res Commun 2008; 365: 29-34.

[32] Hazai E, Bikadi Z. Homology modeling of breast cancer resistance protein (ABCG2). J Struct Biol 2008; 162: 63-74.

[33] O'Mara ML, Tieleman DP. P-glycoprotein models of the apo and ATP-bound states based on homology with Sav1866 and MalK. FEBS Lett 2007; 581: 4217-22.

[34] Dawson RJ, Locher KP. Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. FEBS Lett 2007; 581: 935-8.

[35] Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci USA 2007; 104: 19005-10.

[36] Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, *et al.* Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 2009; 323: 1718-22.

[37] Pajeva IK, Globisch C, Wiese M. Combined pharmacophore modeling, Docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. Chem Med Chem 2009.

[38] Loo TW, Clarke DM. The transmembrane domains of the human multidrug resistance P-glycoprotein are sufficient to mediate drug binding and trafficking to the cell surface. J Biol Chem 1999; 274: 24759-65.

[39] Loo TW, Clarke DM. Mutational analysis of ABC proteins. Arch Biochem Biophys 2008; 476: 51-64.

[40] Shilling RA, Venter H, Velamakanni S, Bapna A, Woebking B, Shahi S, *et al.* New light on multidrug binding by an ATP-binding-cassette transporter. Trends Pharmacol Sci 2006; 27: 195-203.

[41] Loo TW, Clarke DM. Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein. J Biol Chem 2002; 277: 44332-8.

[42] Qu Q, Sharom FJ. Proximity of bound Hoechst 33342 to the ATPase catalytic sites places the drug binding site of P-glycoprotein within the cytoplasmic membrane leaflet. Biochemistry 2002; 41: 4744-52.

[43] Martin C, Berridge G, Higgins CF, Mistry P, Charlton P, Callaghan R. Communication between multiple drug binding sites on P-glycoprotein. Mol Pharmacol 2000; 58: 624-32.

[44] Shapiro AB, Fox K, Lam P, Ling V. Stimulation of P-glycoprotein-mediated drug transport by prazosin and progesterone. Evidence for a third drug-binding site. Eur J Biochem 1999; 259: 841-50.

[45] Loo TW, Bartlett MC, Clarke DM. Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket. Biochem J 2006; 399: 351-9.

[46] Loo TW, Clarke DM. Identification of residues in the drug-binding site of human P-glycoprotein using a thiol-reactive substrate. J Biol Chem 1997; 272: 31945-8.

[47] Clark R, Kerr ID, Callaghan R. Multiple drugbinding sites on the R482G isoform of the ABCG2 transporter. Br J Pharmacol 2006; 149: 506-15.

[48] Deeley RG, Cole SP. Substrate recognition and transport by multidrug resistance protein 1 (ABCC1). FEBS Lett 2006; 580: 1103-11.

[49] Matsson P, Pedersen JM, Norinder U, Bergstrom CA, Artursson P. Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP and MRP2 among registered drugs. Pharm Res 2009; 26: 1816-31.

[50] Ramaen O, Leulliot N, Sizun C, Ulryck N, Pamlard O, Lallemand JY, *et al.* Structure of the human multidrug resistance protein 1

nucleotide binding domain 1 bound to Mg2+/ATP reveals a non-productive catalytic site. J Mol Biol 2006; 359: 940-9.

[51] Mares-Samano S, Badhan R, Penny J. Identification of putative steroid-binding sites in human ABCB1 and ABCG2. Eur J Med Chem 2009; 44: 3601-11.

[52] Badhan R, Penny J. In silico modelling of the interaction of flavonoids with human P-glycoprotein nucleotide-binding domain. Eur J Med Chem 2006; 41: 285-95.

[53] Aszalos A. Role of ATP-binding cassette (ABC) transporters in interactions between natural products and drugs. Curr Drug Metab 2008; 9: 1010-8.

[54] Dahan A, Amidon GL. Grapefruit juice and its constituents augment colchicine intestinal absorption: potential hazardous interaction and the role of p-glycoprotein. Pharm Res 2009; 26: 883-92.

[55] Dasgupta A, Reyes MA, Risin SA, Actor JK. Interaction of white and pink grapefruit juice with acetaminophen (paracetamol) *in vivo* in mice. J Med Food 2008; 11: 795-8.

[56] Paine MF, Widmer WW, Pusek SN, Beavers KL, Criss AB, Snyder J, *et al.* Further characterization of a furanocoumarin-free grapefruit juice on drug disposition: studies with cyclosporine. Am J Clin Nutr 2008; 87: 863-71.

[57] Wu CY, Benet LZ. Predicting drug disposition *via* application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. Pharm Res 2005; 22: 11-23.

[58] Custodio JM, Wu CY, Benet LZ. Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. Adv Drug Deliv Rev 2008; 60: 717-33.

[59] Tsuji A, Terasaki T, Takabatake Y, Tenda Y, Tamai I, Yamashima T, *et al.* P-glycoprotein as the drug efflux pump in primary cultured bovine brain capillary endothelial cells. Life Sci 1992; 51: 1427-37.

[60] Borst PS, A. H. In: W. M. Partridge Ed. ^Eds. , Introduction to the blood-brain barrier. Cambridge; Cambridge University Press 1998; 198-206.

[61] Chan LM, Lowes S, Hirst BH. The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. Eur J Pharm Sci 2004; 21: 25-51.

[62] Ecker GFCP. In: G. F. ; Chiba P. Ecker Ed. ^Eds. , Transporters as Drug Carriers. Weinheim; Wiley-VCH 2009; 349-59.

[63] Szakacs G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, Bussey KJ, *et al.* Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. Cancer Cell 2004; 6: 129-37.

[64] Labute P. A widely applicable set of descriptors. J Mol Graph Model 2000; 18: 464-77.

[65] Kaiser D, Terfloth L, Kopp S, Schulz J, de Laet R, Chiba P, *et al.* Self-organizing maps for identification of new inhibitors of P-glycoprotein. J Med Chem 2007; 50: 1698-702.

[66] Zdrazil BP, C. ; Ecker, G. F. NCI60 screening data - a versatile tool for in silico models predicting substrate properties for ABC-transporter. At: ACS 234th National Meeting & Exposition. Boston 2007

[67] Sakurai A, Onishi Y, Hirano H, Seigneuret M, Obanayama K, Kim G, *et al.* Quantitative structure--activity relationship analysis and molecular dynamics simulation to functionally validate nonsy-nonymous polymorphisms of human ABC transporter ABCB1 (P-glycoprotein/MDR1). Biochemistry 2007; 46: 7678-93.

[68] Rosenberg MF, Kamis AB, Callaghan R, Higgins CF, Ford RC. Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding. J Biol Chem 2003; 278: 8294-9.

[69] Rosenberg MF, Callaghan R, Modok S, Higgins CF, Ford RC. Three-dimensional structure of P-glycoprotein: the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state. J Biol Chem 2005; 280: 2857-62.

[70] Rosenberg MF, Mao Q, Holzenburg A, Ford RC, Deeley RG, Cole SP. The structure of the multidrug resistance protein 1 (MRP1/ABCC1). crystallization and single-particle analysis. J Biol Chem 2001; 276: 16076-82.

[71] McDevitt CA, Collins RF, Conway M, Modok S, Storm J, Kerr ID, *et al.* Purification and 3D structural analysis of oligomeric human multidrug transporter ABCG2. Structure 2006; 14: 1623-32.

18

Using Structural and Mechanistic Information to Design Novel Inhibitors/Substrates of P-Glycoprotein

Freya Klepsch[1], Thomas Stockner[2], Thomas Erker[1], Markus Müller[3], Peter Chiba[2,]* and Gerhard F. Ecker[1,]*

[1]*University of Vienna, Department of Medicinal Chemistry, Althanstrasse 14, 1090 Wien, Austria, *[2]*Medical University of Vienna, Institute of Medical Chemistry, Waehringer Strasse 10, 1090 Wien, Austria, *[3]*Medical University of Vienna, Department of Clinical Pharmacology, Währinger Gürtel 18-20, 1090 Wien, Austria*

Abstract: Design of inhibitors of P-glycoprotein still represents a challenging task for medicinal chemists. The polyspecificity of the transporter combined with the limited structural information renders rational drug design approaches rather ineffective. Within this article we will exemplify how recent insights into structure and mechanism of P-glycoprotein may aid in design of potent inhibitors.

Keywords: MDR, multidrug resistance, P-gp, P-glycoprotein, CYP, cytochrome P450-3A4, SPECT, single-photon emission computed tomography, PET, positron emission tomography, BCRP, breast cancer related protein, BBB, blood brain barrier, CNS, central nervous system, TMD, transmembrane domain, NBD, nucleotide binding domain, ABC, ATP-binding cassette, CPPI, cyclic peptide P-glycoprotein inhibitor, PLIF, protein-ligand interaction fingerprints, SERT, serotonine reuptake transporter, DAT, dopamine transporter, NET, norepinephrine transporter, SAR structure activity relationship.

INTRODUCTION

P-glycoprotein (P-gp), encoded by the MDR1 gene, is a transmembrane, ATP-driven transporter that acts as a drug efflux pump. P-gp is physiologically expressed in epithelial cells of the kidney, liver, pancreas, and colon, underscoring its role in maintaining concentration gradients of (toxic) compounds at physiologically important barriers. It is also expressed at the Blood-Brain Barrier (BBB) where it is thought to act as an active defence mechanism by restricting the penetration of lipophilic substances into the brain. However, the attribute that is most often associated with P-gp is Multi Drug Resistance (MDR), which is the ability of cancer cells to develop resistance to multiple classes of structurally and functionally diverse drugs [1]. MDR represents a major impediment for successful cancer- and also antimicrobial therapy. Therefore it is necessary to identify inhibitors of drug efflux transporters that are able to re-establish drug sensitivity of resistant cells. Several inhibitors of P-gp related MDR have been evaluated in clinical studies, some of which progressed to phase III. Nevertheless, none of these compounds has reached the market so far, raising concerns about the druggability of P-gp [2]. However, increasing knowledge about structural requirements that make compounds P-gp substrates allows designing in or designing out substrate properties. In this review we will highlight the potential impact of recent achievements in structural genomics of drug transporter on the design of inhibitors and substrates of P-gp.

P-gp-A DRUGGABLE TARGET?

The discovery of the fundamental role of P-gp in tumour cell MDR in 1976 raised high hopes for the clinical treatment of drug resistant tumours [3]. Once the role of P-gp as a drug efflux pump was understood, the concept of coadministration of cytotoxic drugs with inhibitors of P-gp was developed. However, since the identification of the P-gp inhibitory potential of verapamil [4] more than 25 years have passed and still no P-gp inhibitor has entered the market. In drug discovery and development the average time between patent application and market introduction is 10-12 years, pointing towards a potential problem with the druggability of P-gp. Within the past decade considerable progress has been made in unravelling the physiological function of P-gp and other ABC-transporters. Results clearly demonstrate the multiple involvement of several members of the ABC-transporter family in physiological and pathophysiological processes such as drug resistance, steroid transport, bile acid transport, and brain uptake. However, their major role in drug-uptake, -disposition and –elimination [5] as well as the substantial overlap with the substrate profiles of several cytochrome P450 isoenzymes [6] also raised concerns about their druggability and might render them off-targets rather than classical targets suited for drug therapy.

Numerous compounds have been tested clinically as P-gp inhibitors with the aim to overcome MDR in tumours. Briefly, first generation inhibitors relied on already marketed drugs, which, due to the polyspecificity of P-gp, also showed activity as drug efflux inhibitors. However, all of them failed because of their inherent pharmacological activity which became dose limiting for their application as MDR modulators [7].

Also second generation modulators, which either followed the concept of chiral switching (dexverapamil, dexniguldipin) or in which the original pharmacological activity was designed out (e.g. the cyclosporine analogue valspodar and the FK506

*Address correspondence to these authors at the Medical University of Vienna, Institute of Medical Chemistry, Waehringer Strasse 10, 1090 Wien, Austria; Tel: +43-1-4277-60806; Fax: +43-1-4277-9608; E-mail: peter.chiba@meduniwien.ac.at; gerhard.f.ecker@univie.ac.at

derivative biricodar), could not fulfil the expectations. These compounds interfered with the metabolism of anti-cancer drugs at the level of cytochrome P450-3A4 (CYP), resulting in prolonged half life and increased plasma levels of the anticancer drug used in clinical coadministration protocols. Dose-reduction protocols failed because of large interindividual variation in anticancer drug metabolism.

In consequence, third generation modulators of P-gp, such as tariquidar, zosuquidar and elacridar have been developed to avoid the interference of coadministered anticancer drugs and P-gp inhibitors at the level of CYP-450 3A4. These third generation modulators, however, did not show a statistically significant benefit in clinical studies. In conclusion, the therapeutic concept of coadministration of anticancer drugs and P-gp inhibitors might not meet the high rising expectations spurred by early *in vitro* experiments and preliminary clinical studies.

Within the past years the concept of developing and using biomarkers as specific molecular probes in assessing and visualising functional P-gp has become increasingly important. Radiolabeled substrates of P-gp also aid in understanding the *in vivo* function of this drug efflux pump under physiological and pathophysiological conditions. Both Single-Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) ligands have been synthesized to facilitate non-invasive imaging techniques both for assessing the localization and function of P-gp [8]. Radiolabeled P-gp substrates should be selective, produce a large signal after P-gp inhibition, and generate only few or, in an ideal case, no radiometabolites. Besides the SPECT tracer [^{99m}Tc]sestamibi, the PET ligands [^{11}C]carvedilol, [^{11}C]colchicine, [^{11}C]daunorubicin, [^{11}C]loperamide, [^{11}C]-*N*-desmethyl-loperamide, [^{18}F] paclitaxel, and [^{11}C]verapamil were developed [8-10]. R*ac*-[^{11}C]-verapamil, *R*-[^{11}C]-verapamil, [^{11}C]-loperamide, and [^{11}C]-*N*-desmethyl-loperamide were also used in humans. However, low brain uptake is a major limitation of currently available radiotracers like [^{11}C]loperamide, or (*R*)-[^{11}C]verapamil which are high-affinity substrates of P-gp. A promising alternative might be the use of radiolabeled non-transported inhibitors with nanomolar binding affinity. To date several third-generation Pgp inhibitors are under investigation as potential PET tracers, like [^{11}C]laniquidar, [^{11}C]MC-18 (6,7-dimethoxy-2-{3-[4-methoxy-3,4-dihydro-2*H*-naphthalen-(1*E*)-ylidene]-propyl}-1,2,3,4-tetrahydro-isoquinoline), [^{11}C]elacridar and [^{11}C]tariquidar [11-14]. For the *in vivo* use of a prospective radioligand a B_{max} (target expression)/K_d (equilibrium dissociation constant) ratio of >1 is considered to be required [15]. A B_{max}/K_d ratio of ~300 for tariquidar and elacridar therefore indicates that a suitable P-gp-specific signal may be generated by these tracers *in vivo*. Very recently [^{11}C]elacridar has been synthesized and evaluated in small-animal PET studies for its ability to assess the distribution of P-gp at the BBB. Studies on both *Mdr1a/b*$^{(-/-)}$ and *Bcrp1*$^{(-/-)}$ mice demonstrated that [^{11}C]-elacridar shows specificity for Pgp over BCRP and thus might represent a versatile tool compound for visualisation of cerebral P-gp [12].

In light of the instrumental role, which P-gp plays in maintaining the BBB, inhibition of the transporter has been advocated as a strategy to improve delivery of coadministered drugs to their site of action in the CNS [16], or to design out P-gp substrate properties of drugs. In a recent review, Oprea and colleagues also outline the concept of considering P-gp as a target with respect to CNS side effects of drugs [17]. Analysis of 64 launched histamine H1 receptor antagonists showed that two major central nervous side effects of these drugs - sedation and orthostatic hypotension - can be explained on basis of their interaction with P-gp. This further emphasizes the important role of P-gp for bioavailability and distribution of drugs as well as for their toxicity and side effects. This also prompted several authors to develop in silico prediction models for substrates/non substrates of P-gp [18]. However, due to lack of structural information all models developed so far relied on ligand information only. With the recent publication of the structure of mouse P-gp in complex with two enantiomeric ligands the situation may have changed and structure-based approaches may become an alternative [19].

THE IMPACT OF STRUCTURAL GENOMICS

The general architecture of ABC transporters and P-gp in particular comprise two transmembrane domains (TMDs) and two nucleotide binding domains (NBDs). Since the NBDs are responsible for the energy supply of the transporters, they are highly conserved among all ABC transporters. Sequence identity is generally lower for the TMDs, which are responsible for substrate specificity, than for the NBDs, which by ATP-binding and hydrolysis provide the energy for substrate transport. During the catalytic cycle P-gp adopts at least three different states, which are associated with ATP-binding and subsequent hydrolysis (reviewed in [20]; Fig. (**1**)). The apo or "open-inward" facing conformation is considered the ground state. In this conformation the protein shows an inverted "V" shape which is open to the cytosol and competent to bind substrates. The second state shows a more closed arrangement of the TMDs and the NBDs and the third conformation which is suggested by crystal structures of bacterial homodimeric transporters (Sav1866 and MsbA) is the nucleotide-bound form, which is considered open to the extracellular space.
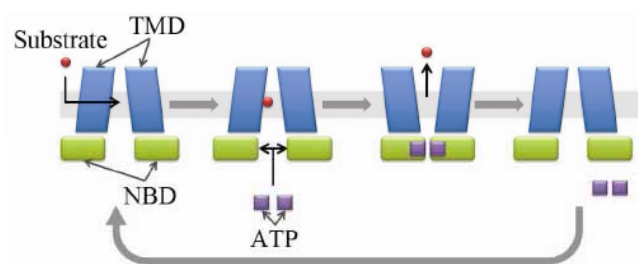
Fig. (1). Schematic illustration of the catalytic cycle of ABCB1. The two different conformations are depicted before and after drug binding.

PROTEIN HOMOLOGY MODELS OF P-gp

P-gp and analogous ABC transporters are embedded in the membrane, which poses a considerable challenge for crystallisation and subsequent structural elucidation at atomic resolution. Therefore, protein homology modeling based on templates of bacterial homologues was the method of choice for first structure-based studies. However, earlier structures of

bacterial MsbA published in 2001 [21], 2003 [22] and 2005 [23] were retracted, which, in consequence, rendered the respective homology models based on these structures obsolete. Due to its high resolution the crystal structure of the *Staphylococcus aureus* transporter SAV1866 (PDB code: 2HYD, resolution: 3.00 Å) [24] in the nucleotide bound "outward-facing" conformation served as a template for a number of modeling attempts [25-28]. Several structures of MsbA obtained in different catalytic states were also resolved [29]. In March 2009 the first X-ray structure of a eukaryotic ABC efflux transporter, mouse MDR1A, was published [19] (PDB code: 3G5U, resolution: 3.8Å). It serves as a good template for homology modelling, because it has an 87% sequence identity with human P-gp [30]. Two additional structures were published at lower resolution in complex with two enantiomeric cyclic peptide inhibitors (CPPIs; QZ59-RRR and QZ59-SSS; PDB-codes: 3G60, 3G61 Fig. (**2**)) [19]. This information may pave the way for first concrete insights into ligand binding of P-gp.
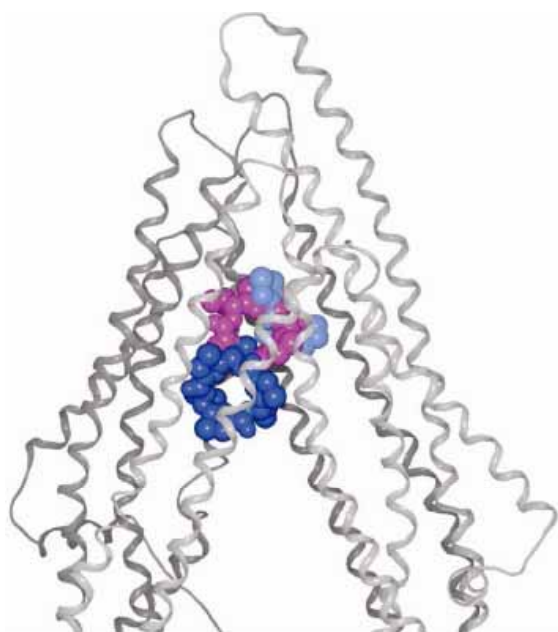
Fig. (2). ABCB1 co-crystallized with cyclic P-gp inhibitors (CPPIs) QZ59-RRR (magenta) QZ59-SSS (blue).

BINDING SITES

The functional unit of ABC-transporters minimally consists of two TM and two NB domains. Earlier studies showed that ABC transporters consisting only of the TMDs were able to bind ligands in the absence of NBDs [31]. This corroborated the notion that drug binding occurs in the membrane spanning region.

Polyspecificity is considered to rely on the presence of a complement of redundant amino acid residues which are able to react with different ligands within an expansive binding pocket with partially overlapping binding sites. Numerous experimental studies were performed to characterise and analyse drug binding sites of P-gp, comprising

among others cysteine and arginine scanning and photoaffinity labeling [32, 33] (reviewed in [20]). The general thinking is that P-gp possesses interaction regions, which lie at the TMD/TMD interface with prominent contributions from TM helices 6 and 12. This conclusion is also strengthened by the structures of ABCB1 co-crystallized in the presence of the CPPIs QZ-59-RRR and QZ59-SSS [19]. The structures show three poses of the enantiomeric inhibitors in what seems to be a rather expansive binding pocket. Partially overlapping interacting amino acid residues are identified, some of which are identical with those involved in rhodamine 123 and verapamil binding [34, 35].

LIGAND DOCKING

In silico docking of ligands has become a standard tool in virtual screening protocols. However, this method has clear limits, which are quite often neglected. First, the docking process is separated into two consecutive steps: placement and scoring. Whereas initial ligand placement into putative binding sites is implemented in several software packages, scoring of the poses still remains a challenge and is subject of intense research. An additional difficulty is that binding of substrates to P-gp has been described to follow an induced fit mechanism [32] and therefore protein flexibility both with respect to side chain flexibility and to backbone movements has to be considered. In case of multidrug transporters, these challenges are surmounted by the fact that the transporter is promiscuous and might possess redundant interaction regions with similar affinities. Finally, one must consider that a structure of human P-gp is not available and that the resolution of the available mouse template structures is low. Thus, in our view conventional docking/scoring approaches have to be validated carefully when applied to P-gp or any other multispecific ABC-efflux transporter.

Recent docking experiments with a series of quinazolinones place the compounds in positions which are comparable with those of the CPPIs (Fig. (**3**)) [30]. The docking poses suggest a hydrogen bonding interaction between the ligands and amino acid residue Tyr307 (TM5). Protein-ligand interaction fingerprints (PLIF) reveal that residues Phe336 (TM6), Tyr953 (TM11) and Phe957 (TM11) also form contact interactions.

Similar results were also obtained in our group when performing docking studies with a homology model of ABCB1 and propafenone derivatives [36]. Our results also show interactions with TM5, TM6 and TM11, confirming the concept of a large binding pocket at the TMD interface. This potentially indicates overlapping quinazolinone and propafenone binding sites.

IS TARGET-BASED INHIBITOR DESIGN POSSIBLE?

Target-based drug discovery strongly depends on the availability of high quality, high resolution protein structures, preferably with bound ligands. This might allow identification of the major interacting amino acids and also the positioning of the amino acid side chains in the ligand bound conformation. In principle, one of these requirements is fulfilled by the mouse P-gp structure as it is resolved in complex with ligands QZ59-RRR and QZ59-SSS. However, the resolution is only 3.8Å, which is too low for an unambiguous assignment of side
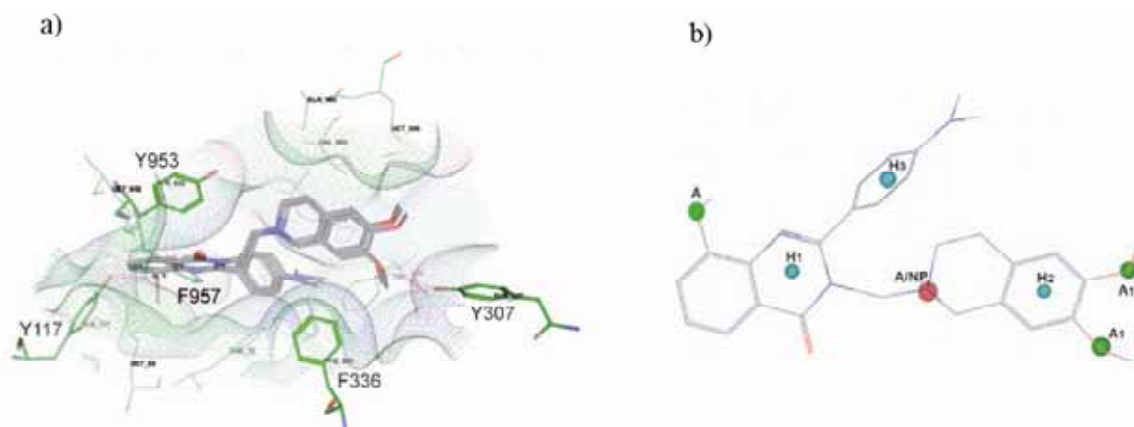
Fig. (3). a) Docking poses of quinazolinones in a human ABCB1 homology model, **b**) Pharmacophore model (taken from [30]).

chain positions; and the ligand in the "upper" site is only partially resolved. In addition, the linker region between NBD1 and TMD2, the phosphorylation of which influences ATPase stimulation by several substrates, is missing [37].

In docking experiments carried out in our group a large number of ligand poses was obtained within the central cavity of the transporter [36]. However, proper ranking of the poses remained challenging. The encountered difficulties might have several origins: (i) data support the notion that distinct P-gp ligands bind to preferred sub-sites within the central cavity [19, 38], (ii) the binding region is promiscuous and (iii) it has been shown that more than one ligand can bind to P-gp simultaneously [39] adding to the complexity of the docking task. In addition, (iv) the classical lock-and-key principle might not be applicable to P-gp and related MDR transporters. Surmounting these challenges will most likely require either refined (target tailored) scoring functions or a workflow which is less dependent on scoring functions.

In light of our own structure-based design studies on off-targets and transporters such as P-gp, the hERG potassium channel, the GABAA receptor, and the SERT, DAT and NET transporters we are actively pursuing a concept based on experiment guided docking and common scaffold clustering. In this approach a small series of structurally analogous ligands with a clear SAR pattern are docked applying an exhaustive sampling of the pose space. Subsequently, these poses are clustered according to RSMD of the basic scaffold of the ligands. Only those clusters which contain poses for all ligands are considered for further evaluation. Prioritisation of these clusters is achieved by a combination of consensus scoring, data from mutagenesis experiments, information from SAR and pharmacophore modeling. First successful implementations for P-gp, SERT and the GABAA receptor have been presented recently [36, 40, 41]. However, in case of a highly diverse set of compounds, which normally is the basis for structure-based screening attempts, this approach is no longer feasible. Thus, our attempts to classify a large and diverse data set of 1.400 compounds, taken from the NCI-60 screening set [42], into substrates and non-substrates by

docking the whole library into a homology model of P-gp completely failed. No scoring function was able to properly separate substrates from non-substrates (data not shown).

In conclusion, a routine application of target-based design methods, such as virtual screening or de novo design, is currently not suited for identification of new inhibitors of P-gp or other MDR ABC-efflux transporters of pathophysiological relevance. This is mainly due to the multispecific interaction pattern and the low resolution of available crystal structures.

MECHANISM BASED INHIBITOR DESIGN

Mechanistic studies of P-gp clearly indicate the importance of ATP binding and hydrolysis for substrate transport. Thus, an additional possibility to design inhibitors of P-gp would be to develop compounds which are targeted towards the ATP-binding site. These are highly conserved, not membrane embedded and high resolution structures are available. This prompted several authors to perform docking studies into the ATP-binding sites. For a recent overview please refer to Klepsch *et al.* [43]. However, up to now it is not clear how to address the problem of selectivity for a single ABC-transporter as compared to other ATP-binding proteins, including the large family of tyrosine kinases (TKs). Notably, several TK-inhibitors such as nilotinib, dasatinib and bosutinib have been shown to be substrates of P-gp and ABCG2 rather than ATP-binding site inhibitors [44]. This indicates clear differences in the shape and geometry of different ATP-binding sites. Thus, design of competitive ATP inhibitors might be a versatile strategy for development of new P-gp inhibitors. Moreover, the knowledge gained in the kinase field might be directly applicable to other ATPases [45].

A completely different and highly promising approach might be the design of ligands with long residence times. There is increasing evidence that compounds with long residence times might be advantageous with respect to both duration of the pharmacological effect and the target selectivity [46]. An analogous concept for substrate selection has very recently also been hypothesised by Ernst *et al.* for the ABC-transporter Pdr5 [47]. Two compounds showing equal binding affinity might be transported with different efficiencies. If the transporter remains in the apo state only for a short period of time, a compound with slow on and off-kinetics might be

transported only once, whereas a compound with fast kinetics may be transported several fold more efficiently. Once the kinetics of the transporter are fully understood on a molecular level, this might pave the way for designing in and designing out substrate properties, a hot topic especially in CNS and anticancer drug development.

CONCLUSIONS AND OUTLOOK

The past few years showed considerable progress in our structural knowledge on ABC-transporters. Templates for several states of the catalytic cycle are available now for protein homology modeling and the recent publication of the mouse P-gp structure bound to two enantiomeric ligands revealed first insights into the molecular basis of drug-transporter interaction [48]. However, the resolution of these structures is too low for structure-based drug design studies such as docking of large compound libraries. Furthermore, crystal structures represent snapshots of a highly flexible protein which undergoes major conformational rearrangements when progressing through the transport cycle. Thus both, structures with remarkably increased resolution as well as snapshots of different states of the transport cycle are needed to better understand the molecular basis of ligand-protein interaction and to enable the use of structure-based drug design approaches for identification of new ligands of P-gp.

ACKNOWLEDGEMENT

REFERENCES

[1]　Gottesman, M.M.; Fojo, T.; Bates, S.E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer*, **2002**, *2*(1), 48-58.

[2]　Ecker, G.F.; Chiba, P. In: *Transporters as Drug Carriers*, G. F. Ecker and P. Chiba, Eds.; Wiley-VCH: Weinheim, **2009**; Vol. *44*, pp. 349-359.

[3]　Juliano, R.L.; Ling, V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta*, **1976**, *455*(1), 152-162.

[4]　Goldberg, H.; Ling, V.; Wong, P.Y.; Skorecki, K. Reduced cyclosporin accumulation in multidrug-resistant cells. *Biochem. Biophys. Res. Commun.*, **1988**, *152*(2), 552-558.

[5]　Szakacs, G.; Varadi, A.; Ozvegy-Laczka, C.; Sarkadi, B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov. Today*, **2008**, *13*(9-10), 379-393.

[6]　Benet, L.Z.; Cummins, C.L. The drug efflux-metabolism alliance: biochemical aspects. *Adv. Drug Deliv. Rev.*, **2001**, *50* (Suppl 1), S3-11.

[7]　Raderer, M.; Scheithauer, W. Clinical trials of agents that reverse multidrug resistance. A literature review. *Cancer*, **1993**, *72*(12), 3553-3563.

[8]　Kannan, P.; John, C.; Zoghbi, S.S.; Halldin, C.; Gottesman, M.M.; Innis, R.B.; Hall, M.D. Imaging the function of P-glycoprotein with radiotracers: pharmacokinetics and *in vivo* applications. *Clin. Pharmacol. Ther.*, **2009**, *86*(4), 368-377.

[9]　Sharma, V. Radiopharmaceuticals for assessment of multidrug resistance P-glycoprotein-mediated drug transport activity. *Bioconjug. Chem.*, **2004**, *15*(6), 1464-1474.

[10]　Elsinga, P.H.; Hendrikse, N.H.; Bart, J.; Vaalburg, W.; van Waarde, A. PET Studies on P-glycoprotein function in the blood-brain barrier: how it affects uptake and binding of drugs within the CNS. *Curr. Pharm. Des.*, **2004**, *10*(13), 1493-1503.

[11]　Bauer, F.; Mairinger, S.; Dörner, B.; Kuntner, C.; Stundner, G.; Bankstahl, J.P.; Wanek, T.; Stanek, J.; Müller, M.; Langer, O.; Erker, T. Synthesis and μPET evaluation of the radiolabelled P-glycoprotein inhibitor [11C]tariquidar. *Eur. J. Nucl. Med. Mol. Imaging*, **2009**, *36*, 222.

[12]　Dörner, B.; Kuntner, C.; Bankstahl, J.P.; Bankstahl, M.; Stanek, J.; Wanek, T.; Stundner, G.; Mairinger, S.; Loscher, W.; Muller, M.; Langer, O.; Erker, T. Synthesis and small-animal positron emission tomography evaluation of [11C]-elacridar as a radiotracer to assess the distribution of P-glycoprotein at the blood-brain barrier. *J. Med. Chem.*, **2009**, *52*(19), 6073-6082.

[13]　Luurtsema, G.; Schuit, R.C.; Klok, R.P.; Verbeek, J.; Leysen, J.E.; Lammertsma, A.A.; Windhorst, A.D. Evaluation of [11C]laniquidar as a tracer of P-glycoprotein: radiosynthesis and biodistribution in rats. *Nucl. Med. Biol.*, **2009**, *36*(6), 643-649.

[14]　van Waarde, A.; Ramakrishnan, N.K.; Rybczynska, A.A.; Elsinga, P.H.; Berardi, F.; de Jong, J.R.; Kwizera, C.; Perrone, R.; Cantore, M.; Sijbesma, J.W.; Dierckx, R.A.; Colabufo, N.A. Synthesis and preclinical evaluation of novel PET probes for P-glycoprotein function and expression. *J. Med. Chem.*, **2009**, *52*(14), 4524-4532.

[15]　Eckelman, W.C.; Mathis, C.A. Targeting proteins *in vivo*: *in vitro* guidelines. *Nucl. Med. Biol.*, **2006**, *33*(2), 161-164.

[16]　Miller, D.S.; Bauer, B.; Hartz, A.M. Modulation of P-glyco-protein at the blood-brain barrier: opportunities to improve central nervous system pharmacotherapy. *Pharmacol. Rev.*, **2008**, *60*(2), 196-209.

[17]　Broccatelli, F.; Carosati, E.; Cruciani, G.; Oprea, T.I. Transporter-mediated efflux influences CNS side effects: ABCB1, from antitarget to target. *Mol. Inf.*, **2010**, *29*, 16-26.

[18]　Demel, M.A.; Kramer, O.; Ettmayer, P.; Haaksma, E.E.; Ecker, G.F. Predicting ligand interactions with ABC transporters in ADME. *Chem. Biodivers*, **2009**, *6*(11), 1960-1969.

[19]　Aller, S.G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P.M.; Trinh, Y.T.; Zhang, Q.; Urbatsch, I.L.; Chang, G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, **2009**, *323*(5922), 1718-1722.

[20]　Seeger, M.A.; van Veen, H.W. Molecular basis of multidrug transport by ABC transporters. *Biochim. Biophys. Acta*, **2009**, *1794*(5), 725-737.

[21]　Chang, G.; Roth, C.B. Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science*, **2001**, *293*(5536), 1793-1800.

[22]　Chang, G. Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. *J. Mol. Biol.*, **2003**, *330*(2), 419-430.

[23]　Reyes, C.L.; Chang, G. Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science*, **2005**, *308*(5724), 1028-1031.

[24]　Dawson, R.J.; Locher, K.P. Structure of a bacterial multidrug ABC transporter. *Nature*, **2006**, *443*(7108), 180-185.

[25]　Becker, J.P.; Depret, G.; Van Bambeke, F.; Tulkens, P.M.; Prevost, M. Molecular models of human P-glycoprotein in two different catalytic states. *BMC Struct. Biol.*, **2009**, *9*(3).

[26]　Globisch, C.; Pajeva, I.K.; Wiese, M. Identification of putative binding sites of P-glycoprotein based on its homology model. *Chem. Med.Chem.*, **2008**, *3*(2), 280-295.

[27]　Stockner, T.; de Vries, S.J.; Bonvin, A.M.; Ecker, G.F.; Chiba, P. Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. *FEBS J.*, **2009**, *276*(4), 964-972.

[28]　O'Mara, M.L.; Tieleman, D.P. P-glycoprotein models of the apo and ATP-bound states based on homology with Sav1866 and MalK. *FEBS Lett.*, **2007**, *581*(22), 4217-4222.

[29]　Ward, A.; Reyes, C.L.; Yu, J.; Roth, C.B.; Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl. Acad. Sci. USA*, **2007**, *104*(48), 19005-19010.

[30]　Pajeva, I.K.; Globisch, C.; Wiese, M. Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. *Chem. Med. Chem.*, **2009**, *4*(11), 1883-1896.

[31]　Loo, T.W.; Clarke, D.M. The transmembrane domains of the human multidrug resistance P-glycoprotein are sufficient to mediate drug binding and trafficking to the cell surface. *J. Biol. Chem.*, **1999**, *274*(35), 24759-24765.

[32]　Loo, T.W.; Clarke, D.M. Mutational analysis of ABC proteins. *Arch. Biochem. Biophys.*, **2008**, *476*(1), 51-64.

[33]　Shilling, R.A.; Venter, H.; Velamakanni, S.; Bapna, A.; Woebking, B.; Shahi, S.; van Veen, H.W. New light on multidrug binding by an

ATP-binding-cassette transporter. *Trends Pharmacol. Sci.*, **2006**, *27*(4), 195-203.

[34] Loo, T.W.; Bartlett, M.C.; Clarke, D.M. Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket. *Biochem. J.*, **2006**, *399*(2), 351-359.

[35] Loo, T.W.; Clarke, D.M. Identification of residues in the drug-binding site of human P-glycoprotein using a thiol-reactive substrate. *J. Biol. Chem.*, **1997**, *272*(51), 31945-31948.

[36] Klepsch, F.; Chiba, P.; Ecker, G.F. Docking studies to identify the binding mode of propafenones to p-glycoprotein. *Drugs Future*, **2009**, *34*(Suppl A), 164.

[37] Szabo, K.; Bakos, E.; Welker, E.; Muller, M.; Goodfellow, H.R.; Higgins, C.F.; Varadi, A.; Sarkadi, B. Phosphorylation site mutations in the human multidrug transporter modulate its drug-stimulated ATPase activity. *J. Biol. Chem.*, **1997**, *272*(37), 23165-23171.

[38] Schumacher, M.A.; Miller, M.C.; Brennan, R.G. Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *EMBO J.*, **2004**, *23*(15), 2923-2930.

[39] Loo, T.W.; Bartlett, M.C.; Clarke, D.M. Simultaneous binding of two different drugs in the binding pocket of the human multidrug resis-tance P-glycoprotein. *J. Biol. Chem.*, **2003**, *278*(41), 39706-39710.

[40] Richter, L.; Ernst, M.; Sieghart, W.; Ecker, G.F. Combining multiple approaches in pose evaluation - Application to the GABAA receptor. *Sci. Pharm.*, **2009**, *77*(1), 181.

[41] Weissensteiner, R.; Sitte, H.H.; Freissmuth, M.; Ecker, G.F. Docking of substrates and inhibitors into new models of neurotransmitter: Sodium symporter. *Sci. Pharm.*, **2009**, *77*(1), 205.

[42] Szakacs, G.; Annereau, J.P.; Lababidi, S.; Shankavaram, U.; Arciello, A.; Bussey, K.J.; Reinhold, W.; Guo, Y.; Kruh, G.D.; Reimers, M.; Weinstein, J.N.; Gottesman, M.M. Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. *Cancer Cell*, **2004**, *6*(2), 129-137.

[43] Klepsch, F.; Jabeen, I.; Chiba, P.; Ecker, G.F. Pharmacoinformatic approaches to design natural product type ligands of ABC-transporters. *Curr. Pharm. Des.*, **2010**, *16*(15), 1712-1752.

[44] Hegedus, C.; Ozvegy-Laczka, C.; Apati, A.; Magocsi, M.; Nemet, K.; Orfi, L.; Keri, G.; Katona, M.; Takats, Z.; Varadi, A.; Szakacs, G.; Sarkadi, B. Interaction of nilotinib, dasatinib and bosutinib with ABCB1 and ABCG2: implications for altered anti-cancer effects and pharmacological properties. *Br. J. Pharmacol.*, **2009**, *158*(4), 1153-1164.

[45] Chene, P. The ATPases: a new family for a family-based drug design approach. *Expert Opin. Ther. Targets*, **2003**, *7*(3), 453-461.

[46] Copeland, R.A.; Pompliano, D.L.; Meek, T.D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.*, **2006**, *5*(9), 730-739.

[47] Ernst, R.; Kueppers, P.; Stindt, J.; Kuchler, K.; Schmitt, L. Multidrug efflux pumps: substrate selection in ATP-binding cassette multidrug efflux pumps--first come, first served? *FEBS J.*, **2010**, *277*(3), 540-549.

[48] Pajeva, I.K.; Globisch, C.; Wiese, M. Comparison of the inward- and outward-open homology models and ligand binding of human P-glycoprotein. *FEBS J.*, **2009**, *276*(23), 7016-7026.

24

Impact of the Recent Mouse P-Glycoprotein Structure for Structure-Based Ligand Design

Freya Klepsch[a] and Gerhard F. Ecker*[a]
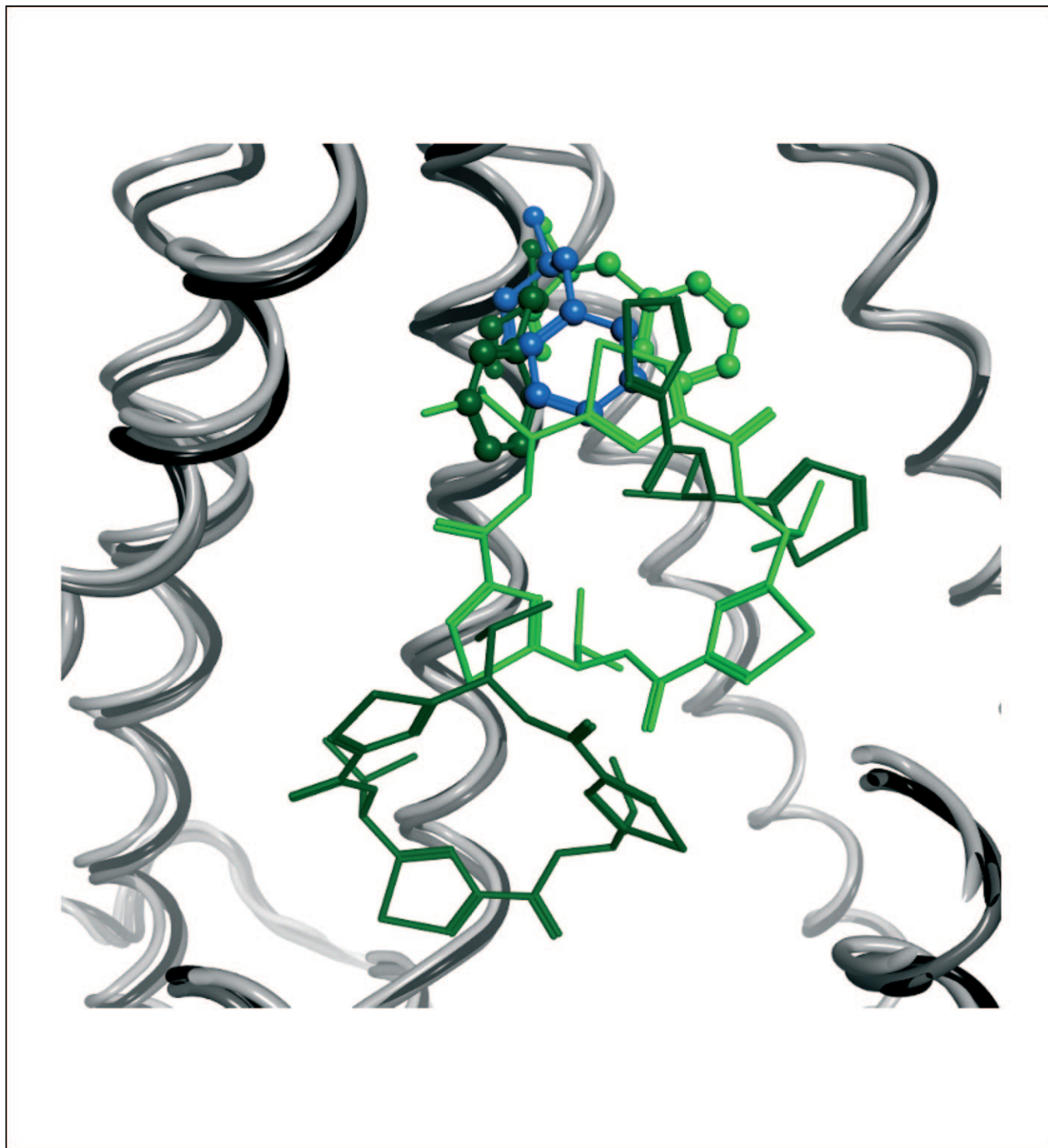
Abstract: P-Glycoprotein (P-gp), a transmembrane, ATP-dependent drug efflux transporter, has attracted considerable interest both with respect to its role in tumour cell multidrug resistance and in absorption-distribution and elimination of drugs. Although known since more than 30 years, the understanding of the molecular basis of drug/transporter interaction is still limited, which is mainly due to the lack of structural information available. However, within the past decade X-ray structures of several bacterial homologues as well as very recently also of mouse P-gp have become available. Within this review we give an overview on the current status of structural information available and on its impact for structure-based drug design.

Keywords: Proteins · Molecular modelling · Drug design, Ligands

1 Introduction

More than 30 years ago a membrane protein could be linked to the phenomenon of decreased uptake of vincristine in multidrug resistant tumour cells.[1] The protein was termed P-Glycoprotein (P-gp, ABCB1), as it shows a glycosylation site and seems to affect permeability of the cell membranes (P therefore accounts for permeability). P-gp is an ATP-driven, membrane bound protein transporting a wide variety of structurally and functionally diverse drugs out of tumor cells.[2] Only a few years later verapamil was identified as being able to reverse tumour cell drug resistance by blocking P-gp.[3] For the first time, the complex phenomenon of multidrug resistance was linked to a distinct protein, which could be targeted by drugs.[4] The verapamil induced restoration of cytotoxicity was also observed for anthracyclines and subsequently extended to other calcium channel blockers, such as benzothiazepines and 1,4-dihydropyridines.[5] Other therapeutically used drugs which also showed P-gp inhibitory activity comprised phenothiazines, quinine, tamoxifen and cyclosporine A.[6] However, 30 years and numerous clinical studies later there is still no compound on the market and there are serious concerns about the druggability of this ATP-dependent, transmembrane transport protein.[7] Within the past decade considerable progress has been made in unravelling the physiological function of P-gp and the other 47 human ABC-transporters (ABC accounts for ATP Binding Cassette).[7] P-gp and some of its analogs are expressed in the intestine, liver, kidney, and the blood-brain barrier and there is an overwhelming amount of data that clearly demonstrate their multiple involvement in drug-uptake, -disposition and -elimination[8] rendering them antitargets rather than classical targets suited for drug therapy.[9] One of the classic examples is the quinidine-digoxin interaction, where the P-gp inhibitor quinidine increased the digoxin absorption rate by 30%, the peak plasma concentration by 81%, and the plasma AUC by 77%.[10] Another example is the brain accumulation of a range of drugs (which normally do not enter the brain) observed in *mdr1a* double knockout mice.[11] Undoubtedly, P-pg is an important impediment for the entry of hydrophobic drugs into the brain. Recently also the breast cancer resistance protein (ABCG2) has been reported as playing a role in the brain uptake of a variety of compounds.[12] This physiological function of P-gp and ABCG2

at the blood-brain barrier challenges the medicinal chemists in two ways: (i) compounds which should not enter the brain should be designed as P-gp substrates; (ii) CNS active compounds must pass the blood-brain barrier and therefore should be poor substrates of P-gp. With this respect P-gp might now again be considered as target, as designing-in/designing-out substrate properties becomes a major task for optimising the pharmacokinetics and tissue distribution of drug candidates.[13] This is also emphasized in a very recent review of the International Transporter Consortium, which published guidelines how to include all these information on transporters in the drug development process.[14] At the beginning of the new millennium, the first X-ray structure of a bacterial homologue of P-gp was published,[15] and since last year the structure of the first murine ABC-transporter – mouse P-gp – is available.[16] The latter, for the very first time, showed an ABC-transporter complexed with a ligand. After a short overview on ligand-based studies we will outline the impact of these structural genomics attempts on our understanding of drug/transporter interaction and on consequences for structure-based inhibitor design.

2 Ligand-Based Studies

2.1 Inhibitor Design

P-glycoprotein and its homologues belong to the large group of membrane-bound proteins, which lack considerable structural information. Thus, inhibitor-design had to rely on classical ligand-based approaches, such as quantitative structure-activity relationship (QSAR) studies and pharmacophore modelling. Especially verapamil analogues, triazines, acridonecarboxamides, phenothiazines, thioxanthenes, flavones, dihydropyridines, propafenones and cyclosporine derivatives have been extensively studied, and the results are summarized in several excellent reviews.[17,18] Main results obtained from QSAR analyses clearly indicate the major importance of lipophilicity for high P-gp inhibito-

[a] F. Klepsch, G. F. Ecker
University of Vienna, Department of Medicinal Chemistry
Althanstraße 14, 1090 Wien, Austria
phone: +43-1-4277-55110; fax: +43-1-4277-9551
*e-mail: Gerhard.f.ecker@univie.ac.at
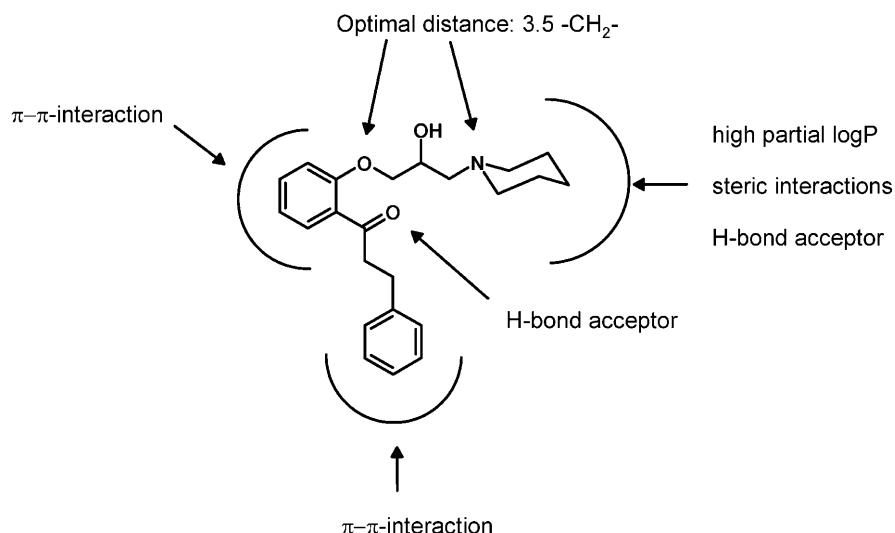
Figure 1. Summary of the results of structure–activity relationship studies on propafenone-type inhibitors of P-gp.

Freya Klepsch received a Master's Degree in Biotechnology with a specialization in Chemistry of Active Substances from the University of Applied Sciences, FH Campus Wien, in 2008 under the supervision of Prof. Ernst Urban. Now she is PhD student in the pharmacoinformatics research group at the Department of Medicinal Chemistry of the University of Vienna. Her work focuses on the evaluation of binding modes of small molecules in ABC transporters.

Gerhard Ecker heads the Pharmacoinformatics Research Group at the Department of Medicinal Chemistry, University of Vienna. He also coordinates the research focus "Computational Life Sciences" of the Faculty of Life Sciences. He studied pharmacy and received his doctorate in natural sciences from the University of Vienna under the supervision of Wilhelm Fleischhacker and Christian Noe. After his postdoctoral training at the group of J. Seydel in Borstel (Germany) he was appointed associate professor for medicinal chemistry at the University of Vienna in 1998 and full professor for pharmacoinformatics in 2009. He has published more than 90 articles related to SAR and QSAR studies on P-glycoprotein and his main scientific interests include pharmacoinformatic approaches to target drug efflux pumps, in silico high throughput screening methods for promiscuous targets and antitargets, and nonlinear methods in drug design. He is currently vice-president of the Austrian Pharmaceutical Society and President of the European Federation for Medicinal Chemistry. He also coordinates the EUROPIN PhD Programme in Pharmacoinformatics.

ry activity. However, this might be due to the fact that the interactions of the compounds with the transporter most probably take place within the membrane bilayer rather than in the intracellular compartment. This gave also rise to the notation of P-gp as hydrophobic vacuum cleaner.[19] Interestingly, Wiese and co-authors could convincingly show that the hydrophobicity of the ligand should be treated as a space directed property rather than as an overall feature of the compounds.[20] For a series of *ortho*-, *meta*-, and *para*-substituted aryloxypropanolamines we could demonstrate that the hydrophobic moment of the compounds is a better predictor than calculated log*P* values.[21] Both studies point towards a distinct drug/protein interaction rather than an unspecific hydrophobic attachment to the transporter. Moreover, in the group of propafenone analogs a clear SAR could be deduced (Figure 1).[22] Briefly, highly active compounds should have a highly lipophilic, but small substituent at the nitrogen atom (best up to now is 4-xylylpiperazine), should be *ortho*-substituted at the central aromatic ring and should have electron donating substituents at the two aromatic rings, preferentially located in *para*-position.

In addition to these extensive QSAR studies, also numerous pharmacophore models were published, both for substrates and for inhibitors.[23] Main features identified comprise hydrophobic, H-bond acceptor and positive ionisable. However, the models, although thoroughly validated and predictive in virtual screening runs, show only minor overlap in the spatial arrangement of the pharmacophoric features. This prompted several groups to postulate multiple binding sites at P-gp, which has been further evidenced by experimental data.[24] Finally, although consistent and statistically valid models were achieved, all these attempts could only minor contribute to the understanding of the function and the molecular basis of ligand-polyspecificity of P-gp and related transporters.

2.2 Prediction of Substrate Properties

As outlined above, P-gp is expressed in several organs, such as kidney, liver, intestine and also at the blood brain barrier (BBB). P-gp substrates therefore show poor oral absorption, enhanced renal and biliary excretion and usually do not enter the brain.[8,25] Furthermore, they are likely to be affected by the MDR phenotype and are thus not suitable as anticancer agents. This prompted the development of in silico models for predicting P-gp substrate properties of compounds of interest.[26] Models developed relied both on simple rule-based classifications and on more advanced methods such as support vector machines and artificial neural networks. First rule-based methods came up more than 10 years ago when Seelig postulated that substrate recognition is particularly based on one of two specific hydrogen bonding patterns. Her analysis suggests that substrates contain either two hydrogen bonding features in a spatial separation of approximately 2.5 Å or three hydrogen bonding features with a spatial separation of the outer two features of approximately 4.6 Å.[27,28] Later on, Didziapetris et al. formulated the "rule of four", which states that compounds with the number of hydrogen bond acceptors in a molecule $(N+O) \geq 8$, and a molecular weight $(MW) > 400$ Da and most acidic $pK_a > 4$ are likely to be ABCB1 substrates whereas compounds with $(N+O) \leq 4$, $MW < 400$, and most basic $pK_a < 8$ probably are non-substrates.[29] Even simpler than the rule of four is the "Gombar-Polli Molecular E-state (MolES) Rule", which states that molecules with MolES > 110 seem to be substrates and those with MolES < 48 seem to be non-substrates.[30] Finally, using only four simple ADMET descriptors (molecular weight, logP, positive ionizable and negative ionizable) Gleeson et al. could demonstrate that neutral or basic molecules showing a $MW > 400$ and a logP value > 4 are more likely to be transported by ABCB1 than acidic or zwitterionic compounds.[31]

One of the main problems of all these studies is the fact that data sets available are rather small and sometimes also inconsistent.[32] Few years ago the group of Gottesman published a comprehensive study where they correlated the cellular toxicity of 1400 selected compounds from the NCI60 screen with the mRNA levels of the 48 human ABC-transporter over the range of 60 human tumour cell lines.[33] An inverse correlation between transporter mRNA levels and compound toxicity indicates that a compound is a substrate for the respective transporter. Based both on this data set as well as on a set of 259 compounds compiled from the literature we used simple ADME-type descriptors (such as logP, number of rotatable bonds, number of H-bond donors and acceptors), van der Waals surface area descriptors[34] and 2D autocorrelation vectors as input matrix for several classification algorithms. When comparing binary QSAR and support vector machines, the latter gave more robust models with total accuracies in the range of 80%. Generally, the prediction of non-substrates performs better than those for substrates.[35] Very recently, based on a large data set provided by Boehringer Ingelheim, we developed a method for classifying rules (RuleFit) based on simple, interpretable physicochemical descriptors.[36] Interpretation of the best performing model indicates that P-gp substrates show a higher number of H-bond acceptors, more rotatable bonds and higher logP values than non-substrates. Although these features are quite general, the respective models showed a sensitivity of 81% and a specificity of 98% for an external test set. Thus, RuleFit modelling might be a versatile tool for establishing predictive and interpretable classification models also for other ABC-transporter.

3 ABC Transporter Structures

3.1 Topology of P-Glycoprotein

P-Glycoprotein is a pseudosymmetrical heterodimer where each monomer consists of a transmembrane (TM) as well as a nucleotide binding (NB) domain (Figure 2c). As the latter is responsible for the ATP-binding and hydrolysis it shows high sequence similarity throughout the ABC-transporter family. In contrast, the TM domains, which comprise 2×6 TM helices, are responsible for the respective substrate profile of the ABC-transporters and thus show only low sequence similarities among different transporters. The six helices of each TM domain (TMD) are connected by three extracellular and 2 intracellular loops. In addition, the intracellular loops comprise coupling helices which are responsible for the TMD-NB domain (NBD) interaction.

The principle of the topology described above is consistent throughout the whole family of human ABC-transporters. However, there are differences concerning the TMD and NBD arrangements. The ABCC transporter subfamily for instance possesses a third TMD at the N-terminus (TMD0) comprised of five helices, which is directly connected with TMD1.[37,38] Members of the ABCG subfamily, half transporters that undergo homodimerization to gain full functionality,[39] show an inverse topology with the NBD at the N-terminus and the TMD at the C-terminal end.[40]

As could be shown by cryo-electron microscopy and biochemical experiments, where P-gp was trapped in different states of the catalytic cycle (using the non-hydrolysable ATP analog AMP-PNP and ADP-V₁), P-gp undergoes large conformational changes during the catalytic cycle.[41] The mechanism of the energy driven drug transport, rendering the high-affinity into a low-affinity binding site, is currently hypothesized in two different ways (extensively reviewed in[42]): The ATP switch model interprets the NBD dimerisation as the power stroke that is needed for altering the affinity for the substrates. Upon ATP binding the substrate is released and the subsequent hydrolysis of both ATP molecules results in the regeneration of the initial apo state, where another drug molecule can bind.[25] The second theory favours the sequential occlusion and hydrolysis of ATP molecules, where the occlusion of one ATP molecule is

28
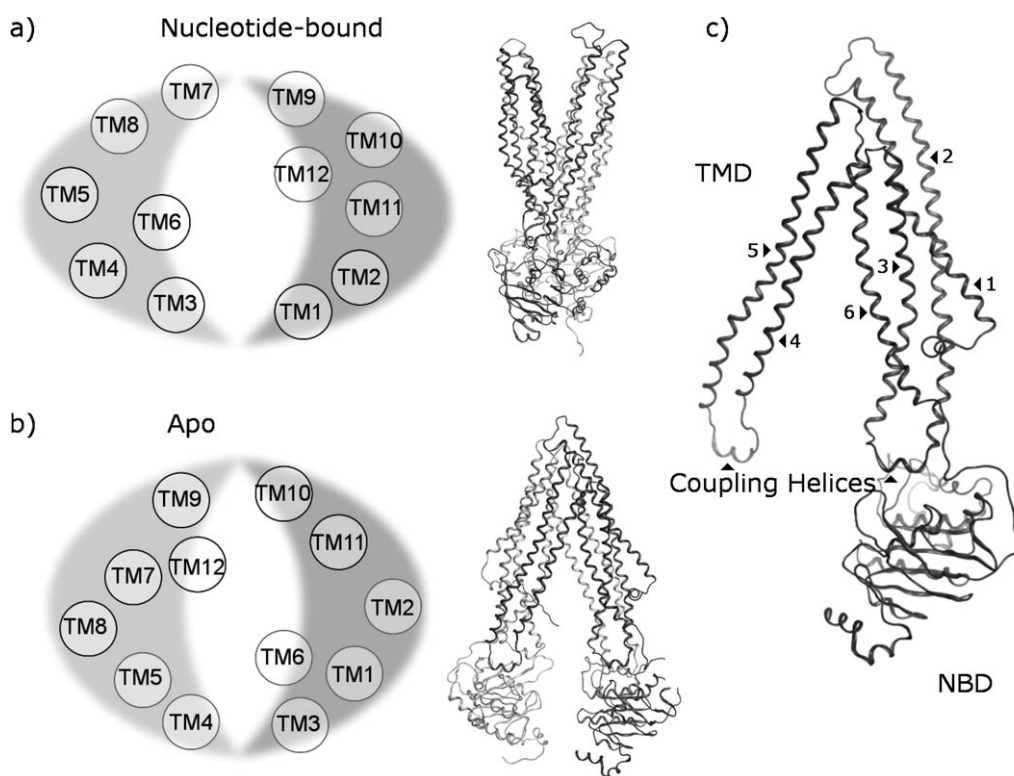
Figure 2. ABCB1 architecture: a) wing-like helix-arrangement of P-gp on the basis of the ADP-bound 2HYD structure; b) helix-arrangement on the basis of the apo-structure 3G5U; c) visualization of the different domains of the N-terminal half of P-gp.

sufficient for the conformational change resulting in drug release (Figure 3).[43] Therefore, only 1 mol ATP/mol P-gp is sufficient for drug transport, whereas hydrolysis of in total two ATP molecules is needed to recover the transporter.[44–46]
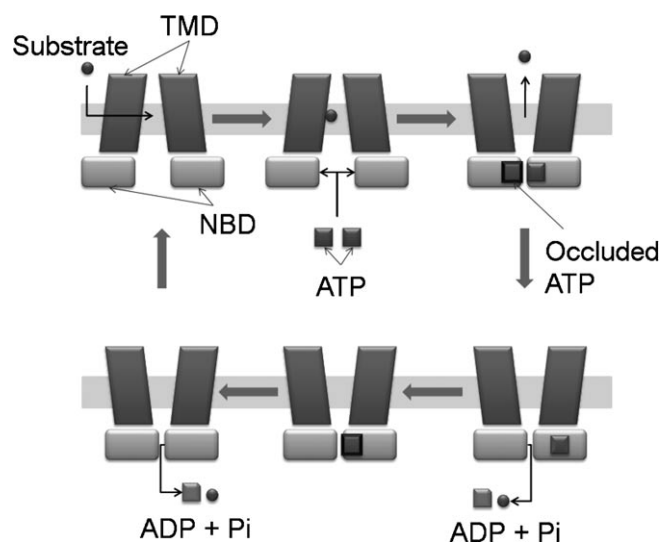
Figure 3. Occlusion-induced switch model.

3.2 Available X-Ray Structures

The availability of high resolution structures of targets is essential for understanding the molecular basis of their function and obviously also for performing structure-based design studies. While the entries in the protein data bank (PDB) are rising exponentially, the structure determination of membrane proteins is still problematic and only relatively few structures have been resolved up to now. Thus, the X-ray structures of *E. coli* MsbA (PDB code: 1JSQ, resolution: 4.5 Å), a lipid A transporter, raised a lot of interest in the ABC-transporter field.[15] However, even higher attention provided the retraction of these structures in 2006, which was due to an error in the data processing.[47]

3.2.1 Sav1866 and Domain Swapping

In 2006 Dawson et al. published the X-ray structure of the multi-drug transporter Sav1866 of *Staphylococcus aureus* in complex with ADP[48] (PDB code: 2HYD, resolution: 3.00 Å).[49] In contrast to P-gp, Sav1866 is a half-transporter. As already mentioned with the ABCG subfamily such transporters consist of two identical monomers which have to homodimerize to yield a functional transporter unit, comprising two TMDs and two NBDs.

This bacterial ABC transporter brought up an interesting arrangement of the transmembrane helices which has not been expected and which finally gave rise to the retraction of the MsbA structures.[47] Instead of having the TMDs on separated halves, building the sides of the large inner cavity (as originally proposed by the retracted MsbA structures), two TM helices (TMH) of one TMD cross the portal thus interacting with the opposite NBD (Figure 2). The parts of the protein that consist of TMH1 and TMH2 from one monomer and TMH3-TMH6 from the other represent a wing like structure (Figure 2a).

The close proximity of the NBDs in the Sav1866 structure was consistent with the electron microscope structure of ABCB1 previously published by Lee et al.,[50] which is also in agreement with the "nucleotide-sandwich dimer" in MJ0796.[51] In addition cross-linking experiments with P-gp showed that TMH 5 and 8 as well as 2 and 11 are located closely together.[52–54] Other cross-linking experiments suggested that the area enclosed by both TMDs of P-glycoprotein is funnel-shaped, with a wide opening at the extracellular side.[55] All this data are consistent with the SAV1866 structure.[56] Half a year later the same group crystallized Sav1866 in complex with the non-hydrolysable ATP-analog AMP.PNP (PDB code: 2ONJ, resolution: 3.40 Å)[57] in a slightly worse resolution. The comparison with the ADP bound state (2HYD) showed no significant differences among these structures.[57] This suggests that 2HYD most likely also resembles the energized, ATP-bound state. However when comparing the nucleotide bound X-ray structures that are open to the extracellular space, with the catalytic cycle depicted in Figure 3, it can be noticed that these states show low affinity for substrates and thus represent a state where the major conformational change already has occurred. The high affinity state therefore seems to be represented by the apo, inverted V-shape like state.[58, 59]

3.2.2 MsbA – the Corrected Structures

Ward et al. fulfilled the need of the nucleotide-free ABC transporter structure and published four different X-ray structures of MsbA in 2007,[60] two nucleotide bound and two in the absence of a nucleotide. One of the two apo structures was captured in a cytoplasmic-facing open state (PDB code: 3B5W, *E. coli*, resolution: 5.30 Å), with the two NBDs located ~50 Å apart from each other. The other apo structure also represented a cytoplasmic-facing, but closed conformation of MsbA (PDB code: 3B5X, *V. cholerae*, resolution: 5.50 Å). As mentioned before, electron microscopy experiments suggested that the NBDs of P-glycoprotein are located close together even in a nucleotide-free state.[50] Moreover, when considering the large hydrophobic pocket of the apo-structure, it might be questioned whether it would be filled with lipid molecules or with water. The closure of this pocket and the associated displacement of these molecules anyway seems rather unlikely.[52] Therefore the MsbA-apo-closed conformation is better qualified to be

taken as a homology modelling template for P-gp than the open conformation. On the other hand, this structure was resolved at a resolution of 5.50 Å only representing the Cα-trace, which renders the modelling process quite difficult.

The nucleotide-bound MsbA structures represented complexes with AMP.PNP (PDB code: 3B5Y, *S. typhimurium*, resolution: 4.50 Å) and ADP·V$_i$ (PDB code: 3B5Z, *S. typhimurium*, resolution: 4.20 Å). At this resolution the complexes are identical, showing a RMSD of <0.65 Å between the Cα positions. Although the resolution of these MsbA X-ray structures are rather low and thus insufficient for a detailed investigation of drug-transporter interactions, they provided substantial insight into the catalytic cycle of ABC transporters and further confirmed the domain swapping topology suggested by the Sav1866 structures.

3.2.3 Mouse P-Glycoprotein

The first mammalian X-ray structure of an ABC transporter was published last year by Aller et al.[16] The publication comprised three nucleotide-free structures of murine P-glycoprotein, without a ligand (PDB code: 3G5U, resolution: 3.80 Å) and in complex with two enantiomeric cyclic peptide P-gp inhibors (CPPI, RRR- and SSS-QZ59; PDB codes: 3G60/3G61, resolution: 4.40 Å/4.35 Å).

Also in this case, the two transmembrane halves, where each of it consists of TMH 1–3 and TMH 6 of one monomer and TMH 4–5 from the other monomer, form an inverted V-shape structure (Figure 2b). For the very first time also co-crystals with inhibitors were available and provided new insights into possible binding areas. The interactions between the protein and the QZ59 isomers have recently been reviewed by Gutmann et al.[61] Even though both enantiomers showed distinct binding regions, the interacting TM helices were almost the same. Considering all residues within 4 Å of the bound CPPIs, both stereoisomers showed contacts with TMH5, TMH6, TMH7 and TMH12. While the RRR-isomer also interacted with TMH11, the SSS-enantiomer showed interactions with TMH 1.

When superposing the Cα positions of the three crystal structures (3G5U: apo, 3G60: co-crystal with QZ59-RRR, 3G61: co-crystal with QZ59-SSS) the RMSD calculated amounts to 0.56 Å, with 0.61 Å difference between the apo and the complexed structure. This information suggests that the binding of these large inhibitors hardly affects the protein structure, with at least almost no backbone movement involved. However, the behaviour of the side chains remains unclear. Therefore, we expanded the superposition by all atoms. The calculated RMSD in this case was raised to a value of 0.76 Å, which suggests that the positions of the protein atoms are highly similar among these structures. This fact is surprising as the polyspecificity of P-gp was always connected with its flexibility, Loo et al. even proposed that the protein binds its ligands via an induced fit mechanism.[62] This assumption should have resulted in a

30

higher RMSD between the crystal structures of the complexes in comparison to the apo structure.

Regarding the amino acid residues that line the drug binding pocket, only one residue (Phe974) showed an average RMSD of >2 Å between the three crystal structures. This Phe974 is part of TMH12 and directly extends into the binding cavity (Figure 4a). As can be seen in Figure 4b the different rotations of this residue are essential for the different binding of the CPPIs. The pose of QZ59-RRR (dark green) would not be possible with the Phe974 rotamers of the protein structures of 3G61 or 3G5U. In case of the isomer QZ59-SSS, two molecules can be seen in the structure.[16] While for the lower positioned molecule the rotation of Phe974 should have no influence, the upper one would be clearly inhibited by the 3G5U and 3G60 rotamers.

The murine P-gp structure shares highest similarity with the closed-apo MsbA structure (PDB code: 3B5X) and therefore is also consistent with electron microscope investigations mentioned before.[50] However, the MsbA structure is more opened on the top of the protein, whereas the NBDs lie closer together than with P-gp.

4 Homology Models for Structure-Based Design

Before the publication of the mouse P-gp structure last year the structural investigations on human P-glycoprotein heavily relied on homology models based on bacterial transporters. As Kerr et al. recently stated in a review,[63] the high resolution structures of ABC transporters and therefore the templates for homology modelling can be divided into two phases, the pre-Sav1866 (2001–2006) and the post-Sav1866 (2006 – present) phase. The same scheme can be applied with homology models, since some of them unfortunately relied on the retracted MsbA structure (Table 1). Nevertheless, since the beginning of the post-Sav1866 phase a considerable number of new homology models has been published.

Table 1 gives an overview on homology models of P-gp and the templates they are based on. So far most homology models rely on the 2HYD structure, since this is the best resolved ABC-transporter structure available. In addition, they fulfill most of the structural restraints obtained by cross-linking studies.[69,70,72] Models on basis of the MsbA structures were mainly used for exploring the conformational changes during the catalytic cycle or for performing docking studies. However, these structures unfortunately possess resolutions far from being suitable for docking experiments, with some templates only showing Cα atoms.

It should be stressed that also the homology models based on the retracted MsbA structures could fulfil a substantial amount of biochemical data.[52,65] Also photoaffinity labelling data obtained with benzophenone-analogous propafenone derivatives could convincingly be mapped onto a homology model of P-gp based on the retracted
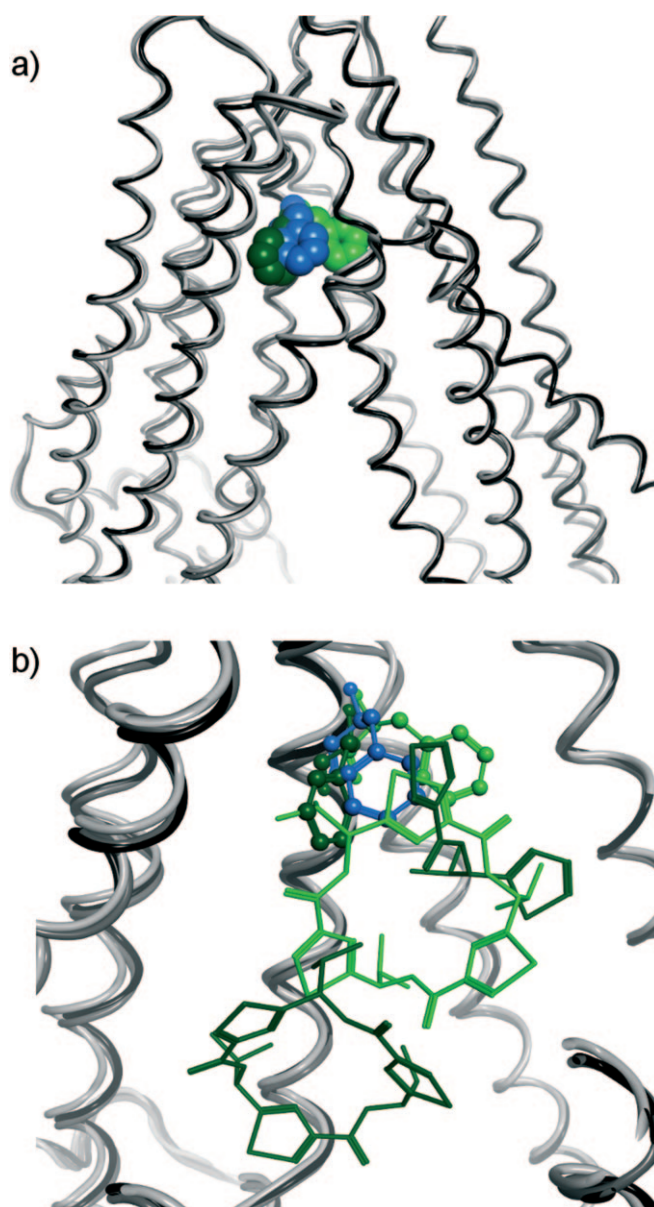
Figure 4. a) Different rotamers of F974 of the different murine X-ray structures. 3G5U (blue), 3G60 (light green), 3G61 (dark green); b) together with co-crystallized QZ59 isomers. QZ59-RRR (light green), QZ50-SSS (dark green).

MsbA structure.[67] This stresses the importance of a careful validation of the X-ray structures and the need for high resolution (<2 Å) structures. Especially in case of these highly flexible, highly promiscuous membrane transporters cysteine cross link studies and ligand photoaffinity labelling could be interpreted in several ways and thus might lead to convincing, sound hypotheses even when based on partially wrong assumptions on the structure of the protein.

With the mouse P-gp structure published last year a high number of homology models for human P-gp are expected to be published this and the following years. The access to

31

Table 1. List of templates that were used for homology models of P-glycoprotein.

Template	Organsim	Sequence Identitiy/ Similarity[a]	Co-crystal[b]	PDB Code	Resolution [Å]	Reference	Homology Models
MsbA	*E. coli*	36%/57%	Apo-open[c]	1JSQ	4.50	[15] *Retracted*	[52,64,65]
MsbA	*V. cholerae*	33%/55%	Apo-closed[d]	1PF4	3.80	[66] *Retracted*	[67]
MsbA	*S. typhimurium*	37%/57%	ADP·V$_i$	1Z2R	4.20	[68] *Retracted*	
Sav1866	*S. aureus*	34%/52%	ADP	2HYD	3.00	[48]	[69–74]
MsbA	*E. coli*	36%/57%	Apo-open	3B5W	5.30	[60]	[75]
MsbA	*V. cholerae*	33%/55%	Apo-closed	3B5X	5.50	[60]	[72]
MsbA	*S. typhimurium*	37%/57%	AMP-PNP	3B5Y	4.50	[60]	
MsbA	*S. typhimurium*	37%/57%	ADP·V$_i$	3B5Z	4.20	[60]	
MsbA	*S. typhimurium*	37%/57%	AMP-PNP	3B60	3.70	[60]	[72]
MalK	*E. coli*	31%/50%	Apo-semi open	1Q1B	2.80	[73]	[73]
MalK	*E. coli*	31%/50%	Apo-open	1Q1E	2.90	[73]	[73]
ABCB1	*M. musculus*	87%/93%	Apo-closed	3G5U	3.80	[16]	[76,77]
ABCB1	*M. musculus*	87%/93%	QZ59-RRR	3G60	4.40	[16]	
ABCB1	*M. musculus*	87%/93%	QZ59-SSS	3G61	4.35	[16]	[78]

[a] Sequence Identity/Similarity with human P-glycoprotein[79]. [b] Co-crystallized molecules (Nucleotides, drugs or apo). [c] Apo-open describes the nucleotide-free protein with the NBDs far apart. [d] Apo-closed describes the nucleotide-free protein with NBDs that lie close together

an X-ray structure of P-glycoprotein, although not in perfect resolution, represents a huge step forward for structure-based studies on this transporter. Although bacterial homologues share sequence identities with ABCB1 of about 35% (Table 1), one has to bear in mind that this incidence relies on the high conservation of the NBDs (>50% sequence identity). The sequence in the TM domains possess only about 20% identity[79] and is therefore in the so called "twilight-zone" concerning homology modelling.[80]

5 Docking Studies

5.1 Binding Sites

Docking is a prevalent tool to identify the binding mode of drugs in the target protein and to use this information for identifying new hits in structure-based virtual screening runs. Concerning ABC transporters in general and P-glycoprotein in particular, we face the problem that hardly any binding sites for known P-gp ligands have been identified unambiguously. Current it seems common sense that there is a large binding cavity in the transmembrane region[81] which comprises distinct active sites. However, due to the polyspecificity of P-gp, there is still only limited knowledge on concrete interaction sites. Furthermore, cysteine-scanning mutagenesis studies showed that the protein is able to bind at least two different molecules simultaneously.[82] A more detailed characterisation of concrete binding sites for distinct substrates utilised techniques like cysteine and arginine scanning, photoaffinity labelling, or hypothesis driven mutagenesis (reviewed in References[24,42,76,83]). This led to the characterization of the interaction regions of Rhodamine 123 and Hoechst 33342, named R- and the H-site,[84,85] together with a regulatory site, which binds prazosin/progesterone.[86] Nevertheless, the release of the P-gp/CPPI-complexes presented another step forward in eluci-

dating drug/P-gp interactions. Since the co-crystallized enantiomers showed distinct binding patterns, this information raised the assumption of stereoselectivity of P-gp in its ligand binding quality.[16] Stereoselectivity has also been shown for flupentixol[87] and propafenone derivatives. However, this has to be taken with a grain of salt and there are also ample reports on equal activity of enantiomes. Thus, as for niguldipine and verapamil both enantiomers showed equivalent activities,[88,89] the distomers with respect to cardiovascular activity were used for clinical studies.

5.2 Docking

As the resolution of the hitherto available templates used for constructing protein homology models is quite low, only very few docking studies have been conducted so far. In a recently published paper, Pajeva et al.[90] docked quinazolinones into a homology model of human P-gp based on 3G61, which is in complex with SSS-QZ59. The binding site they used was defined by the co-crystallized ligands and was extended by 14 Å. The results suggested interaction with TM helices 5, 6 and 11 and were further confirmed by a pharmacophore model.

Becker et al. performed docking studies of the P-glycoprotein modulators colchicine, rhodamine B, verapamil and vinblastine into a homology model based on the closed-apo MsbA structure 3B5X.[91] The binding site was defined as a 30 Å^3 cube which covered the complete central cavity. The resultant poses predicted that all ligands were able to interact with residues that were experimentally identified as important for ligand binding, strongly involving TM helices 5, 6, 7, 11 and 12. However, none of the drugs was able to contact every identified residue, which favours the hypothesis of distinct interactions sites forming one binding cavity.

Based on our extensive data from SAR studies on propa-fenones, we selected a small set of compounds for docking into a homology model based on 3G5U (mouse P-gp without QZ59 isomer).[77] The structure of the apo protein was chosen due to the better resolution compared to those of the complexes. As pointed out already, the structures do not seem to differ to a large extent, except for the amino acid residue Phe974. This residue corresponds to the hP-gp residue Phe978 in the alignment proposed by Aller et al.[16] and does not extend into the binding pocket in the homology model. Assuming a similar binding mode of the propafenone derivatives, poses were prioritized on basis of common scaffold clustering and protein-ligand interaction fingerprints. Interestingly, our results proposed similar interacting regions as have been identified for quinazolinones, involving especially TM helices 5, 6, 7, 8 and 12.[77] This might indicate a common binding region for these two compound classes.

6 Conclusions – a Personal View

Structure-based drug design with low-resolution X-ray structures has to be done very cautiously. Using protein homology models on basis of these structures is even more risky.[92] After almost 30 years of "structural blindness", topped by the retraction of five X-ray structures, the recently published structure of mouse P-gp raised considerable hopes that structure-based drug design to elucidate the molecular basis of transporter/ligand binding becomes possible. However, docking studies performed so far kept the protein rigid. The problem of considering flexible receptors in docking experiments is to find the right balance between the accuracy of the method and the computational cost (reviewed in References[93, 94]). In case of small, defined binding sites with a co-crystallized ligand which is structurally similar to the compounds to be docked, rigid receptor docking would be appropriate. This is definitely not the case for P-gp, so we strongly need to question whether it is adequate to keep the protein rigid, although it is known that P-glycoprotein is highly flexible and, in addition, the resolution of the X-ray structure is as low as 3.80 Å! Furthermore, the X-ray structures of mouse P-gp, although for the first time co-crystallized with inhibitors, represent only singly snapshots of the transporter on its way through the transport cycle. Thus, there are still many drawbacks and pitfalls, which render docking/scoring approaches quite risky. On our personal opinion structure-based design approaches in this area might be useful for creating hypotheses, but definitely not for identifying new hit structures. Intense studies on the structures available, including molecular dynamics simulations of the different templates embedded in phospholipid bilayers, as well as careful validation of the homology models using the plethora of information available from cysteine and alanine scanning as well as cross-linking studies and mutagenesis will be necessary to provide accurate starting structures for docking experiments. The latter, due to the polyspecificity of the transporter, definitely will have to rely on information from ligand-based structure-activity relationship studies, pharmacophore modelling and photoaffinity labeling to be able to provide sound binding hypotheses for selected ligands and to aid in a deeper understanding of the molecular basis of transporter/drug interaction. However, even if the binding mode of several ligands could be resolved, there is still a long way to go for explaining the polyspecificity of the transporter on one side, and the presence of clear structure-activity relationships within structurally related compound series on the other side. As outlined in the introductory section, P-gp inhibitors show clear SAR, but this is observed for numerous scaffolds. Finally, there is experimental evidence that some of the compounds described as inhibitors are stimulating the ATPase activity of P-gp, thus indicating that they are substrates (propafenones,[95, 96] verapamil,[97] cyclosporine[98]). This adds another layer of complexity, especially when considering the fact that within the structurally analogous series of propafenone analogs some compounds inhibited the ATPase activity, some stimulated it in low concentrations and inhibited it in high concentrations and some compounds showed only ATPase stimulation, but no inhibition.[96] To explain these data on a structural basis, intense studies on the dynamics of the transporter combined with detailed investigations of the coupling of ATP-binding to TMD movement need to be performed. Thus, there is definitely still a long way to go for full understanding of the structure and function of this important drug efflux pump and its 47 human homologues.

Acknowledgements

References

[1] R. L. Juliano, V. Ling, *Biochim. Biophys. Acta.* **1976**, *455*, 152–162.

[2] M. M. Gottesman, T. Fojo, S. E. Bates, *Nat. Rev. Cancer.* **2002**, *2*, 48–58.

[3] T. Tsuruo, H. Iida, S. Tsukagoshi, Y. Sakurai, *Cancer Res.* **1981**, *41*, 1967–1972.

[4] M. M. Gottesman, V. Ling, *FEBS Lett.* **2006**, *580*, 998–1009.

[5] G. H. Mickisch, J. Kossig, R. K. Tschada, G. Keilhauer, E. Schlick, P. M. Alken, *Urol. Int.* **1991**, *47*, 118–125.

[6] M. Raderer, W. Scheithauer, *Cancer.* **1993**, *72*, 3553–3563.

[7] G. F. Ecker, P. Chiba, *Transporters as Drug Carriers*, Wiley-VCH, Weinheim, **2009**.

[8] G. Szakacs, A. Varadi, C. Ozvegy-Laczka, B. Sarkadi, *Drug Discov. Today.* **2008**, *13*, 379–393.

[9] G. F. Ecker, *Chem. Today.* **2005**, *23*, 3–6.

[10] M. F. Fromm, R. B. Kim, C. M. Stein, G. R. Wilkinson, D. M. Roden, *Circulation.* **1999**, *99*, 552–557.

33

[11] P. Borst, W. J. van Blitterswijk, J. Borst, A. D. Tepper, A. H. Schinkel, *Drug. Resist. Updat.* **1998**, *1*, 337–339.

[12] H. Kodaira, H. Kusuhara, J. Ushiki, E. Fuse, Y. Sugiyama, *J. Pharmacol. Exp. Ther.* **2010**, published ahead of print.

[13] F. Broccatelli, E. Carosati, G. Cruciani, T. I. Oprea, *Mol. Inf.* **2010**, *29*, 16–26.

[14] K. M. Giacomini, S. M. Huang, D. J. Tweedie, L. Z. Benet, K. L. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K. M. Hillgren, K. A. Hoffmaster, T. Ishikawa, D. Keppler, R. B. Kim, C. A. Lee, M. Niemi, J. W. Polli, Y. Sugiyama, P. W. Swaan, J. A. Ware, S. H. Wright, S. W. Yee, M. J. Zamek-Gliszczynski, L. Zhang, *Nat. Rev. Drug Discov., 9*, 215–236.

[15] G. Chang, C. B. Roth, *Science.* **2001**, *293*, 1793–1800.

[16] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch, G. Chang, *Science.* **2009**, *323*, 1718–1722.

[17] T. J. Raub, *Mol. Pharm.* **2006**, *3*, 3–25.

[18] K. Pleban, G. F. Ecker, *Mini Rev. Med. Chem.* **2005**, *5*, 153–163.

[19] Y. Raviv, H. B. Pollard, E. P. Bruggemann, I. Pastan, M. M. Gottesman, *J. Biol. Chem.* **1990**, *265*, 3975–3980.

[20] I. Pajeva, M. Wiese, *J. Med. Chem.* **1998**, *41*, 1815–1826.

[21] G. Koenig, P. Chiba, G. F. Ecker, *Monatsh. Chem.* **2008**, *139*, 401–405.

[22] G. F. Ecker, P. Chiba, *Recent Res. Devel. Medicinal. Chem.* **2001**, *1*, 121–137.

[23] S. Ekins, in *Transporters as Drug Carriers, Volume 44* (Eds.: G. F. Ecker and P. Chiba), WILEY-VCH, Weinheim, **2009**, pp. 215–221.

[24] T. W. Loo, D. M. Clarke, *Arch. Biochem. Biophys.* **2008**, *476*, 51–64.

[25] C. F. Higgins, K. J. Linton, *Nat. Struct. Mol. Biol.* **2004**, *11*, 918–926.

[26] M. A. Demel, R. Schwaha, O. Kraemer, P. Ettmayer, E. E. J. Haaksma, G. F. Ecker, *Expert Opin. Drug Metab. Toxicol.* **2008**, *4*, 1167–1180.

[27] A. Seelig, *Eur. J. Biochem.* **1998**, *251*, 252–261.

[28] A. Seelig, E. Landwojtowicz, *Eur. J. Pharm. Sci.* **2000**, *12*, 31–40.

[29] R. Didziapetris, P. Japertas, A. Avdeef, A. Petrauskas, *J Drug Target.* **2003**, 391–406.

[30] V. K. Gombar, J. W. Polli, J. E. Humphreys, S. A. Wring, C. S. Serabjit-Singh, *J. Pharm. Sci.* **2004**, *93*, 957–968.

[31] M. P. Gleeson, *J. Med. Chem.* **2008**, *51*, 817–834.

[32] G. F. Ecker, P. Chiba, in *Transporters as Drug Carriers, Volume 44* (Eds.: G. F. Ecker and P. Chiba), WILEY-VCH, Weinheim, **2009**, pp. 349–362.

[33] G. Szakacs, J. P. Annereau, S. Lababidi, U. Shankavaram, A. Arciello, K. J. Bussey, W. Reinhold, Y. Guo, G. D. Kruh, M. Reimers, J. N. Weinstein, M. M. Gottesman, *Cancer Cell.* **2004**, *6*, 129–137.

[34] P. Labute, *J. Mol. Graph. Model.* **2000**, *18*, 464–477.

[35] B. Zdrazil, C. Prakusvudhisarn, G. F. Ecker, **2007**, presented at *ACS 234th National Meeting & Exposition,* Boston.

[36] M. A. Demel, O. Kraemer, P. Ettmayer, E. Haaksma, G. F. Ecker, *Mol. Inf.* **2010**, *29*, 233–242.

[37] E. Bakos, T. Hegedus, Z. Hollo, E. Welker, G. E. Tusnady, G. J. Zaman, M. J. Flens, A. Varadi, B. Sarkadi, *J. Biol. Chem.* **1996**, *271*, 12322–12326.

[38] D. R. Hipfner, K. C. Almquist, E. M. Leslie, J. H. Gerlach, C. E. Grant, R. G. Deeley, S. P. Cole, *J. Biol. Chem.* **1997**, *272*, 23623–23630.

[39] K. Kage, S. Tsukahara, T. Sugiyama, S. Asada, E. Ishikawa, T. Tsuruo, Y. Sugimoto, *Int. J. Cancer.* **2002**, *97*, 626–630.

[40] R. Allikmets, L. M. Schriml, A. Hutchinson, V. Romano-Spica, M. Dean, *Cancer Res.* **1998**, *58*, 5337–5339.

[41] M. F. Rosenberg, G. Velarde, R. C. Ford, C. Martin, G. Berridge, I. D. Kerr, R. Callaghan, A. Schmidlin, C. Wooding, K. J. Linton, C. F. Higgins, *EMBO J.* **2001**, *20*, 5615–5625.

[42] M. A. Seeger, H. W. van Veen, *Biochim. Biophys. Acta.* **2009**, *1794*, 725–737.

[43] Z. E. Sauna, S. V. Ambudkar, *Mol. Cancer. Ther.* **2007**, *6*, 13–23.

[44] A. B. Shapiro, V. Ling, *Eur. J. Biochem.* **1998**, *254*, 189–193.

[45] Z. E. Sauna, S. V. Ambudkar, *Proc. Natl. Acad. Sci. U S A.* **2000**, *97*, 2515–2520.

[46] Z. E. Sauna, K. Nandigama, S. V. Ambudkar, *J. Biol. Chem.* **2006**, *281*, 26501–26511.

[47] G. Chang, C. B. Roth, C. L. Reyes, O. Pornillos, Y. J. Chen, A. P. Chen, *Science.* **2006**, *314*, 1875.

[48] R. J. Dawson, K. P. Locher, *Nature.* **2006**, *443*, 180–185.

[49] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.

[50] J. Y. Lee, I. L. Urbatsch, A. E. Senior, S. Wilkens, *J. Biol. Chem.* **2002**, *277*, 40125–40131.

[51] P. C. Smith, N. Karpowich, L. Millen, J. E. Moody, J. Rosen, P. J. Thomas, J. F. Hunt, *Mol. Cell.* **2002**, *10*, 139–149.

[52] D. R. Stenham, J. D. Campbell, M. S. Sansom, C. F. Higgins, I. D. Kerr, K. J. Linton, *FASEB J.* **2003**, *17*, 2287–2289.

[53] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2004**, *279*, 7692–7697.

[54] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2004**, *279*, 18232–18238.

[55] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2001**, *276*, 36877–36880.

[56] J. K. Zolnerciks, C. Wooding, K. J. Linton, *FASEB J.* **2007**, *21*, 3937–3948.

[57] R. J. Dawson, K. P. Locher, *FEBS Lett.* **2007**, *581*, 935–938.

[58] C. Martin, C. F. Higgins, R. Callaghan, *Biochemistry.* **2001**, *40*, 15733–15742.

[59] P. Aanismaa, E. Gatlik-Landwojtowicz, A. Seelig, *Biochemistry.* **2008**, *47*, 10197–10207.

[60] A. Ward, C. L. Reyes, J. Yu, C. B. Roth, G. Chang, *Proc. Natl. Acad. Sci. U S A.* **2007**, *104*, 19005–19010.

[61] D. A. Gutmann, A. Ward, I. L. Urbatsch, G. Chang, H. W. van Veen, *Trends Biochem. Sci., 35*, 36–42.

[62] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2003**, *278*, 13603–13606.

[63] I. D. Kerr, P. M. Jones, A. M. George, *FEBS J.* **2010**, *277*, 550–563.

[64] M. Seigneuret, A. Garnier-Suillerot, *J. Biol. Chem.* **2003**, *278*, 30115–30124.

[65] S. Vandevuer, F. Van Bambeke, P. M. Tulkens, M. Prevost, *Proteins.* **2006**, *63*, 466–478.

[66] G. Chang, *J. Mol. Biol.* **2003**, *330*, 419–430.

[67] K. Pleban, S. Kopp, E. Csaszar, M. Peer, T. Hrebicek, A. Rizzi, G. F. Ecker, P. Chiba, *Mol. Pharmacol.* **2005**, *67*, 365–374.

[68] C. L. Reyes, G. Chang, *Science.* **2005**, *308*, 1028–1031.

[69] C. Globisch, I. K. Pajeva, M. Wiese, *Chem. Med. Chem.* **2008**, *3*, 280–295.

[70] T. Stockner, S. J. de Vries, A. M. Bonvin, G. F. Ecker, P. Chiba, *FEBS J.* **2009**, *276*, 964–972.

[71] A. Sakurai, Y. Onishi, H. Hirano, M. Seigneuret, K. Obanayama, G. Kim, E. L. Liew, T. Sakaeda, K. Yoshiura, N. Niikawa, M. Sakurai, T. Ishikawa, *Biochemistry.* **2007**, *46*, 7678–7693.

[72] J. P. Becker, G. Depret, F. Van Bambeke, P. M. Tulkens, M. Prevost, *BMC Struct. Biol.* **2009**, *9*, 3.

[73] M. L. O'Mara, D. P. Tieleman, *FEBS Lett.* **2007**, *581*, 4217–4222.

34

F. Klepsch, G. F. Ecker

[74] A. W. Ravna, I. Sylte, G. Sager, *Theor. Biol. Med. Model.* **2007**, *4*, 33.

[75] A. W. Ravna, I. Sylte, G. Sager, *Theor. Biol. Med. Model.* **2009**, *6*, 20.

[76] Z. Parveen, C. Bentele, T. Stockner, S. Pferschy, M. Kraupp, M. Freissmuth, G. F. Ecker, P. Chiba, **2010**, Submitted.

[77] F. Klepsch, P. Chiba, G. F. Ecker, *Drugs of the Future.* **2009**, *34*, 164.

[78] I. K. Pajeva, C. Globisch, M. Wiese, *FEBS J.* **2009**, *276*, 7016–7026.

[79] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

[80] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

[81] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **1999**, *274*, 24759–24765.

[82] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2003**, *278*, 39706–39710.

[83] E. Crowley, R. Callaghan, *FEBS J.* **2010**, *277*, 530–539.

[84] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2002**, *277*, 44332–44338.

[85] Q. Qu, F. J. Sharom, *Biochemistry.* **2002**, *41*, 4744–4752.

[86] A. B. Shapiro, K. Fox, P. Lam, V. Ling, *Eur. J. Biochem.* **1999**, *259*, 841–850.

[87] S. Dey, P. Hafkemeyer, I. Pastan, M. M. Gottesman, *Biochemistry.* **1999**, *38*, 6630–6639.

[88] G. Luurtsema, C. F. Molthoff, A. D. Windhorst, J. W. Smit, H. Keizer, R. Boellaard, A. A. Lammertsma, E. J. Franssen, *Nucl. Med. Biol.* **2003**, *30*, 747–751.

[89] V. Höllt, M. Kouba, M. Dietel, G. Vogt, *Biochem. Pharmacol.* **1992**, *43*, 2601–2608.

[90] I. K. Pajeva, C. Globisch, M. Wiese, *Chem. Med. Chem.* **2009**, *4*, 1883–1896.

[91] J. P. Becker, G. Depret, F. Van Bambeke, P. M. Tulkens, M. Prevost, *BMC Struct Biol.* **2009**, *9*, 3.

[92] C. Oshiro, E. K. Bradley, J. Eksterowicz, E. Evensen, M. L. Lamb, J. K. Lanctot, S. Putta, R. Stanton, P. D. Grootenhuis, *J. Med. Chem.* **2004**, *47*, 764–767.

[93] P. Cozzini, G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, C. A. Sotriffer, *J. Med. Chem.* **2008**, *51*, 6237–6255.

[94] A. M. Henzler, M. Rarey, *Mol. Inf.* **2010**, *29*, 164–173.

[95] I. Bachmakov, S. Rekersbrink, U. Hofmann, M. Eichelbaum, M. F. Fromm, *N-S. Arch. Pharmacol.* **2005**, *371*, 195–201.

[96] D. Schmid, D. L. Staudacher, H. G. Loew, P. G. Spieckermann, G. F. Ecker, S. Kopp, P. Chiba, *J. Pharmacol. Exp. Ther.* **2003**, *307*, 589–596.

[97] T. Litman, T. Zeuthen, T. Skovsgaard, W. D. Stein, *Biochim. Biophys. Acta.* **1997**, *1361*, 169–176.

[98] U. Schramm, G. Fricker, R. Wenger, D. S. Miller, *Am. J. Physiol.* **1995**, *268*, F46–F52.

35

2.2 Structure-based studies on membrane proteins

Contribution of the thesis author

Molecular Modeling and Simulation of Membrane Transport Proteins (ISBN 978-953-51-0513-8) A. Jurik, F. Klepsch and B. Zdrazil
The author gathered literature about the topic of this book chapter and wrote the sections *2.2 Molecular docking* and *3.1 ABC Transporters and multidrug resistance.*

Molecular Modeling and Simulation of Membrane Transport Proteins

Andreas Jurik, Freya Klepsch and Barbara Zdrazil
University of Vienna, Department of Medicinal Chemistry
Pharmacoinformatics Research Group
Austria

1. Introduction

Membranes fulfill the essential need of all living species to separate different compartments. On the other hand, in a cell the homeostatic environment can only be maintained by the cellular membrane acting as a selective 'filter', which allows the cell to continuously communicate with other cells. Mechanisms which facilitate the translocation of materials across the membrane regulate the entrance and disposal of ions, amino acids, nutrients, and signaling molecules.

This selective transport across cellular membranes is carried out by two broad classes of specialized proteins, which are associated with or embedded in those lipid bilayers: channels and transmembrane transporters. They work by different mechanisms: Whereas channels catalyze the passage of ions (or water and gas in the case of the aquaporin channel) (Agre, 2006) across the membrane through a watery pore spanning the membrane-embedded protein, transporters are working via a cycle of conformational changes that expose substrate-binding sites alternately to the two sides of the membrane (Theobald & Miller, 2010).

If we regard the force that drives the transport process there is also a huge difference in the way ion channels and transporters act. Channels assist a downhill movement along a concentration gradient (passive diffusion), whereas in transporters it is usually directed against a concentration gradient of the substrate. Thus, in order to comply with their business, transporters are dependent on another source of the cellular energy. Secondary active transporters rely on ionic gradients. In the case of primary active transporters ATP is the driving force (Wang et al., 2010).

A comprehensive list of all annotated transport proteins is freely available online on the TCDB website (http://www.tcdb.org). This Transporter Classification Database uses an International Union of Biochemistry and Molecular Biology (IUBMB) approved system of nomenclature for transport protein classification. The TC system is analogous to the Enzyme Commission (EC) system for classification of enzymes, except that it incorporates both functional and phylogenetic information (Saier et al., 2006; Saier et al., 2009).

According to the TCDB system Membrane Transport proteins can be classified as follows (List of families and subfamilies of the TC system):

1. Pores and channels
    a. Helical channels
    b. Strand porins
    c. Pore-forming toxins
    d. Non-ribosomally synthesized channels
    e. Holins
2. Electrochemical-potential-driven transporters
    a. Transporters or carriers (uniporters, symporters and antiporters)
    b. Non-ribosomally synthesized transporters
3. Primary active transporters
    a. P-P-bond-hydrolysis-driven transporters
    b. Decarboxylation-driven transporters
    c. Methyl-transfer-driven transporters
    d. Oxidoreduction-driven transporters
    e. Light-driven transporters
4. Group translocators
5. Transmembrane electron carriers
6. Accessory factors involved in transport
7. Incompletely characterized transport systems

Our special interest focuses on transmembrane transport proteins ('transporters'). Excellent manuscripts on membrane channels have been provided by other groups (Gumbart et al., 2005; Schmidt et al., 2006).

To date more than 400 membrane proteins have been annotated in the human genome. Two major superfamilies - which are also intensively investigated in our group - are the **ATP- binding cassette transporters** (ABC, e.g. P-glycoprotein), and the **solute carrier superfamily** (SLC). They will be discussed into more detail in section 3 of this chapter.

1.1 Transporters as pharmacological targets

Transport proteins are playing important roles in the whole drug discovery and development process. They regulate absorption, distribution, and excretion of drugs and therefore influence drug disposition, therapeutic efficacy and adverse drug reactions in the human body. This has to be taken into account in pharmacological studies (Giacomini et al., 2010).

It is estimated that transporters account for about 50% of drug targets. However, their modes of (selective) transport are only poorly understood. This is due to difficulties in membrane protein purification, expression, and crystallization (Caffrey, 2003), which is still in its childhood. As a consequence there exists a striking disproportion between the number of entries of resolved structures of soluble proteins and membrane proteins in the protein databank (PDB, http://www.pdb.org/). To date, only about 2% (1462 by Sept 2011) of the structures are from transmembrane proteins (75594 structures in total). Out of these there are only 302 unique structures (proteins of same type but from different species are included) (Irvine). Moreover, a significant number of the membrane protein structures determined are at relatively low resolutions (Lindahl & Sansom, 2008).

However, there are tremendous efforts as to ameliorate the methods in order to obtain atomic resolution structures of membrane protein molecules (Newby et al., 2009). Thus, for

the past two decades, the number of available structures of membrane proteins has been climbing the exponential foot of the growth curve, since it is doubling every three years (Theobald & Miller, 2010).

For a medicinal chemist, the availability of a growing number of structures, paves the way for further *in silico* studies. Some are very promising in a way that they can be used as templates in order to build up a homology model of the membrane protein of interest. Those models, which may further be applied for Molecular Dynamics (MD) Simulation and Docking Studies, give us the opportunity to gather new insights into the molecular structure and function of the protein under investigation and the behaviour of certain ligands in the binding site.

As translocation always involves a dynamic process, which cannot easily be studied by mere experimental techniques, above all, the application of long-term MD simulations should be implemented into the whole process of drug discovery and development. Due to significant increase in computational power and improvements in parallelization techniques, nowadays simulations of membrane transport proteins may stretch up to microseconds - that is, to physiologically relevant time scales. In this review we are describing the theory and methodology related to computational techniques used in the modeling of transporters and we will outline the recent developments in the field of ABC transporters and neurotransmitter transporters.

2. Methods

2.1 Homology modeling

2.1.1 Basic concepts

Despite the enormous increase of published structural data for proteins, the particular availability for a protein of interest can vary from the sheer presence of the amino acid sequence to a multitude of high-resolution X-ray structures. Fortunately, Mother Nature was not too generous in providing unique structural folds for functionally related proteins, as the structural arrangement within a family of homologous proteins is much higher conserved than the respective amino acid sequences (Lesk & Chothia, 1980). Thus, in many cases the combination of a primary sequence on the one hand and one or more reasonably well-resolved homologue structures on the other hand can result in homology models surprisingly well representing the molecular reality, paving the way to successful comparative modeling studies. The process of predicting the 3D structure of a protein can be achieved by four main steps: fold assignment, target-template sequence alignment, building and evaluation of the models (Cavasotto, 2011).

2.1.2 Assignment of the basic fold and sequence alignment

The first step towards a good model is the identification and careful selection of structurally related template proteins. Although generally a high percentage of global sequence identity is a good indicator for the model quality to be expected, it must be kept in mind that the identity in the area of interest, i.e. the binding site(s), can differ significantly from overall values. One should strive to put the main focus during template selection on facilitating a maximum of accuracy in modeling those vital regions.

Maybe the best source for structural templates is the Protein Data Bank (Berman et al., 2000). As mentioned earlier, it offers the coordinates of structurally resolved proteins, including large amounts of surplus information like primary sequence, experimental settings or co-bound ligands and ions. Search tools like BLAST and FASTA (Altschul et al., 1997; Pearson, 1990) usually do reasonably well in identifying the correct fold of a protein. The second, even trickier step is the subsequent alignment step of target and template sequence, as it is possible with T-Coffee or CLUSTAL W (Notredame et al., 2000; Thompson et al., 1994). Conserved residues and regions of experimentally determined proximity need to be aligned as accurately as possible. Multiple alignments of sequences belonging to the same gene family can significantly enhance the performance of the search for conserved residues or even regions, but thorough literature search and manual adjustment of the alignment are inevitable in order to achieve best results.

A good example for the importance of taking a comprehensive look at the research subject is the meanwhile annual GPCR Dock competition (Kufareva et al., 2011), where prior to the release of newly resolved G-Protein Coupled Receptors (GPCR) 3D structures, modeling groups get the chance to submit their best efforts of predicting the correct receptor and ligand conformation. It turned out, that advanced modeling tools and human intervention contributed about equally to the success of the individual approaches.

2.1.3 Building and refinement of the models

Once it is assumed that the alignment meets all available experimental data, it can be started to calculate coordinates for the target residues. Although automated homology modeling methods exist, the yielded models tend to lack accuracy, especially in cases of low sequence identity (Dalton & Jackson, 2007).

Usually, the crude model is built by aligning the basic backbone framework, the so-called structurally conserved regions (SCRs). Conserved secondary structural elements like α-helices or β-sheets are inherited, being responsible for the general shape of the model. Several homology modeling approaches also try to include information about known ligands into the binding site construction in order to meet its particular geometry (Evers & Klebe, 2004; Sherman et al., 2006).

Subsequently, assignment of the side chain conformations needs to be done according to steric and energetic constraints. Identical residues usually can be considered to be oriented similarly, likewise highly similar amino acids. For non related residues, rotamer libraries can provide initial geometrical guesses (Schrauber et al., 1993), although other effects like packing energies may lead to significant deviations. Up to 30% of side chain conformations in X-ray structures do not correspond to usual rotamers, yet adopting energetically allowed conformations. Naturally, selecting the most probable side chain orientation solely according to statistical criteria is problematic, so methods including structural features of the local environment have been developed (Deane & Blundell, 2001). Still, some cases require manual adjustment, for instance the incorporation of known disulfide bridges, specific internal hydrogen bonds or ion binding pockets.

The major challenge in comparative modeling is the treatment of structurally variable regions (SVRs). Especially flexible loop regions lacking a structural template are difficult to predict, since the calculation time increases nearly exponentially with the degrees of

freedom added by every flexible residue. There are several strategies to meet this issue. Knowledge-based strategies try to find structural guesses by automated database search for related sequence sections in other proteins, possibly not even close to being genetically similar. From a computational point of view, conformational searches by *ab initio* calculation of the desired region are more costly, but recent approaches yielded reasonably good results for a loop length up to 17 residues (Mehler et al., 2002; Zhao et al., 2011). For significantly longer loops, in case that the problematic region is remote from the actual binding site(s) and not considered being directly linked to the binding process, it can be viable to leave it up to the standard modeling software how to build the respective flexible region, and hope for subsequent MD simulations to find a near-to-native conformation (Amaro & Li, 2010).

The model building process bears numerous sources of unfavorable steric strain energies, calling for an appropriate minimization procedure. As one can imagine, this step has to be carefully balanced in order to overcome steric clashes without compromising tediously elaborated side chain orientations or, even worse, entire conserved regions. Instead of global optimization attempts good minimization protocols start with local treatment of clashes with initially fixed backbone atoms. Thus, solvent molecules, ions and hydrogen atoms possibly responsible for large initial forces can adopt energetically more favorable positions. Gradually, initial tethering forces are reduced, avoiding artificial distortions (Höltje et al., 2008). This can be facilitated by molecular mechanics calculations using different force fields like CHARMM (Brooks et al., 1983), OPLS (Jorgensen & Rives, 1988) or AMBER (Weiner et al., 1984). In contrast to force fields for small molecules, they have to handle huge systems, therefore being usually somehow simplified regarding the treatment of long-distance non-bonding interactions or non-polar hydrogen atoms, called united-atom models.

Energy minimized models may be further optimized by molecular dynamics (MD) simulations, as reported in-depth in section 2.3.

2.1.4 Model evaluation

Predictions including as many degrees of freedom as homology models desperately need reliable tools to estimate their quality, as the accuracy of a 3D model is responsible for the amount of information that can be gained by it (Marti-Renom et al., 2000). Several evaluation programs exist; many of them are available online on server-basis. The SWISS-MODEL (Arnold et al., 2006) (http://swissmodel.expasy.org) and the SAVES server (http://nihserver.mbi.ucla.edu/SAVES), for instance, offer a variety of local and global quality estimation tools (Benkert et al., 2011; Hutchinson & Thornton, 1996; Zhou & Zhou, 2002). It is important to look at both, as they are not necessarily mutually related.

A comprehensive stereochemistry check can be carried out using the Procheck suite (Laskowski et al., 1993), searching for geometrically unusual residues in a given protein structure by comparison with stereochemical parameters of high-quality benchmark structures. Likewise, the What_Check module of the WHATIF package and the VADAR web server do a similar job (Vriend, 1990; Willard et al., 2003). The features examined include the planarity of aromatic ring system and peptide bonds, bond lengths and basic checks like C_α-chirality.

Once the yielded quality statistics are acceptable within the limitations of the possible, a model can be considered ready for the further use in, for instance, docking studies.

2.2 Molecular docking

Molecular Docking is a versatile tool in structure based drug design. This technique is able to predict possible orientations of one molecule to another. In this section we will focus on protein-ligand docking, describing the interaction of a small molecule in a binding pocket of the protein of interest.

In principle molecular docking is comprised of three consecutive steps: i) the definition of the binding site, ii) the placement of the ligand inside the defined site, and iii) the ensuing evaluation of this placement, called scoring.

2.2.1 Binding site identification

The right determination of the binding site of the ligand is essential for the subsequent docking process. If the active site is not known there are several algorithms that are able to detect potential binding pockets (extensively reviewed in (Henrich et al., 2010)). These programs scan the protein surface for cavities that fulfill certain geometrical constraints, which mark them as possible ligand binding sites. While the program LIGSITE uses a grid for the surface scan (Hendlich et al., 1997), the PASS algorithm utilizes layers of spheres that should describe buried cavities (Stouten & Brady, 2000).

In order to consider also physico-chemical criteria in binding site detection the surface of the protein can be scanned with fragments of ligands with subsequent calculation of their complementarity. Another approach to get an idea about potential active sites can be achieved by comparing the query protein with homologues, as proteins with related function share similar binding sites. For this purpose the program CAVBASE (Kuhn et al., 2007), which relies on the LIGSITE algorithm, has been developed.

As soon as the binding site is known, it has to be characterized in order to get information about specific binding possibilities through non-covalent interactions. To detect these hot spots atom probes, ligand fragments or whole small molecules are positioned inside the binding pocket. The program GRID is able to detect interactions and solvation effects by calculating the interaction energy between grid points of the binding pocket and certain atom probes (Reynolds et al., 1989). On the other hand, the multiple copy simultaneous search (MCSS) method places thousands of probe copies inside the pocket. After energy minimization those probes cluster at certain local minima defining the hot spots (Caflisch et al., 1993).

2.2.2 Search algorithms

The role of the search algorithm is the correct placement of the ligand in the binding pocket. Ideally it should therefore consider all possible degrees of freedom, which leads to higher accuracy. However, due to limitations regarding computer power, this is penalized in favor of higher speed by reducing the number of the degrees of freedom (Sousa et al., 2006).

Although in protein-protein docking the rigid-body approximation is still applied (Kuntz et al., 1982), in protein-ligand docking the small molecule is treated flexible.

Approaches that try to explore all degrees of freedom of the ligand systematically comprise conformational search methods, fragmentation methods or database methods. By applying

conformational search methods, every rotatable bond of the ligand is rotated in fixed increments. As this can easily lead to a combinatorial explosion, this technique can only be applied for small or rigid ligands. More prevalently used are the so-called fragmentation methods that place fragments of the ligand in the binding pocket, which are subsequently fused. Depending on the fragmentation and placing of the ligand *place and join* and *incremental approaches* can be distinguished. Popular docking programs utilizing this type of search algorithms include LUDI (Bohm, 1992), FlexX (Rarey et al., 1996), DOCK (Ewing et al., 2001), ADAM (Mizutani et al., 1994) or Hammerhead (Welch et al., 1996).

A computationally efficient way to search for possible orientations forms the database method. For this protocol a conformational library of the ligand is prepared which is docked rigidly into the binding site.

Besides systematic search algorithms there are programs that prefer stochastic principles for binding mode prediction. At this the flexibility of the ligand is provided by random conformational changes that are either kept or rejected on basis of a direct evaluation of the conformation. Among others genetic algorithms present a convenient tool for this purpose. With this optimizing procedure a random population of possible ligand poses is generated, where the characteristics (degrees of freedom) of each are stored in its genetic code (chromosome). By applying genetic operations, like cross-over or mutation, new poses are generated and subsequently scored. Depending on this fitness score the pose is either rejected or it replaces the least fit member of the population. This procedure is conducted over thousands of cycles which ends up in highly optimized ligand orientations. This protocol is included in the popular docking programs GOLD (Jones et al., 1995; Verdonk et al., 2003) and AUTODOCK (Goodsell et al., 1996; Goodsell & Olson, 1990; Olson et al., 1998).

Another possibility to consider ligand flexibility is presented by molecular dynamics simulation of the ligand in the binding pocket. However, this is mainly used in combination with other search algorithms (Kitchen et al., 2004).

In the last years not only the flexibility of the ligand but also protein movements due to ligand binding gained more and more importance (B-Rao et al., 2009; Cozzini et al., 2008). Although it is known that some proteins undergo large structural changes, even domain rearrangements, upon ligand binding, by now it is not possible to cover that in reasonable time and effort. However, since docking a ligand into the right conformation of the binding site is extremely important for the quality of the resulting orientations, efficient workarounds have been developed. Soft docking is one possibility to account small movements of the protein side chains during docking (Jiang & Kim, 1991). For this technique soft potentials are applied on certain side chain atoms in the binding pocket, which therefore tolerate overlap with ligand atoms. The merit of this technique is the easy implementation, as only scoring parameters have to be adapted. On the other hand only small changes can be considered and there might be a bias towards the starting structure.

With the help of rotamer libraries, movements of side chains are included in the search algorithm (Leach, 1994). Depending on the size of the library this method calls on moderate computational power and is able to adapt to certain ligand conformations. Nevertheless, as the backbone is kept rigid large structural movements cannot be covered.

Docking into multiple protein structures (MPS) is therefore highly appreciated as they allow flexibility of the protein during the docking process. Different protein conformations (X-ray structures or taken from MD simulations) are selected and multiple docking runs are performed. As this approach is extremely costly, the more efficient method of ensemble docking should be used preferentially. Therefore an average receptor grid is generated and used for docking (Knegtel et al., 1997).

A hybrid technique that is commonly used to encounter protein flexibility is the induced fit docking protocol of the Schrödinger Suite (Sherman et al., 2006). This method turns major attention on the ligand-induced conformational changes of the protein residues surrounding the binding site. Therefore, the ligand is docked into the rigid binding pocket, amino acid residues that are within a certain radius of the resulting poses are removed and rebuilt using the Schrödinger homology modeling program Prime. After energy minimization of the complex the ligand is redocked into the modified binding pocket.

2.2.3 Scoring functions

The application of a scoring function is important to assess the quality of ligand orientations in the binding pocket that resulted from docking experiments. Basically there are three areas of use for scoring functions. In order to understand the interaction between a defined molecule and the target protein the scoring function needs to be able to identify the true pose among the plethora of orientations, generated by the search algorithm. For lead optimization in particular a scoring function should correctly determine the affinity between the ligand and the protein. However, for virtual screening of large compound databases scoring should provide correct ranking. As there are still limitations regarding computer power, the right balance between accuracy and speed has to be chosen, which is strongly dependent on the field of application (reviewed in (Huang et al., 2010)).

Force field based scoring functions use terms that describe the free energy of binding for evaluating binding poses. In that regard bond stretching, angle bending and dihedral angle forces for the ligand, but also non-bonded VDW and electrostatic interactions with the protein are calculated (Huang et al., 2006). Furthermore the accuracy of these methods depends on their treatment of the solvent. More accurate techniques, like thermodynamic integration or free energy pertubation, treat water molecules explicitly. As these are the most expensive affinity prediction methods, more simplified and computationally less expensive versions are linear interaction energy (LIE) models, where two additional empirical parameters can be used to reduce the number of simulations needed. On the other hand, MM/PBSA and MM/GBSA methods gain speed by using implicit solvent models.

However all of these methods are still not applicable for virtual screening as they are computationally too expensive.

Empirical scoring functions are therefore a fast alternative. They assess the quality of binding by a number of weighted terms that are derived by fitting data of complexes to known affinities (Bohm, 1994; Bohm, 1998). Numerous commonly used scoring functions belong to this group, including ChemScore (Eldridge et al., 1997) and X-Score (Wang et al., 2002). Nevertheless, a disadvantage of this method would be the dependence on the training set, as complexes with binding affinity are essential.

Thus, knowledge-based scoring functions may be preferred in this regard. These scoring functions make use of the statistical occurrence of protein-ligand interactions of complex databases. In contrast to empirical functions they do not aim at reproducing binding-affinities, but experimentally determined structures, wherefore a much larger training set can be used (Tanaka & Scheraga, 1976). Representatives of this group of scoring functions are among others ITScore (Huang & Zou, 2006; Huang & Zou, 2006) and DrugScore (Gohlke et al., 2000). A further development of the ITScore by Zou et al. ITScore/SE managed to include solvation and entropic effects into the scoring function (Huang & Zou, 2008), which lead to a strong increase in scoring accuracy.

As the choice of the scoring function strongly depends on the research query, the combination of several functions, so-called consensus scoring, has been suggested (Charifson et al., 1999).

2.3 Molecular dynamics

It is obvious that the mechanism of action by which certain nutrients or drugs are translocated by a transporter implicates the protein to be flexible. In order to be able to allow for a sufficient comprehension of the dynamics of the transport protein, we can not only rely on experimental techniques. In addition, biomolecular simulations can provide a detailed description of particles in motion as a function of time. Thus, they are an important tool for understanding the physical basis of the structure and function of proteins, and biological macromolecules in general. However, experimental validation should always serve to test the accuracy of the calculated results and also to provide a basis for improving the methodology (Karplus & McCammon, 2002).

It is almost 35 years ago, since for the first time McCommon, Gelin and Karplus have studied the dynamics of the pancreatic trypsin inhibitor by solving the equations of motion for the atoms with an empirical potential energy function (McCammon et al., 1977). In this very beginning of Molecular Dynamics simulations, the calculations were still restricted to the picosecond timescale. However, according to Moore's Law computer power is doubling approximately every two years (Moore, 1965). Thus, MD simulations of biomolecules now are able to stretch up to microseconds. For the study of biological relevant phenomena like enzyme catalysis or even protein folding, MD has become a standard tool – always complementary to experimental techniques.

2.3.1 Theory, fields of application, strengths and limitations of MD simulations

By integrating the Newtonian Equations of Motion, Molecular Dynamics simulations are able to describe the behavior of particles in a certain system within the observed period of time. The interaction of the atoms is described by the potential energy function of the given force field [e.g. Amber (Cornell et al., 1995), CHARMM (Brooks et al., 1983), GROMOS (Scott et al., 1999), OPLS (Jorgensen & Rives, 1988)]. Nowadays, there is an ongoing effort to ameliorate these parameters in a need for models being as less artificial as possible.

The field of application of biomolecular simulations is manifold. It reaches from validation and optimization of previously built homology models, refinement of crystal structures, to the prediction of protein-ligand, and protein-protein interactions, to the study of

functional properties of biological systems at the atomic level (e.g. protein-folding, destabilization or structural change of a protein upon mutation), to even *de novo* design of proteins (Park et al., 2005).

Despite obvious drawbacks of classical MD simulations, which include limitations in time scales that can be studied but also certain inaccuracies of the force field, for instance with respect to polarization effects, the ability of bringing molecular structures alive also allows the researcher to sample the conformational space. This is especially interesting in ligand-docking applications. On one hand, various extracted snapshots from a previously MD-equilibrated protein-ligand complex may serve in order to perform an ensemble docking which is said to outperform docking into only one sample structure (Knegtel et al., 1997). Secondly, MD may also be used in order to refine certain poses and study the ligand-protein interactions on a molecular basis as a function of time.

2.3.2 Simulations in a membrane

The setup of a simulation system, which includes a protein embedded into a lipid bilayer requires additional efforts in comparison to a system with a soluble protein. There are different choices the researcher has to make regarding to the nature of the phospholipid bilayer used, the temperature at which the simulations should be performed (this also depends on the nature of the bilayer), the force field, the water model (e.g. SPC, SPC/E, TIP3P, TIP4P, TIP5P; this also depends on the choice of the force field), and many more.

One of the most challenging parts is the correct parameterization of the ligands. According to the force field that has been selected there are diverging approaches, ranging from a pure manual assignment of partial charges and force constants, to the use of scientific software like Gaussian (www.gaussian.com), to an automated procedure by taking advantage of platforms such as the Automated Topology Builder and Repository (Malde et al., 2011). However, it has to be stated clearly that a manual inspection and refinement of suchlike obtained topologies will always be needed.

Membrane proteins should be placed in a bilayer which is as similar as possible to its native environment. There is a diverse spectrum of phosphlipid bilayers available – differing mainly in the charges of their polar head groups, lengths and saturation of their acyl chains. If lipids play key roles in the proteins function, different combinations of lipids will probably better represent the *in vivo* conditions. It should always be kept in mind that in order to simulate the membrane in a liquid-crystalline state the temperature of the simulation needs to be above the melting temperature of the chosen lipids (phase-transition temperature).

The protocols for setting up MD simulations of membrane proteins are manifold. In any case, however, one needs a pre-equilibrated bilayer, which can be retrieved from different groups around the world (e.g. Peter Tieleman, Scott Feller, Helmut Heller, Mikko Karttunen) or an individual bilayer may be generated and equilibrated with regard to the respective size and nature of the protein to be studied.

When it comes to the insertion of the respective protein into the pre-equilibrated bilayer, again no standard protocol has been established up to now. In any case, it is of utmost importance to obtain a system with a tightly packed bilayer around the protein, so that the

consecutive equilibration time for the membrane can be kept quite short. Protocols like inflategro (Kandt et al., July 2009; Kandt et al., 2007) or g_membed (Wolf et al., 2010) seem most suitable. Whereas, inflategro inflates the lipid bilayer, insert the protein and then deflate the lipid bilayer again, g_membed does it the other way around. It grows a protein into an already hydrated and equilibrated lipid bilayer during a short MD simulation. A special case of insertion procedure certainly is the use of coarse-grained simulations. Here the lipids are able to self-assemble around the protein. However, as this type of simulations use a very simplified description of interactions, for a lot of investigations the relevant information might not be captured.

An idea of a general protocol for the set up of a MD simulation can be found here:

1. Choose a force field for which you have parameters for the protein and lipids.
2. Insert the protein into the membrane.
3. Solvate the system and add ions to neutralize excess charges and adjust the final ion concentration
4. Energy minimize.
5. Let the membrane adjust to the protein. Typically run MD for ~5-10ns with restraints on all protein heavy atoms.
6. Equilibrate without restraints (gradually release the protein).
7. Run production MD.
8. Analysis.

As seen from this overview, after the insertion of the membrane protein it is inevitable to properly equilibrate the lipid bilayer again. This is done by restraining the protein (plus eventually existing ligands conserved water molecules, ions, and cofactors) during a MD run where the membrane is able to adjust to the protein. Subsequently, the whole system has to undergo an extensive equilibration procedure. The end point of the equilibration phase and simultaneous starting point for the MD production run can be determined mainly by evaluation of system parameters (e.g. total energy, temperature) and parameters concerning the protein (e.g. backbone root mean square deviation). A production run for membrane proteins typically resides somewhere in between 50 ns and hundreds of nanoseconds.

2.3.3 Enhanced sampling techniques

As already mentioned in chapter 2.3.1, classical MD simulations are confronted with their limitations in time scales. The limiting factor is the maximum timestep that can be used for the integration, determined by the fastest motion in the system (e.g. bond vibrations).

Thus, it is not able to study 'slow' biological processes without taking advantage of enhanced sampling techniques. This includes of course always a method, which works at the expense of fidelity.

As outlined in an excellent review of Christen and van Gunsteren (Christen & van Gunsteren, 2008) we have to distinguish three different types of search and sampling enhancement techniques: deformation or smoothening of the potential energy surface (e.g. Coarse-graining the model by reduction of the number of interaction sites), scaling of system parameters (e.g. simulated temperature annealing), and multi-copy searching and sampling (e.g. replica-exchange algorithm).

If we want to study membrane proteins and especially their interactions with ligands sampling along transition pathways will be needed. Such pathways are often characterized by high-energy barriers separating meta-stable states along the ligand/substrate transition. Here, methods like pulling or steered MD (SMD) and targeted MD (TMD) may be used in order to drive the sampling to a specific direction. In the SMD approach external forces are applied on certain atoms in order to accelerate processes that are otherwise too slow to model (Isralewitz et al., 2001). Subsequently, the potential of mean force required to induce the transition can be used to estimate free energy barriers. This method is well established and has been used in many applications (Isralewitz et al., 2001; Lu & Schulten, 1999).

In the TMD method, the reaction coordinate is defined by a single mass-weighted root mean-square 'target distance' between a known initial structure and a fixed final (target) structure. By gradually reducing the constrained target distance to zero, the system is driven from the reactant to product state without explicitly defining the reaction pathway (Schlitter et al., 1994).

3. Recent developments in transporter research – Examples

3.1 ABC Transporters and multidrug resistance

ABC (ATP binding cassette) transporters are ubiquitous proteins that are expressed by prokaryotic and eukaryotic organisms. About 50 human ABC transporters are known, which are divided into seven different subfamilies, designated A-G.

Depending on ABC subfamily substrates include among others drugs, lipids, bile salts, peptides, ions and amino acids. Additionally some ABC proteins are known for transporting a broad variety of chemically diverse molecules, which are therefore referred as multidrug transporters. Besides their physiologically important protecting function of exporting xenotoxins, these efflux pumps affect pharmacokinetic profiles of many drugs. Furthermore the acquisition of multidrug resistance (MDR) can often be traced back to elevated expression of multidrug transporters in the affected cells.

The three ABC transporters mostly associated with MDR are P-glycoprotein (P-gp, ABCB1), multidrug resistance protein 1 (MRP1, ABCC1) and breast cancer resistance protein (BCRP, ABCG2).

P-gp is encoded by the *mdr1* gene and is expressed in epithelial cells of the blood brain barrier, liver, kidney and intestine, where it is located at the apical side of the membrane (Szakacs et al., 2008) (Fig. 1).

The cells of the blood brain barrier (BBB) are closely linked by tight junctions, which practically prevent hydrophilic molecules to diffuse between the cells into the central nervous system (CNS). However, as hydrophobic substances might diffuse through the membrane, it is the role of P-gp to keep those out as well (Neuhaus & Noe, 2009). The protecting function of P-gp at the BBB has been observed with *mdr1* knock-out mice and the dog breed collie, which naturally lacks functional P-gp because of a mutated *mdr1* gene. Collies are extremely susceptible to neurotoxic drugs and thus show dramatic adverse reactions after treatment with the antiparasitic drug ivermectin (Mealey et al., 2001).

This detoxifying role of P-gp can be observed at other barriers as well (e.g. the fetal-maternal barrier), but also in faster clearance of administered drugs as it exports substrates form the hepatocytes into the bile, and from the intestinal epithelium into the intestinal lumen (Schinkel & Jonker, 2003).
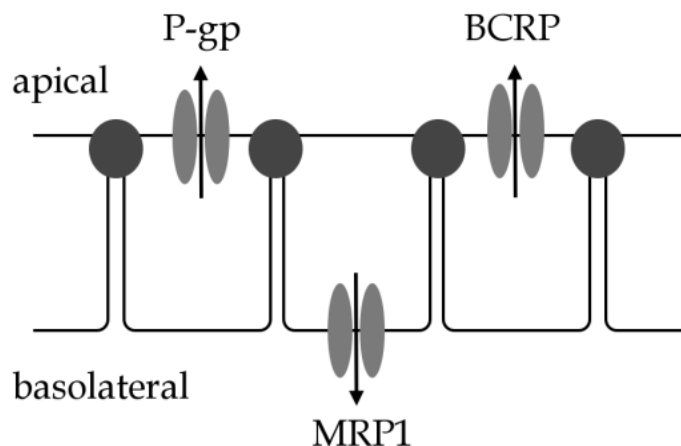
Fig. 1. Localization of the three most important multidrug-transporters.

However, in drug research P-gp poses a large problem, since it highly influences pharmacokinetic properties of drugs. Because of the efflux behavior in the intestinal epithelium the oral bioavailability of drugs is hindered. There are a number of compounds that are able to modulate P-gp activity, which results in modified P-gp concentrations in the target tissue. As a consequence this can lead to adverse drug-drug interactions, when therapeutics are administered at the same time. Furthermore, elevated expression of P-gp, (as it is the case in cancer cells), is one major reason for the acquisition of MDR. One way to overcome these negative effects associated with P-gp activity would be the development of P-gp inhibitors that should restore sensitivity to therapeutics. Already 30 years ago, the reversal of resistance against the vinca alkaloids vincristine and vinblastine by the calcium-channel blocker verapamil was identified (Tsuruo et al., 1981). However, since then no inhibitor reached the market so far. This can be explained by its important physiological functions, rendering them rather antitargets than targets (Ecker & Chiba, 2009).

Another ABC transporter that is highly associated with MDR belongs to the ABCC subfamily. MRP1/ABCC1 is located at the basolateral membrane of epithelial cells of the lung, kidney and the intestine (Fig. 1).

Although the substrate specificity of MRP1 shows some overlap with P-gp especially in terms of hydrophobic substances, MRP1 preferably binds to anionic substances in contrast to the positive charged substrates of P-gp (Borst & Elferink, 2002). Furthermore MRP1 is known for the export of hydrophilic substances as glutathione (GSH) conjugates. Therefore it is not only responsible for preventing xenotoxins entering the cell, but MRP1 also effluxes toxic metabolic compounds, which is highly important for faster clearance.

The role of MRP1 in the acquisition of MDR has particular impact on non-small cell lung carcinoma, a very aggressive cancer type, where high concentrations of MRP1 could be detected in the cancer cells.

The development of MRP1 specific inhibitors faces immense problems as MRP1 substrates and inhibitors show anionic properties, which lack good cell penetration properties (Schinkel & Jonker, 2003).

In 1998 Doyle et al. identified another ABC transporter that conferred resistance to the anthracenedione mitoxantrone, which is a poor substrate for P-gp and MRP1 (Doyle et al., 1998). BCRP belongs to the G or white subfamily of ABC transporters and received the name BCRP because of its isolation from a breast cancer cell line.

As P-gp, BCRP is located at the apical membrane of epithelial cells in the intestine, kidney and placenta (Schinkel & Jonker, 2003) (Fig. 1). Regarding substrate profiles, BCRP shows some overlap with P-gp and MRP1 but does not confer resistance to taxols, cis-platin and verapamil, or vinca alkaloids and anthracyclines. On the other hand, BCRP is known for transporting positively and negatively charged drugs (Sharom, 2008).

A specific inhibitor of BCRP is the tremorgenic mycotoxin fumitremorgin C (FTC). FTC blocked mitoxantrone transport by BCRP without affecting P-gp or MRP1-mediated drug resistance (Rabindran et al., 2000). However, due to neurotoxic effects *in vivo* application is still not possible.

3.1.1 Structure of ABC transporters

The minimal functional unit of an ABC transporter consists of two (pseudo)-symmetric halves that comprise a transmembrane (TM) and a nucleotide binding domain (NBD) (Fig. 2). In the case of P-gp and most other eukaryotic ABC transporters, these subdomains are fused to one polypeptide chain. On the contrary BCRP is a so-called half-transporter (Fig. 2). Half-transporters express each protein half separately and thus need to homo-dimerize to yield functional full transporters.
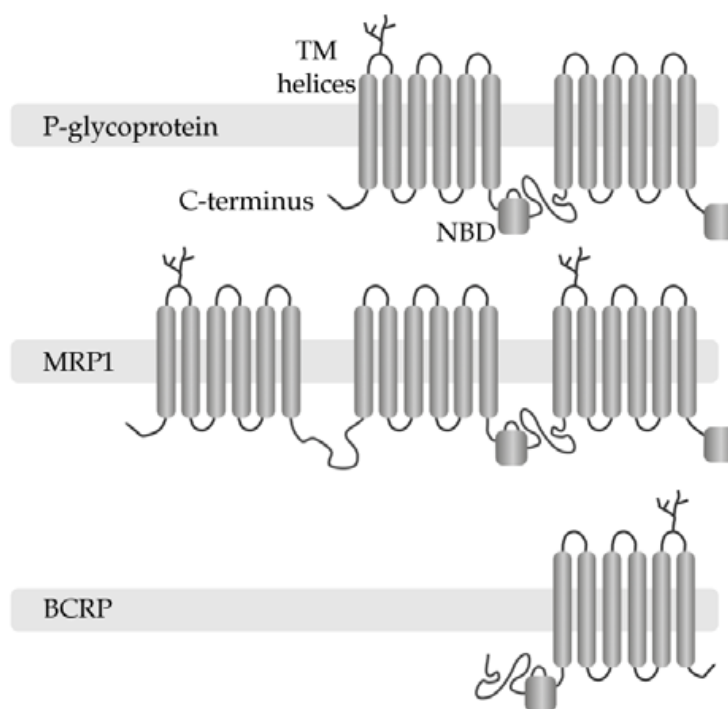
Fig. 2. Topology of the three most relevant multidrug transporters.

The NBDs are responsible for the binding and hydrolysis of ATP, which is needed for the active transport of substrates. On each NBD sequence the characteristic ABC domain consisting of the Walker A and B region, as well as the "signature" or C motif, can be found (Fig. 3). One ATP molecule is supposed to be sandwiched between the Walker A and B of one NBD and the C motif of the other NBD (Fig. 3). As they are highly conserved, the NBDs show large sequence identity among ABC transporters.
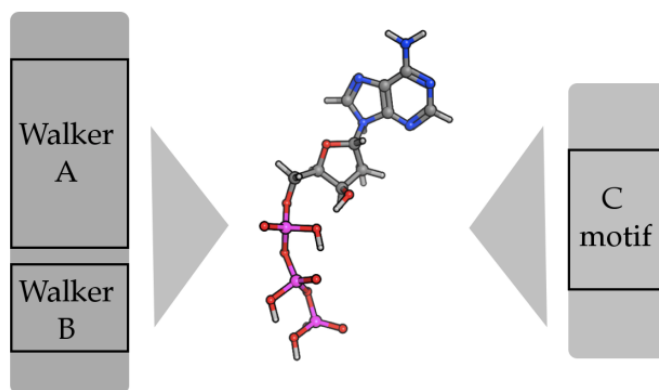
Fig. 3. Position of one ATP molecule in one nucleotide binding domain.

However, substrate binding and transport occurs at the TMDs. Each TMD consists of six TM helices, although this number varies between ABC transporters. The TMDs are much less conserved which leads to a large diversity in substrate profiles among ABC transporters.

During drug transport P-gp and its homologues undergo large conformational changes, converting an open-inward drug-binding state into an open-outward drug-releasing state (Rosenberg et al., 2001). This assumption was confirmed by cryo-electron microscopy and biochemical experiments, where P-gp was trapped in different states of the catalytic cycle (using the non-hydrolysable ATP analog AMP-PNP and ADP-Vi). The detailed mechanism of the energy driven drug transport, rendering the high-affinity into a low-affinity binding site, is currently hypothesized in two different ways and has been extensively reviewed in (Seeger & van Veen, 2009).

3.1.2 Homology modeling of P-gp

As already described in the introduction of this chapter, entries in the protein data bank (PDB) raise exponentially, but the structure determination of membrane proteins is still problematic and only relatively few structures have been resolved up to now. Thus, homology modeling is essential for performing docking or MD studies on most of the ABC transporters.

In 2001 the publication of X-ray structures of *E. coli* MsbA (PDB code: 1JSQ, resolution: 4.5Å), a lipid A transporter, raised a lot of interest in the ABC-transporter community (Chang & Roth, 2001).

The lipid flippase MsbA is an ABC protein that is responsible for the transport for lipid A and lipopolysaccharide (LPS). A non-functional MsbA leads to accumulation of lipopolysaccharide and phospholipids in the inner membrane of gram-negative bacteria.

According to the X-ray structure published in 2001 the association of the two TMDs was interpreted as a chamber that provides alternating access for potential ligands during the catalytic cycle. The theory of MsbA switching between different conformations was confirmed by the subsequent publications of the X-ray structures of Vibrio cholerae MsbA in 2003 (PDB code: 1PF4, resolution: 3.80Å) (Chang, 2003) and of Salmonella typhimurium MsbA in complex with ADP·Vi (PDB code: 1Z2R, resolution: 4.20Å) (Reyes & Chang, 2005). The former presents the apo protein in a closed state, while the latter captures the Protein/ADP·Vi complex in the posthydrolytic state.

At this time these structures were the only source for structure-based design on MsbA and its homologues. Numerous ABCB1 homology models were generated relying on these MsbA templates (Pleban et al., 2005; Seigneuret & Garnier-Suillerot, 2003; Shilling et al., 2003; Stenham et al., 2003; Vandevuer et al., 2006).

However, with the publication of the X-ray structure of the *Staphylococcus aureus* transporter Sav1866, an MsbA homologue (Dawson & Locher, 2006), the previous MsbA and two additional EmrE structures had to be retracted (Chang et al., 2006). According to Chang, an error in the *in-house* software that should process the crystallographic data resulted in a sign change and therefore to a momentous misinterpretation of the data (Matthews, 2007). This incident became the center of numerous discussions, often referred to as the "pentaretraction" (Davis et al., 2008; Penders et al., 2007). In contrast to the retracted MsbA models, the architecture of Sav1866 shows a helix arrangement that is analogous to domain swapping in other enzymes. Thus, TM helices of one TMD are in close contact with the opposite NBD via so-called coupling helices.

One year later, in 2007, Ward et al. published the corrected MsbA structures (Ward et al., 2007), which are in agreement with the SAV1866 architecture.

As SAV1866 (PDB code: 2HYD, resolution: 3.00Å) is one of the best resolved ABC exporters it has been often used as template for further modeling studies. In addition, this structure also fulfills most of the structural restraints that were obtained by cross-linking studies. However SAV1866 was crystallized in the nucleotide-bound conformation, which represents the ligand-releasing state of the protein. Thus the suitability of this template for docking studies can be questioned. In this respect Stockner et al. generated a data-driven homology model on the basis of SAV1866 that should represent the ligand-binding state of the protein by applying structural restraints in TM helices 6 and 12 obtained by cross-linking data on the model (Stockner et al., 2009).

The corrected MsbA coordinates cover different catalytic states, including a nucleotide-free ligand-binding conformation. Unfortunately these structures are resolved at resolutions far from being suitable for docking experiments, with some templates only represented by Cα atoms. Models on basis of the MsbA structures were therefore mainly used for exploring the conformational changes during the catalytic cycle.

With the publication of murine P-gp in March 2009 the first X-ray structure of a eukaryotic ABC exporter was available (Aller et al., 2009) (PDB code: 3G5U, resolution: 3.8Å). Two additionally published structures that include co-crystallized enantiomeric cyclic peptide inhibitors (CPPIs; QZ59-RRR and QZ59-SSS) highlight the binding-competence of these conformations and thus their great value for further docking studies. Furthermore the high sequence identity of 87% with human P-gp highly facilitates the modeling process.

3.1.3 Docking and MD studies

The definition of a binding site is an essential preparation step for docking studies. Regarding P-gp and other ABC transporters, we face the problem that hardly any binding sites for known P-gp ligands have been identified unambiguously. So far, it has been assumed that there is a large binding cavity in the transmembrane region (Loo & Clarke, 1999), which comprises distinct active sites. Furthermore, cysteine-scanning mutagenesis studies showed that the protein is able to bind at least two different molecules simultaneously (Loo et al., 2003). By using biochemical techniques a more detailed characterization of concrete binding sites for distinct substrates was possible (extensively reviewed in (Crowley et al., 2010; Loo & Clarke, 2008; Seeger & van Veen, 2009)). This led to the characterization of the interaction regions of Rhodamine 123 and Hoechst 33342, named R- and the H- site (Loo & Clarke, 2002; Qu & Sharom, 2002), together with a regulatory site, which binds prazosin/progesterone (Shapiro et al., 1999). Furthermore, the release of the P-gp/CPPI-complexes presented another step forward in elucidating drug/P-gp interactions.

Since the co-crystallized enantiomers showed distinct binding patterns, this information raised the assumption of stereoselectivity of P-gp in its ligand binding quality (Aller et al., 2009). Stereoselectivity has also been shown for flupentixol (Dey et al., 1999) and propafenone derivatives (Jabeen et al., 2011). On the other hand there are also ample reports on equal activity of enantiomeres. Thus, as for niguldipine and verapamil both enantiomers showed equivalent activities (Hollt et al., 1992; Luurtsema et al., 2003), the distomers with respect to cardiovascular activity were used for clinical studies.

As the resolution of the hitherto available templates used for constructing protein homology models is quite low, only very few docking studies have been conducted so far. Shortly after the publication of mouse P-gp, Pajeva et al. docked quinazolinones into a homology model of human P-gp based on the murine homologue, which is in complex with the cyclopeptide SSS-QZ59 (Pajeva et al., 2009). The binding site they used was defined by the co-crystallized ligands and was extended by 14Å. The results suggested interaction with TM helices 5, 6 and 11 and were further confirmed by a pharmacophore model.

Becker et al. performed docking studies of the P-glycoprotein modulators colchicine, rhodamine B, verapamil and vinblastine into a homology model based on the nucleotide-free corrected MsbA structure (Becker et al., 2009). The resultant poses predicted that all ligands were able to interact with residues that were experimentally identified as important for ligand binding, strongly involving TM helices 5, 6, 7, 11 and 12. However, none of the drugs was able to contact every identified residue, which favors the hypothesis of distinct interactions sites forming one binding cavity.

Recently Dolghih et al. published a docking approach that was able to discriminate between known P-gp binders and non-binding metabolites (Dolghih et al., 2011). In this study there was a major interest in considering the high flexibility of P-gp. Therefore the induced fit protocol of the Schrödinger Suite was applied (Sherman et al., 2006). However, the discrimination between binders and non-binders can be more efficiently performed on basis of physicochemical properties than different binding mechanisms.

In our group, docking into a homology model based on mouse P-gp was used for explaining the stereoselective P-gp modulating activity of tricyclic benzopyranooxazines (Jabeen et al.,

2011). Besides from activity differences, compounds with 4aS,10bR configuration showed a clear logP-activity correlation (r^2=0.96), which was not the case for the 4aR,10bS series. This characteristic could be partly explained by the received binding hypotheses. The analysis of the docking poses by agglomerative hierarchical clustering resulted in distinct clusters for the different diastereomers. Therefore it has been hypothesized, that activity differences of the diastereomers is due to their different binding modes in the P-gp binding cavity. In addition, molecules with 4aR,10bS chirality were found close to the entry path of the protein, wherefore activity is primarily affected by the molecules' partition coefficient. On the other hand compounds of the 4aS,10bR series also showed docking poses at an active site in the binding pocket of P-gp, thus suggesting that the activity is dependent on multiple factors.

Furthermore, we were able to propose reliable binding hypotheses of propafenone analogs in P-gp by applying a knowledge driven docking protocol (Klepsch et al., 2011). Based on our extensive data from SAR studies on propafenones (Ecker et al., 2008; Pleban et al., 2005), we selected a small set of compounds for docking into a homology model based on mouse P-gp. As propafenone analogs show a clear SAR we assumed a similar binding mode of the docked propafenone derivatives. In that sense the resultant docking poses were clustered according the RMSD of their common scaffold. The clusters were prioritized according a combination of SAR data and protein-ligand interaction fingerprint information. With this protocol a high number of docking poses could be reduced to two reliable binding modes. Key interactions formed by these two clusters were formed with amino acids of TM helices 5, 6, 7 and 12 which were shown previously to be involved in ligand binding (Loo & Clarke, 2008; Seeger & van Veen, 2009).

In contrast to the compounds investigated above steroidal compounds are assumed to bind to the NBD rather than the TMDs. Several docking studies could show ATP-like binding of flavonoids, flavones and chalcones at the ATP-binding site, which is extensively reviewed in (Klepsch et al., 2010).

Regarding P-gp's high flexibility MD simulation represents a convenient technique to consider structural changes of the protein. Unfortunately, a number of MD studies were conducted relying on homology models based on the retracted MsbA X-ray structures (Campbell et al., 2003; Omote & Al-Shawi, 2006; Vandevuer et al., 2006) and are therefore partly no longer valid.

By now MD methods were mainly used for functional investigations of the protein. In order to determine the mechanisms of ATP hydrolysis numerous studies were conducted on isolated NBDs (Campbell & Sansom, 2005; Jones & George, 2007; Jones & George, 2009; Newstead et al., 2009; Wen & Tajkhorshid, 2008), as this comprises the sequence motives essential for ATP-binding. However, recently also studies considering the behaviour of the whole protein upon ATP hydrolysis were published (Becker et al., 2010; Gyimesi et al., 2011; Oliveira et al., 2011). All of those studies relied on the SAV1866 crystal structure, which represents the ligand-releasing and therefore open-outward state of the protein.

While Oliveira et al. (Oliveira et al., 2011) were able to show that replacing both ATP molecules in the NBDs by ADP structural changes in the protein occurred, Gyimesi et al. (Gyimesi et al., 2011) observed structural rearrangements already by exchanging one ATP molecule. This could be of great relevance for heterodimeric ABC proteins like P-gp, where

an asymmetric ATP hydrolysis might be possible. In addition movements in TM helices 3 and 6 could be identified, which is in agreement with MD studies conducted by Becker et al. (Becker et al., 2010). Both groups observed closure of the TM domains after ATP hydrolysis.

The investigation of the drug-binding open-inside conformation of P-gp by MD simulation still faces numerous problems, due to instability of the mouse P-gp structure. In that sense the validity of this model is somewhat doubted (Gyimesi et al., 2011; Loo et al., 2010).

3.2 Neurotransmitter transporters

3.2.1 Biological background of the SLC-6 family

The concerted release and reuptake of transmitter substances is a basic principle of proper signal transduction in the nerve cells. In order to terminate a synaptic signal after neural firing, transporter proteins have to remove about 10^5-fold of basal concentrations (Chen et al., 2004; Gouaux, 2009). The transporters practically have to act as selective molecular vacuum cleaners to deal with such huge loads of neurotransmitters in order to re-establish pre-signaling conditions within milliseconds. A major ion gradient serves as driving force and patron for the protein class: the Neurotransmitter:Sodium Symporters (NSS).

Synonymously called the solute carrier 6 family (SLC-6), NSS members include the sodium- and chloride-dependent transporters for GABA, dopamine, serotonin, norepinephrine and glycine, but also just sodium-dependent transporters of amino acids. Thus, the protein family is of particular medical importance, as many CNS diseases like depression, anxiety or epilepsy can be targeted by inhibiting transporters (Iversen, 1971).

They share a basic scaffold consisting of 12 transmembrane regions (TMs), segments 1-5 and 6-10 forming two pseudosymmetric domains housing the substrate and ion binding sites in partially unwound regions half-way across the membrane (Kanner & Zomot, 2008).

The crystal structure of LeuT, a bacterial orthologue of the eukaryotic NSS members, became available in an occluded state conformation in 2005 and in the open to out orientation in 2008 (Singh et al., 2008; Yamashita et al., 2005), thus revealing first detailed insights into the binding site topology. Furthermore, very recently a double mutant stabilized in an inward-open conformation was published (Krishnamurthy & Gouaux, 2012). These crystallographic snapshots fortify the so-called alternating access model for neurotransmitter membrane transport (Jardetzky, 1966). Various attempts have been made to clarify the exact molecular transport mechanism (Forrest et al., 2008; Shi et al., 2008), yet many questions remain unanswered. Concerning the quaternary structure, it is generally assumed that neurotransmitter:sodium symporters form constitutive oligomers (Forrest et al., 2008; Sitte et al., 2004). Despite a comparably poor average overall sequence identity between eu- and prokaryotic SLC-6 members of slightly above 20%, these structures paved the way to comparative modeling studies. Predominantly the monoamine transporters DAT, NET and SERT, but also GAT, have been modeled and studied extensively. For a comprehensive summary of the state of knowledge about the SLC-6 family, the reader is referred to the recent review by Kristensen et al. (Kristensen et al., 2011).

3.2.2 Examples of studies on the hSERT

As mentioned earlier, especially when dealing with low template-target sequence identity, a very careful sequence alignment including all possible experimental knowledge is crucial

for the construction of reliable homology models. For the main members of the SLC-6 family a lot of effort has been put into this work, resulting in the comprehensive alignment of NSS sequences with the LeuT published by Beuming et al. in 2006 (Beuming et al., 2006). Since then, some new structural insights into the protein class have been gained leading to slightly altered regions, but still the alignments can be considered a good starting point for experiments with NSS models. In the case of the hSERT, the recent work of Sarker et al. (Sarker et al., 2010) provides a good example for the cumulative value of combining molecular modeling methods with mutagenesis experiments in order to verify *in silico* elaborated hypotheses. For investigating the binding mode of tricyclic antidepressants (TCAs) in the serotonin transporter, comparative modeling marked the starting point for subsequent studies. Using the Beuming alignment, homology models of hSERT were built based on the previously mentioned high-resolution open-to-out structure of the LeuT published in 2008 (PDB code 3F3A). Subsequent docking studies of imipramine resulted in three pose clusters of potential binding modes, showing interactions to previously reported key residues (Andersen et al., 2009; Chen & Rudnick, 2000; White et al., 2006). A diagnostic Y95F mutation, a candidate residue for hydrogen bonding with the imipramine diaminopropyl moiety, significantly decreased imipramine affinity without affecting serotonin binding, ruling out one cluster. Further uptake and docking assays demonstrated that carbamazepine, structurally a truncated and slightly more rigid derivative of imipramine, was able to bind mutually non-exclusive with the substrate serotonin, whereas binding of its large-tailed relative is mutually exclusive. This led to the following conclusions: a) the tricyclic ring system of TCAs binds in an outer vestibule, and b) the basic side chain of imipramine points into the actual substrate binding site.
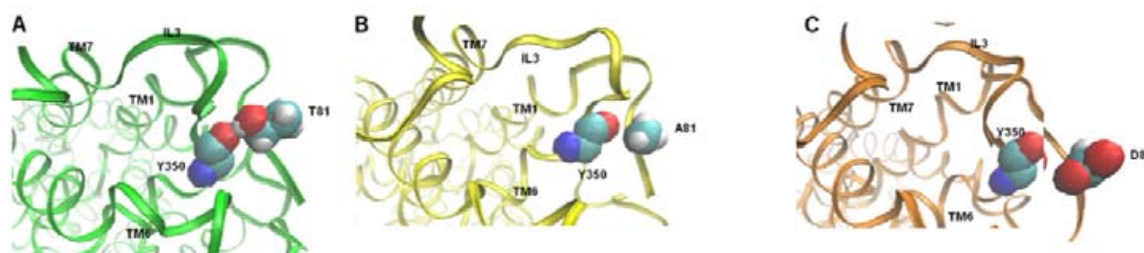
Fig. 4. Molecular dynamics simulations of SERTThr-81 mutants reveal models favoring inward facing states. A, snapshot of wild type SERT after 16 ns of MD simulation. The Thr81 side chain forms a stable H-bond with the backbone carbonyl of Tyr350 in IL3. B, snapshot of SERTT81A after 6 ns of MD simulation; the H-bond is not formed between Ala81 and Tyr350 during the course of the simulation. C, snapshot of SERTT81D after 6 ns of MD simulation; no H-bond is formed between Asp81 and Tyr350 during the course of the simulation. (taken from (Sucic et al., 2010)).

As an example for a more functional study on the SERT, the work of Sucic et al. (Sucic et al., 2010) can be mentioned. As it was analogously reported for the DAT (Guptaroy et al., 2009), the important role of a highly conserved phosphorylation site at the N-terminus of the transporter in mediating the action of amphetamines was studied. Amphetamines are said to induce substrate efflux, but the way they do so is not well understood. Sucic et al. reported that mutating the highly conserved N-terminal residue T81 (a candidate site for phosphorylation by protein kinase C), to alanine or aspartate leads to subsequent fail of the transporter to support amphetamine-induced efflux. As it was also confirmed by molecular

dynamics simulations of the wild type transporter, the *in silico* mutated SERT[T81A] and SERT[T81D], the data suggested that by phosphorylation or *in silico* mutation of T81 the conformational equilibrium of the serotonin transport cycle alters towards the inward facing conformation. As seen in the MD studies, this happens due to a loss of a hydrogen bond network of T81 with Y350 in IL3 by these mutations. Furthermore, an increased distance between the C terminus (i.e. the most distal point of TM12) and the N terminus after *in silico* mutation was observed. This example nicely indicates how functional MD studies might aid in elucidating biological relevant phenomena.

3.2.3 Studies on hGAT models

The four Na^+- and Cl^--dependent GABA transporters, GAT-1-3 and BGT-1 (SLC6A1, A16, A11, A12), provide a similar percentage of sequence identity to the LeuT. The subtype showing the highest quantity in the CNS is GAT-1. It is also the best-investigated, and the only one currently targeted by a marketed drug, the second-line antiepileptic tiagabine (Gabitril®). Accordingly, systematic synthesis studies in order to discover even more selective compounds have been performed mainly on GAT-1. Nevertheless, other subtypes should not be ignored, as they may be the key to a less side-effect afflicted antiepileptic therapy, as tiagabine efficacy as anticonvulsant is limited, and its use was connected to several adverse effects like sedation, agitation, or even seizure induction. Neuronal GABA reuptake, mainly done by GAT-1, leads to subsequent recycling of the transmitter substance. On the contrary, astroglial uptake of GABA leads to degradation, suggesting subtypes predominantly present in glia cells being an interesting target for enhancing overall GABA levels. For example, the lipophilic GABA analog EF-1502, characterized by GAT1 and GAT2 (BGT-1) selectivity, showed synergistic anticonvulsant activity, when administered with tiagabine (Schousboe et al., 2004), although BGT1 levels in the CNS are about 1000-fold lower, and even a recent study with BGT-1 knockout mice did not show any change in seizure susceptibility (Lehre et al., 2011).

In the search for potent selective non-GAT-1 inhibitors, GABA mimetic moieties (like R-nipecotic acid in tiagabine, β-alanine or THPO [4,5,6,7-Tetrahydroisoxazolo(4,5-c)pyridin-3-ol]) were systematically combined with large aromatic side chains, both in order to increase the affinity and to make the compounds blood-brain barrier permeable (Andersen et al., 1993; Andersen et al., 1999; Clausen et al., 2005; Knutsen et al., 1999; Kragler et al., 2008). Unfortunately, up to now no truly selective tools for the evaluation of non-GAT-1 inhibition are available, although the GAT-1/BGT-1 inhibitor EF1502 and SNAP-5114, showing a certain GAT-2/GAT-3 selectivity, mark a good starting point (Madsen et al., 2010). Thus, further insights into the molecular basis of ligand binding are sought by the aid of *in silico* methods.

GAT-1 has been subject of several comparative modeling studies. Initial studies predominantly aimed at clarifying the GABA binding mode in the occluded transporter state, which is quite well documented so far (Pallo et al., 2007; Wein & Wanner, 2009). Though, compounds with large aromatic tails cannot be accommodated in the occluded-state active site, as the entrance to the binding pocket is barred by the two extracellular gate residues R69 and D451, as well as the F294 side chain, forming the binding site "roof". In order to study tiagabine-like ligands, constructing open-to-out models seemed inevitable, as it was done by Skovstrup et al. (Skovstrup et al., 2010). Structures of both states were

modeled and refined exhaustively, as described in section 2.1. The combined use of docking and molecular dynamics simulation was chosen to investigate binding of GABA, its analogue (R)-nipecotic acid and the high active (R)-enantiomer of tiagabine. The results for GABA binding were in line with the earlier mentioned experiments. In case of tiagabine, MD simulations helped to distinguish between the *cis-* and *trans-* conformer, both being possible states due to the protonated state of tiagabine at physiological pH. During the MD, the *trans-* conformer immediately stirred away to the extracellular space, whereas the other one remained stable in the binding site. Summing up, GABA and (R)-tiagabine turned out having two different binding modes, sharing the orientation of the carboxy group towards one of the co-transported sodium ions as a common feature.

For the other GAT subtypes, things are a bit more complicated. Looking at the residues corresponding to LeuT substrate binding site, just a few candidate residues differ significantly, being somehow unlikely to be fully responsible for subtype selective binding. So far, molecular modeling studies have been performed, but highly similar binding sites and the lack of selective ligand data limited their explanatory power (Pallo et al., 2009). Thus, a huge field of activity remains to be explored on the way to fully understand the differences between the GABA subtypes, *in silico* methods being a valuable tool for stepwise adding pieces of information to the big puzzle.

4. Concluding remarks

Membrane transport proteins are responsible for one of the most important processes in living cells: directed transport across barriers. They comprise about 30% of known proteomes and constitute about 50% of pharmacological targets. Although, due to difficulties in expression, purification and crystallization, only about 2% of the high resolution crystal structures in the Protein Data Bank (PDB) are transporters. Thus, computational methods have been utilized extensively to provide significant new insights into protein structure and function. Above all, molecular modeling and molecular dynamics (MD) simulations may deliver atomic level details to reveal the molecular basis of e.g. drug-transporter interactions. As shown on basis of recent research examples, *in silico* methods in many cases can provide additional information to biological experiments, either underpinning pharmacological results or they may even lead to new insight, not being biologically accessible.

5. Acknowledgments

6. References

Agre, P. (2006). The Aquaporin Water Channels. *Proc Am Thorac Soc*, Vol.3, No.1, pp. 5-13

Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L. & Chang, G. (2009). Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, Vol.323, No.5922, pp. 1718-22

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, Vol.25, No.17, pp. 3389-402

Amaro, R.E. & Li, W.W. (2010). Emerging methods for ensemble-based virtual screening. *Curr Top Med Chem*, Vol.10, No.1, pp. 3-13

Andersen, J., Taboureau, O., Hansen, K.B., Olsen, L., Egebjerg, J., Stromgaard, K. & Kristensen, A.S. (2009). Location of the antidepressant binding site in the serotonin transporter: importance of Ser-438 in recognition of citalopram and tricyclic antidepressants. *J Biol Chem*, Vol.284, No.15, pp. 10276-84

Andersen, K.E., Braestrup, C., Gronwald, F.C., Jorgensen, A.S., Nielsen, E.B., Sonnewald, U., Sorensen, P.O., Suzdak, P.D. & Knutsen, L.J. (1993). The synthesis of novel GABA uptake inhibitors. 1. Elucidation of the structure-activity studies leading to the choice of (R)-1-[4,4-bis(3-methyl-2-thienyl)-3-butenyl]-3-piperidinecarboxylic acid (tiagabine) as an anticonvulsant drug candidate. *J Med Chem*, Vol.36, No.12, pp. 1716-25

Andersen, K.E., Sorensen, J.L., Huusfeldt, P.O., Knutsen, L.J., Lau, J., Lundt, B.F., Petersen, H., Suzdak, P.D. & Swedberg, M.D. (1999). Synthesis of novel GABA uptake inhibitors. 4. Bioisosteric transformation and successive optimization of known GABA uptake inhibitors leading to a series of potent anticonvulsant drug candidates. *J Med Chem*, Vol.42, No.21, pp. 4281-91

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, Vol.22, No.2, pp. 195-201

B-Rao, C., Subramanian, J. & Sharma, S.D. (2009). Managing protein flexibility in docking and its applications. *Drug Discov Today*, Vol.14, No.7-8, pp. 394-400

Becker, J.P., Depret, G., Van Bambeke, F., Tulkens, P.M. & Prevost, M. (2009). Molecular models of human P-glycoprotein in two different catalytic states. *BMC Struct Biol*, Vol.9, 3

Becker, J.P., Van Bambeke, F., Tulkens, P.M. & Prevost, M. (2010). Dynamics and structural changes induced by ATP binding in SAV1866, a bacterial ABC exporter. *Journal of Physical Chemistry B*, Vol.114, No.48, pp. 15948-57

Benkert, P., Biasini, M. & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, Vol.27, No.3, pp. 343-50

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res*, Vol.28, No.1, pp. 235-42

Beuming, T., Shi, L., Javitch, J.A. & Weinstein, H. (2006). A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na+ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Mol Pharmacol*, Vol.70, No.5, pp. 1630-42

Bohm, H.J. (1992). LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des*, Vol.6, No.6, pp. 593-606

Bohm, H.J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, Vol.8, No.3, pp. 243-56

Bohm, H.J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des*, Vol.12, No.4, pp. 309-23

Borst, P. & Elferink, R.O. (2002). Mammalian ABC transporters in health and disease. *Annu Rev Biochem*, Vol.71, 537-92

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*, Vol.4, No.2, pp. 187-217

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, Vol.4, No.2, pp. 187-217

Caffrey, M. (2003). Membrane protein crystallization. *J Struct Biol*, Vol.142, No.1, pp. 108-132

Caflisch, A., Miranker, A. & Karplus, M. (1993). Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J Med Chem*, Vol.36, No.15, pp. 2142-67

Campbell, J.D., Biggin, P.C., Baaden, M. & Sansom, M.S. (2003). Extending the structure of an ABC transporter to atomic resolution: modeling and simulation studies of MsbA. *Biochemistry*, Vol.42, No.13, pp. 3666-73

Campbell, J.D. & Sansom, M.S. (2005). Nucleotide binding to the homodimeric MJ0796 protein: a computational study of a prokaryotic ABC transporter NBD dimer. *FEBS Lett*, Vol.579, No.19, pp. 4193-9

Cavasotto, C.N. (2011). Homology models in docking and high-throughput docking. *Curr Top Med Chem*, Vol.11, No.12, pp. 1528-34

Chang, G. (2003). Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. *J Mol Biol*, Vol.330, No.2, pp. 419-30

Chang, G. & Roth, C.B. (2001). Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science*, Vol.293, No.5536, pp. 1793-800

Chang, G., Roth, C.B., Reyes, C.L., Pornillos, O., Chen, Y.J. & Chen, A.P. (2006). Retraction. *Science*, Vol.314, No.5807, pp. 1875

Charifson, P.S., Corkery, J.J., Murcko, M.A. & Walters, W.P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem*, Vol.42, No.25, pp. 5100-9

Chen, J.G. & Rudnick, G. (2000). Permeation and gating residues in serotonin transporter. *Proc Natl Acad Sci U S A*, Vol.97, No.3, pp. 1044-9

Chen, N.H., Reith, M.E. & Quick, M.W. (2004). Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6. *Pflugers Arch*, Vol.447, No.5, pp. 519-31

Christen, M. & van Gunsteren, W.F. (2008). On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *Journal of Computational Chemistry*, Vol.29, No.2, pp. 157-166

Clausen, R.P., Moltzen, E.K., Perregaard, J., Lenz, S.M., Sanchez, C., Falch, E., Frolund, B., Bolvig, T., Sarup, A., Larsson, O.M., Schousboe, A. & Krogsgaard-Larsen, P. (2005). Selective inhibitors of GABA uptake: synthesis and molecular pharmacology of 4-N-methylamino-4,5,6,7-tetrahydrobenzo[d]isoxazol-3-ol analogues. *Bioorg Med Chem*, Vol.13, No.3, pp. 895-908

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. & Kollman, P.A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, Vol.117, No.19, pp. 5179-5197

Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., Orozco, M., Pertinhez, T.A., Rizzi, M. & Sotriffer, C.A. (2008). Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, Vol.51, No.20, pp. 6237-55

Crowley, E., O'Mara, M.L., Kerr, I.D. & Callaghan, R. (2010). Transmembrane helix 12 plays a pivotal role in coupling energy provision and drug binding in ABCB1. *FEBS J*, Vol.277, No.19, pp. 3974-85

Dalton, J.A. & Jackson, R.M. (2007). An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics*, Vol.23, No.15, pp. 1901-8

Davis, A.M., St-Gallay, S.A. & Kleywegt, G.J. (2008). Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov Today*, Vol.13, No.19-20, pp. 831-41

Dawson, R.J. & Locher, K.P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature*, Vol.443, No.7108, pp. 180-5

Deane, C.M. & Blundell, T.L. (2001). CODA: a combined algorithm for predicting the structurally variable regions of protein models. *Protein Sci*, Vol.10, No.3, pp. 599-612

Dey, S., Hafkemeyer, P., Pastan, I. & Gottesman, M.M. (1999). A single amino acid residue contributes to distinct mechanisms of inhibition of the human multidrug transporter by stereoisomers of the dopamine receptor antagonist flupentixol. *Biochemistry*, Vol.38, No.20, pp. 6630-9

Dolghih, E., Bryant, C., Renslo, A.R. & Jacobson, M.P. (2011). Predicting binding to p-glycoprotein by flexible receptor docking. *PLoS Comput Biol*, Vol.7, No.6, pp. e1002083

Doyle, L.A., Yang, W., Abruzzo, L.V., Krogmann, T., Gao, Y., Rishi, A.K. & Ross, D.D. (1998). A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc Natl Acad Sci U S A*, Vol.95, No.26, pp. 15665-70

Ecker, G.F. & Chiba, P. (2009). ABC Transporters - From Targets to Antitargets? Transporters as Drug Carriers. G. F. Ecker and P. Chiba. Weinheim, Wiley-VCH. 1: 349-362

Ecker, G.F., Stockner, T. & Chiba, P. (2008). Computational models for prediction of interactions with ABC-transporters. *Drug Discov Today*, Vol.13, No.7-8, pp. 311-7

Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. & Mee, R.P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to

estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, Vol.11, No.5, pp. 425-45

Evers, A. & Klebe, G. (2004). Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem*, Vol.47, No.22, pp. 5381-92

Ewing, T.J., Makino, S., Skillman, A.G. & Kuntz, I.D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*, Vol.15, No.5, pp. 411-28

Forrest, L.R., Zhang, Y.W., Jacobs, M.T., Gesmonde, J., Xie, L., Honig, B.H. & Rudnick, G. (2008). Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci U S A*, Vol.105, No.30, pp. 10338-43

Giacomini, K.M., Huang, S.M., Tweedie, D.J., Benet, L.Z., Brouwer, K.L., Chu, X., Dahlin, A., Evers, R., Fischer, V., Hillgren, K.M., Hoffmaster, K.A., Ishikawa, T., Keppler, D., Kim, R.B., Lee, C.A., Niemi, M., Polli, J.W., Sugiyama, Y., Swaan, P.W., Ware, J.A., Wright, S.H., Yee, S.W., Zamek-Gliszczynski, M.J. & Zhang, L. (2010). Membrane transporters in drug development. *Nat Rev Drug Discov*, Vol.9, No.3, pp. 215-36

Gohlke, H., Hendlich, M. & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*, Vol.295, No.2, pp. 337-56

Goodsell, D.S., Morris, G.M. & Olson, A.J. (1996). Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition*, Vol.9, No.1, pp. 1-5

Goodsell, D.S. & Olson, A.J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, Vol.8, No.3, pp. 195-202

Gouaux, E. (2009). Review. The molecular logic of sodium-coupled neurotransmitter transporters. *Philos Trans R Soc Lond B Biol Sci*, Vol.364, No.1514, pp. 149-54

Gumbart, J., Wang, Y., Aksimentiev, A., Tajkhorshid, E. & Schulten, K. (2005). Molecular dynamics simulations of proteins in lipid bilayers. *Curr Opin Struct Biol*, Vol.15, No.4, pp. 423-431

Guptaroy, B., Zhang, M., Bowton, E., Binda, F., Shi, L., Weinstein, H., Galli, A., Javitch, J.A., Neubig, R.R. & Gnegy, M.E. (2009). A juxtamembrane mutation in the N terminus of the dopamine transporter induces preference for an inward-facing conformation. *Mol Pharmacol*, Vol.75, No.3, pp. 514-24

Gyimesi, G., Ramachandran, S., Kota, P., Dokholyan, N.V., Sarkadi, B. & Hegedus, T. (2011). ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions. *Biochim Biophys Acta*, Vol.

Hendlich, M., Rippmann, F. & Barnickel, G. (1997). LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics & Modelling*, Vol.15, No.6, pp. 359-63, 389

Henrich, S., Salo-Ahen, O.M., Huang, B., Rippmann, F.F., Cruciani, G. & Wade, R.C. (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit*, Vol.23, No.2, pp. 209-19

Hollt, V., Kouba, M., Dietel, M. & Vogt, G. (1992). Stereoisomers of calcium antagonists which differ markedly in their potencies as calcium blockers are equally effective in modulating drug transport by P-glycoprotein. *Biochem Pharmacol*, Vol.43, No.12, pp. 2601-8

Höltje, H.D., Sippl, W., Rognan, D. & Folkers, G. (2008). Molecular Modeling. Basic Principles and Applications. Weinheim, Germany, Wiley-VCH Verlag GmbH & CoKGaA

Huang, N., Kalyanaraman, C., Irwin, J.J. & Jacobson, M.P. (2006). Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J Chem Inf Model*, Vol.46, No.1, pp. 243-53

Huang, S.Y., Grinter, S.Z. & Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys*, Vol.12, No.40, pp. 12899-908

Huang, S.Y. & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem*, Vol.27, No.15, pp. 1866-75

Huang, S.Y. & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J Comput Chem*, Vol.27, No.15, pp. 1876-82

Huang, S.Y. & Zou, X. (2008). An iterative knowledge-based scoring function for protein-protein recognition. *Proteins*, Vol.72, No.2, pp. 557-79

Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci*, Vol.5, No.2, pp. 212-20

Irvine, S.W.l.a.U. Membrane Proteins of Known 3D Structure, Available from: http://blanco.biomol.uci.edu/mpstruc/listAll/list

Isralewitz, B., Gao, M. & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, Vol.11, No.2, pp. 224-230

Iversen, L.L. (1971). Role of transmitter uptake mechanisms in synaptic neurotransmission. *Br J Pharmacol*, Vol.41, No.4, pp. 571-91

Jabeen, I., Wetwitayaklung, P., Klepsch, F., Parveen, Z., Chiba, P. & Ecker, G.F. (2011). Probing the stereoselectivity of P-glycoprotein-synthesis, biological activity and ligand docking studies of a set of enantiopure benzopyrano[3,4-b][1,4]oxazines. *Chem Commun (Camb)*, Vol.47, No.9, pp. 2586-8

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature*, Vol.211, No.5052, pp. 969-70

Jiang, F. & Kim, S.H. (1991). "Soft docking": matching of molecular surface cubes. *J Mol Biol*, Vol.219, No.1, pp. 79-102

Jones, G., Willett, P. & Glen, R.C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol*, Vol.245, No.1, pp. 43-53

Jones, P.M. & George, A.M. (2007). Nucleotide-dependent allostery within the ABC transporter ATP-binding cassette: a computational study of the MJ0796 dimer. *J Biol Chem*, Vol.282, No.31, pp. 22793-803

Jones, P.M. & George, A.M. (2009). Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle. *Proteins*, Vol.75, No.2, pp. 387-96

Jorgensen, W.L. & Rives, T. (1988). The OPLS Potential Functions For Proteins - Energy Minimizations For Crystals Of Cyclic-Peptides And Crambin. *Journal of the American Chemical Society*, Vol.110, No.6, pp. 1657-1666

Kandt, C., Ash, W. & Tieleman, D.P. (July 2009). InflateGRO, Available from: http://www.csb.bit.uni-bonn.de/inflategro.html

Kandt, C., Ash, W.L. & Tieleman, D.P. (2007). Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, Vol.41, No.4, pp. 475-488

Kanner, B.I. & Zomot, E. (2008). Sodium-coupled neurotransmitter transporters. *Chem Rev*, Vol.108, No.5, pp. 1654-68

Karplus, M. & McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*, Vol.9, No.9, pp. 646-652

Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, Vol.3, No.11, pp. 935-49

Klepsch, F., Chiba, P. & Ecker, G.F. (2011). Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of P-glycoprotein. *PLoS Comput Biol*, Vol.7, No.5, pp. e1002036

Klepsch, F., Jabeen, I., Chiba, P. & Ecker, G.F. (2010). Pharmacoinformatic approaches to design natural product type ligands of ABC-transporters. *Curr Pharm Des*, Vol.16, No.15, pp. 1742-52

Knegtel, R.M., Kuntz, I.D. & Oshiro, C.M. (1997). Molecular docking to ensembles of protein structures. *J Mol Biol*, Vol.266, No.2, pp. 424-40

Knutsen, L.J., Andersen, K.E., Lau, J., Lundt, B.F., Henry, R.F., Morton, H.E., Naerum, L., Petersen, H., Stephensen, H., Suzdak, P.D., Swedberg, M.D., Thomsen, C. & Sorensen, P.O. (1999). Synthesis of novel GABA uptake inhibitors. 3. Diaryloxime and diarylvinyl ether derivatives of nipecotic acid and guvacine as anticonvulsant agents. *J Med Chem*, Vol.42, No.18, pp. 3447-62

Kragler, A., Hofner, G. & Wanner, K.T. (2008). Synthesis and biological evaluation of aminomethylphenol derivatives as inhibitors of the murine GABA transporters mGAT1-mGAT4. *European Journal of Medicinal Chemistry*, Vol.43, No.11, pp. 2404-11

Krishnamurthy, H. & Gouaux, E. (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*, doi:10.1038/nature10737 [Epub ahead of print]

Kristensen, A.S., Andersen, J., Jorgensen, T.N., Sorensen, L., Eriksen, J., Loland, C.J., Stromgaard, K. & Gether, U. (2011). SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol Rev*, Vol.63, No.3, pp. 585-640

Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C., Abagyan, R. & participants, G.D. (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*, Vol.19, No.8, pp. 1108-26

Kuhn, D., Weskamp, N., Hullermeier, E. & Klebe, G. (2007). Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem*, Vol.2, No.10, pp. 1432-47

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. & Ferrin, T.E. (1982). A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, Vol.161, No.2, pp. 269-88

Laskowski, R.A., Macarthur, M.W., Moss, D.S. & Thornton, J.M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, Vol.26, 283-291

Leach, A.R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol*, Vol.235, No.1, pp. 345-56

Lehre, A.C., Rowley, N.M., Zhou, Y., Holmseth, S., Guo, C., Holen, T., Hua, R., Laake, P., Olofsson, A.M., Poblete-Naredo, I., Rusakov, D.A., Madsen, K.K., Clausen, R.P., Schousboe, A., White, H.S. & Danbolt, N.C. (2011). Deletion of the betaine-GABA transporter (BGT1; slc6a12) gene does not affect seizure thresholds of adult mice. *Epilepsy Res*, Vol.95, No.1-2, pp. 70-81

Lesk, A.M. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol*, Vol.136, No.3, pp. 225-70

Lindahl, E. & Sansom, M.S.P. (2008). Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol*, Vol.18, No.4, pp. 425-431

Loo, T.W., Bartlett, M.C. & Clarke, D.M. (2003). Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein. *J Biol Chem*, Vol.278, No.41, pp. 39706-10

Loo, T.W., Bartlett, M.C. & Clarke, D.M. (2010). Human P-glycoprotein is active when the two halves are clamped together in the closed conformation. *Biochem Biophys Res Commun*, Vol.395, No.3, pp. 436-40

Loo, T.W. & Clarke, D.M. (1999). The transmembrane domains of the human multidrug resistance P-glycoprotein are sufficient to mediate drug binding and trafficking to the cell surface. *J Biol Chem*, Vol.274, No.35, pp. 24759-65

Loo, T.W. & Clarke, D.M. (2002). Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein. *J Biol Chem*, Vol.277, No.46, pp. 44332-8

Loo, T.W. & Clarke, D.M. (2008). Mutational analysis of ABC proteins. *Arch Biochem Biophys*, Vol.476, No.1, pp. 51-64

Lu, H. & Schulten, K. (1999). Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Structure, Function, and Bioinformatics*, Vol.35, No.4, pp. 453-463

Luurtsema, G., Molthoff, C.F., Windhorst, A.D., Smit, J.W., Keizer, H., Boellaard, R., Lammertsma, A.A. & Franssen, E.J. (2003). (R)- and (S)-[11C]verapamil as PET-tracers for measuring P-glycoprotein function: in vitro and in vivo evaluation. *Nucl Med Biol*, Vol.30, No.7, pp. 747-51

Madsen, K.K., White, H.S. & Schousboe, A. (2010). Neuronal and non-neuronal GABA transporters as targets for antiepileptic drugs. *Pharmacol Ther*, Vol.125, No.3, pp. 394-401

Malde, A.K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P.C., Oostenbrink, C. & Mark, A.E. (2011). An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J Chem Theory Comput*, Vol.7, No.12, pp. 4026-4037

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, Vol.29, 291-325

Matthews, B.W. (2007). Five retracted structure reports: inverted or incorrect? *Protein Sci*, Vol.16, No.6, pp. 1013-6

McCammon, J.A., Gelin, B.R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, Vol.267, No.5612, pp. 585-590

Mealey, K.L., Bentjen, S.A., Gay, J.M. & Cantor, G.H. (2001). Ivermectin sensitivity in collies is associated with a deletion mutation of the mdr1 gene. *Pharmacogenetics*, Vol.11, No.8, pp. 727-33

Mehler, E.L., Periole, X., Hassan, S.A. & Weinstein, H. (2002). Key issues in the computational simulation of GPCR function: representation of loop domains. *J Comput Aided Mol Des*, Vol.16, No.11, pp. 841-53

Mizutani, M.Y., Tomioka, N. & Itai, A. (1994). Rational automatic search method for stable docking models of protein and ligand. *J Mol Biol*, Vol.243, No.2, pp. 310-26

Moore, G. (1965). Cramming more components onto integrated circuits *Electronics*, Vol.38, No.8, pp.

Neuhaus, W. & Noe, C.R. (2009). Transport at the Blood-Brain Barrier. Transporters as Drug Carriers. G. F. Ecker and P. Chiba. Weinheim, Wiley-VCH. 1: 263-298

Newby, Z.E.R., O'Connell, J.D., Gruswitz, F., Hays, F.A., Harries, W.E.C., Harwood, I.M., Ho, J.D., Lee, J.K., Savage, D.F., Miercke, L.J.W. & Stroud, R.M. (2009). A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat. Protocols*, Vol.4, No.5, pp. 619-637

Newstead, S., Fowler, P.W., Bilton, P., Carpenter, E.P., Sadler, P.J., Campopiano, D.J., Sansom, M.S. & Iwata, S. (2009). Insights into how nucleotide-binding domains power ABC transport. *Structure*, Vol.17, No.9, pp. 1213-22

Notredame, C., Higgins, D.G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, Vol.302, No.1, pp. 205-17

Oliveira, A.S., Baptista, A.M. & Soares, C.M. (2011). Conformational changes induced by ATP-hydrolysis in an ABC transporter: a molecular dynamics study of the Sav1866 exporter. *Proteins*, Vol.79, No.6, pp. 1977-90

Olson, A.J., Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E. & Belew, R.K. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, Vol.19, No.14, pp. 1639-1662

Omote, H. & Al-Shawi, M.K. (2006). Interaction of transported drugs with the lipid bilayer and P-glycoprotein through a solvation exchange mechanism. *Biophysical Journal*, Vol.90, No.11, pp. 4046-59

Pajeva, I.K., Globisch, C. & Wiese, M. (2009). Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. *ChemMedChem*, Vol.4, No.11, pp. 1883-96

Pallo, A., Bencsura, A., Heja, L., Beke, T., Perczel, A., Kardos, J. & Simon, A. (2007). Major human gamma-aminobutyrate transporter: in silico prediction of substrate efficacy. *Biochem Biophys Res Commun*, Vol.364, No.4, pp. 952-8

Pallo, A., Simon, A., Bencsura, A., Heja, L. & Kardos, J. (2009). Substrate-Na+ complex formation: coupling mechanism for gamma-aminobutyrate symporters. *Biochem Biophys Res Commun*, Vol.385, No.2, pp. 210-4

Park, S., Kono, H., Wang, W., Boder, E.T. & Saven, J.G. (2005). Progress in the development and application of computational methods for probabilistic protein design. *Comput Chem Eng*, Vol.29, No.3, pp. 407-421

Pearson, W.R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol*, Vol.183, 63-98

Penders, B., Horstman, K. & Vos, R. (2007). Proper science in moist biology. *EMBO Rep*, Vol.8, No.7, pp. 613

Pleban, K., Kaiser, D., Kopp, S., Peer, M., Chiba, P. & Ecker, G.F. (2005). Targeting drug-efflux pumps -- a pharmacoinformatic approach. *Acta Biochim Pol*, Vol.52, No.3, pp. 737-40

Pleban, K., Kopp, S., Csaszar, E., Peer, M., Hrebicek, T., Rizzi, A., Ecker, G.F. & Chiba, P. (2005). P-glycoprotein substrate binding domains are located at the transmembrane domain/transmembrane domain interfaces: a combined photoaffinity labeling-protein homology modeling approach. *Mol Pharmacol*, Vol.67, No.2, pp. 365-74

Qu, Q. & Sharom, F.J. (2002). Proximity of bound Hoechst 33342 to the ATPase catalytic sites places the drug binding site of P-glycoprotein within the cytoplasmic membrane leaflet. *Biochemistry*, Vol.41, No.14, pp. 4744-52

Rabindran, S.K., Ross, D.D., Doyle, L.A., Yang, W. & Greenberger, L.M. (2000). Fumitremorgin C reverses multidrug resistance in cells transfected with the breast cancer resistance protein. *Cancer Res*, Vol.60, No.1, pp. 47-50

Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, Vol.261, No.3, pp. 470-89

Reyes, C.L. & Chang, G. (2005). Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science*, Vol.308, No.5724, pp. 1028-31

Reynolds, C.A., Wade, R.C. & Goodford, P.J. (1989). Identifying targets for bioreductive agents: using GRID to predict selective binding regions of proteins. *J Mol Graph*, Vol.7, No.2, pp. 103-8, 100

Rosenberg, M.F., Velarde, G., Ford, R.C., Martin, C., Berridge, G., Kerr, I.D., Callaghan, R., Schmidlin, A., Wooding, C., Linton, K.J. & Higgins, C.F. (2001). Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *EMBO J*, Vol.20, No.20, pp. 5615-25

Saier, M.H., Tran, C.V. & Barabote, R.D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res*, Vol.34, No.suppl 1, pp. D181-D186

Saier, M.H., Yen, M.R., Noto, K., Tamang, D.G. & Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic Acids Res*, Vol.37, No.suppl 1, pp. D274-D278

Sarker, S., Weissensteiner, R., Steiner, I., Sitte, H.H., Ecker, G.F., Freissmuth, M. & Sucic, S. (2010). The high-affinity binding site for tricyclic antidepressants resides in the outer vestibule of the serotonin transporter. *Mol Pharmacol*, Vol.78, No.6, pp. 1026-35

Schinkel, A.H. & Jonker, J.W. (2003). Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. *Adv Drug Deliv Rev*, Vol.55, No.1, pp. 3-29

Schlitter, J., Engels, M. & Krüger, P. (1994). Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*, Vol.12, No.2, pp. 84-89

Schmidt, D., Jiang, Q.-X. & MacKinnon, R. (2006). Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature*, Vol.444, No.7120, pp. 775-779

Schousboe, A., Sarup, A., Larsson, O.M. & White, H.S. (2004). GABA transporters as drug targets for modulation of GABAergic activity. *Biochem Pharmacol*, Vol.68, No.8, pp. 1557-63

Schrauber, H., Eisenhaber, F. & Argos, P. (1993). Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J Mol Biol*, Vol.230, No.2, pp. 592-612

Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. (1999). The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A*, Vol.103, No.19, pp. 3596-3607

Seeger, M.A. & van Veen, H.W. (2009). Molecular basis of multidrug transport by ABC transporters. *Biochim Biophys Acta*, Vol.1794, No.5, pp. 725-37

Seigneuret, M. & Garnier-Suillerot, A. (2003). A structural model for the open conformation of the mdr1 P-glycoprotein based on the MsbA crystal structure. *J Biol Chem*, Vol.278, No.32, pp. 30115-24

Shapiro, A.B., Fox, K., Lam, P. & Ling, V. (1999). Stimulation of P-glycoprotein-mediated drug transport by prazosin and progesterone. Evidence for a third drug-binding site. *Eur J Biochem*, Vol.259, No.3, pp. 841-50

Sharom, F.J. (2008). ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics*, Vol.9, No.1, pp. 105-27

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A. & Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem*, Vol.49, No.2, pp. 534-53

Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J.A. (2008). The mechanism of a neurotransmitter:sodium symporter--inward release of Na+ and substrate is triggered by substrate in a second binding site. *Mol Cell*, Vol.30, No.6, pp. 667-77

Shilling, R.A., Balakrishnan, L., Shahi, S., Venter, H. & van Veen, H.W. (2003). A new dimer interface for an ABC transporter. *Int J Antimicrob Agents*, Vol.22, No.3, pp. 200-4

Singh, S.K., Piscitelli, C.L., Yamashita, A. & Gouaux, E. (2008). A competitive inhibitor traps LeuT in an open-to-out conformation. *Science*, Vol.322, No.5908, pp. 1655-61

Sitte, H.H., Farhan, H. & Javitch, J.A. (2004). Sodium-dependent neurotransmitter transporters: oligomerization as a determinant of transporter function and trafficking. *Mol Interv*, Vol.4, No.1, pp. 38-47

Skovstrup, S., Taboureau, O., Brauner-Osborne, H. & Jorgensen, F.S. (2010). Homology modelling of the GABA transporter and analysis of tiagabine binding. *ChemMedChem*, Vol.5, No.7, pp. 986-1000

Sousa, S.F., Fernandes, P.A. & Ramos, M.J. (2006). Protein-ligand docking: current status and future challenges. *Proteins*, Vol.65, No.1, pp. 15-26

Stenham, D.R., Campbell, J.D., Sansom, M.S., Higgins, C.F., Kerr, I.D. & Linton, K.J. (2003). An atomic detail model for the human ATP binding cassette transporter P-

glycoprotein derived from disulfide cross-linking and homology modeling. *FASEB J*, Vol.17, No.15, pp. 2287-9

Stockner, T., de Vries, S.J., Bonvin, A.M., Ecker, G.F. & Chiba, P. (2009). Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. *FEBS J*, Vol.276, No.4, pp. 964-72

Stouten, P.F.W. & Brady, G.P. (2000). Fast prediction and visualization of protein binding pockets with PASS. *Journal of Computer-Aided Molecular Design*, Vol.14, No.4, pp. 383-401

Sucic, S., Dallinger, S., Zdrazil, B., Weissensteiner, R., Jorgensen, T.N., Holy, M., Kudlacek, O., Seidel, S., Cha, J.H., Gether, U., Newman, A.H., Ecker, G.F., Freissmuth, M. & Sitte, H.H. (2010). The N terminus of monoamine transporters is a lever required for the action of amphetamines. *J Biol Chem*, Vol.285, No.14, pp. 10924-38

Szakacs, G., Varadi, A., Ozvegy-Laczka, C. & Sarkadi, B. (2008). The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov Today*, Vol.13, No.9-10, pp. 379-93

Tanaka, S. & Scheraga, H.A. (1976). Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, Vol.9, No.6, pp. 945-50

Theobald, D.L. & Miller, C. (2010). Membrane transport proteins: surprises in structural sameness. *Nat Struct Mol Biol*, Vol.17, 2–3

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, Vol.22, No.22, pp. 4673-80

Tsuruo, T., Iida, H., Tsukagoshi, S. & Sakurai, Y. (1981). Overcoming of vincristine resistance in P388 leukemia in vivo and in vitro through enhanced cytotoxicity of vincristine and vinblastine by verapamil. *Cancer Res*, Vol.41, No.5, pp. 1967-72

Vandevuer, S., Van Bambeke, F., Tulkens, P.M. & Prevost, M. (2006). Predicting the three-dimensional structure of human P-glycoprotein in absence of ATP by computational techniques embodying crosslinking data: insight into the mechanism of ligand migration and binding sites. *Proteins*, Vol.63, No.3, pp. 466-78

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. & Taylor, R.D. (2003). Improved protein-ligand docking using GOLD. *Proteins*, Vol.52, No.4, pp. 609-23

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, Vol.8, No.1, pp. 52-6, 29

Wang, R., Lai, L. & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des*, Vol.16, No.1, pp. 11-26

Wang, Y., Shaikh, S.A. & Tajkhorshid, E. (2010). Exploring transmembrane diffusion pathways with molecular dynamics. *Physiology*, Vol.25, No.3, pp. 142-54

Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. (2007). Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A*, Vol.104, No.48, pp. 19005-10

Wein, T. & Wanner, K.T. (2009). Generation of a 3D model for human GABA transporter hGAT-1 using molecular modeling and investigation of the binding of GABA. *Journal of Molecular Modeling*, Vol.

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *Journal of the American Chemical Society*, Vol.106, No.3, pp. 765-784

Welch, W., Ruppert, J. & Jain, A.N. (1996). Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol*, Vol.3, No.6, pp. 449-62

Wen, P.C. & Tajkhorshid, E. (2008). Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis. *Biophysical Journal*, Vol.95, No.11, pp. 5100-10

White, K.J., Kiser, P.D., Nichols, D.E. & Barker, E.L. (2006). Engineered zinc-binding sites confirm proximity and orientation of transmembrane helices I and III in the human serotonin transporter. *Protein Sci*, Vol.15, No.10, pp. 2411-22

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D. & Wishart, D.S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*, Vol.31, No.13, pp. 3316-9

Wolf, M.G., Hoefling, M., Aponte-Santamaria, C., Grubmüller, H. & Groenhof, G. (2010). g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *J Comput Chem*, Vol.31, 2169–2174

Yamashita, A., Singh, S.K., Kawate, T., Jin, Y. & Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature*, Vol.437, No.7056, pp. 215-23

Zhao, S., Zhu, K., Li, J. & Friesner, R.A. (2011). Progress in super long loop prediction. *Proteins*, Vol.79, No.10, pp. 2920-35

Zhou, H. & Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci*, Vol.11, No.11, pp. 2714-26

Chapter 3

Results and Discussion

Contribution of the thesis author

Probing the stereoselectivity of P-glycoprotein—synthesis, biological activity and ligand docking studies of a set of enantiopure benzopyrano[3,4-b][1,4]oxazines   I. Jabeen, P. Wetwitayaklung, F. Klepsch, Z. Parveen, P. Chiba and G.F. Ecker

F. Klepsch was involved in the computational part of this study. This included the generation of the homology model and guidance throughout the docking study and analysis.

Exhaustive Sampling of Docking Poses Reveals Binding Hypotheses for Propafenone Type Inhibitors of P-Glycoprotein   F. Klepsch, P. Chiba and G.F. Ecker

F. Klepsch conceived and designed the experiments, together with G.F. Ecker and P. Chiba. F. Klepsch performed the experiments and analyzed the data. F. Klepsch and G.F. Ecker wrote the manuscript.

Ligand and structure-based classification models for Prediction of P-glycoprotein inhibitors   F. Klepsch, V. Poongavanam and G.F. Ecker

F. Klepsch and V. Poongavanam contributed equally to this work. F. Klepsch designed and performed the structure-based classification part of the study. V. Poongavanam designed the machine-learning part, which was executed by F. Klepsch. F. Klepsch and G.F. Ecker wrote the manuscript.

ChemComm

This article is part of the

Enzymes & Proteins web themed issue

This issue showcases high quality research in the field of enzymes and proteins.

Please visit the website to access the other articles in this issue:-
http://www.rsc.org/chemcomm/enzymesandproteins
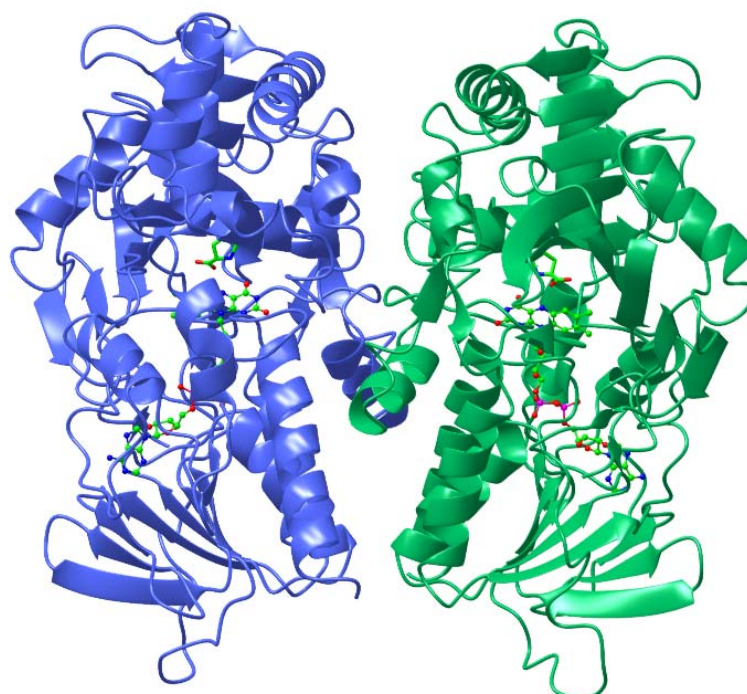
ChemComm

COMMUNICATION

Probing the stereoselectivity of P-glycoprotein—synthesis, biological activity and ligand docking studies of a set of enantiopure benzopyrano[3,4-*b*][1,4]oxazines†‡

Ishrat Jabeen,[a] Penpun Wetwitayaklung,[ab] Freya Klepsch,[a] Zahida Parveen,[c] Peter Chiba[c] and Gerhard F. Ecker*[a]

A series of enantiomerically pure benzopyrano[3,4-*b*][1,4]-oxazines have been synthesised and tested for their ability to inhibit P-glycoprotein. Reducing the conformational flexibility of the molecules leads to remarkable differences in the activity of diastereoisomers. Docking studies into a homology model of human P-gp provide first insights into potential binding areas for these compounds.

P-glycoprotein (P-gp) is a transmembrane, ATP-dependent drug efflux pump which transports a wide variety of structurally and functionally diverse compounds out of cells.[1] P-gp is expressed in epithelial cells of the kidney, liver, pancreas, colon, as well as at the blood–brain barrier,[2] underscoring its role in maintaining concentration gradients of (toxic) compounds at physiological barriers.[3] In addition, it is very often overexpressed in tumor cells and thus is one of the major factors responsible for multiple drug resistance in anticancer therapy. Inhibition of P-gp has therefore been advocated as promising concept to overcome the MDR phenotype. However, although several inhibitors of P-gp have been evaluated in clinical studies, none of them has reached the market so far, which questions the druggability of P-gp.[4]

One of the initial candidates for use as P-gp inhibitor was the calcium channel blocker verapamil. However, clinical studies indicated that the serum concentrations required to reverse MDR lead to severe cardiovascular side effects due to the original biological profile of verapamil. As the cardio-vascular activity is concentrated in the *S*-enantiomer and both enantiomers are equipotent at P-gp, *R*-verapamil was used for further clinical studies. Unfortunately, also this compound failed in clinical phase 3 studies.

Lack of significant stereoselectivity in drug/P-gp interaction was also observed for other compounds, such as niguldipine, nitrendipine, felodipine, carvedilol, propranolol, zosuquidar[5] and propafenone.[5]

However, there are also a few reports of remarkable stereo-specificity.[6] Furthermore, the recently published crystal structure of mouse P-gp co-crystallised with the two enantiomeric cyclopeptides QZ59-*RRR* and QZ59-*SSS* revealed distinct binding sites for the two enantiomers. QZ59-*RRR* binds in the center of the P-gp binding pocket, whereas QZ59-*SSS* binds at two positions: in one position it interacts with hydrophobic residues between TMs 6 and 12, while in the other position it interacts with TMs 8 and 9 and is surrounded by three polar residues. Amino acid residue Val982 plays an important role having close proximity to all three QZ59 sites.[7] Analogous positions of the QZ-isomers were found in docking experiments of the two isomers into a homology model of human P-gp based on the mouse P-gp structure.[8]

In light of our intense structure–activity relationship studies of inhibitors of P-gp, we also synthesized and tested a series of 3-hydroxy-4-amino-dihydrobenzopyranes.[9] These compounds showed biological activities in the low micromolar range, which is comparable to propafenone and verapamil.

In contrast to our main lead compound propafenone, the dihydrobenzopyranes offer the advantage of remarkably reduced conformational flexibility and thus might be versatile molecular tools for probing stereoselective differences of drug/P-gp interaction. Especially annelation of a third ring leading to benzopyrano[3,4-*b*][1,4]oxazines and introduction of large substituents at position 2 of the tricyclic system should lead to compounds with pronounced configurational differences. The compound design is thus based on synthesis of both enantiomers of epoxide **4**, nucleophilic ring opening with *tert*-butyl esters of selected amino acids followed by ester hydrolysis and cyclisation to yield enantiopure target compounds **11–13** (Scheme 1).

Synthesis of the benzopyrane ring system was achieved according to Godfrey *et al.*[10] O-alkylation of 4-hydroxy-benzonitril (**1**) with 3-trifluoroacetyl-3-methyl-but-1-yne

*[a] Department of Medicinal Chemistry, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.
E-mail: gerhard.f.ecker@univie.ac.at; Fax: +43 1-4277-9551;
Tel: +43 1-4277-55110*
[b] Department of Pharmacognosy, Faculty of Pharmacy, Silpakorn University, Nakhon Pathom, Thailand
*[c] Medical University of Vienna, Institute of Medical Chemistry, Weahringerstrasse 10, 1090, Vienna, Austria.
E-mail: peter.chiba@meduniwien.ac.at; Fax: +43 1-4277-60889;
Tel: +43 1-4277-60806*
† This article is part of the 'Enzymes and Proteins' web-theme issue for ChemComm.
‡ Electronic supplementary information (ESI) available: Experimental procedures and spectroscopic data of all compounds biologically tested as well as NMR-spectra of tricycles **11a**–**13b**. See DOI: 10.1039/c0cc03075a
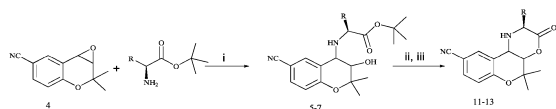
Scheme 1 Synthesis of target compounds **11–13**; (i) 96% ethanol, reflux; (ii) 70% HClO₄; (iii) 4-dimethylaminopyridine, bis-(2-oxo-3-oxazolidinyl)-phosphinic chloride, triethylamine.
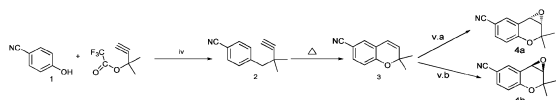
Scheme 2 Synthesis of the benzopyrane ring system and enantiomeric pure (S,S)- and (R,R)-epoxide **4a** and **4b**; (iv) CH₃CN, DBU, CuCl₂, −4 °C, Ar atmosphere; (v.a) (S,S)-Mn(III) Salen NaOCl solution, buffer to pH 11.3, 0 °C; (v.b) (R,R)-Mn(III) Salen NaOCl solution, buffer to pH 11.3, 0 °C.

followed by thermal cyclization gave 6-cyano-2,2-dimethyl-2H-1-benzopyran **3**. Enantioselective epoxidation using a Jacobsons Mn(III) Salen epoxidation catalyst and commercial household bleach (sodium hypochlorite)[11] as the oxygen source gave (S,S)- and (R,R)-epoxide **4a** and **4b** (Scheme 2). Enantiomeric purity of both epoxides was confirmed by HPLC analysis, using a LiChroCART (R,R)-Whelk-01 column (25 × 0.4 cm) and n-hexane/isopropanol (95 : 5) as eluent.

Nucleophilic ring opening of epoxides with L-alanine-, L-valine- and L-phenylalanine-tert-butyl-esters yields optically pure trans-3,4-disubstituted diastereomeric esters **5a,b–7a,b**. These were hydrolysed with 70% HClO₄[12] to yield the corresponding acids **8a,b–10a,b**, which were subsequently cyclised without further purification using bis-(2-oxo-3-oxazoldinyl)-phosphinic chloride, 4-dimethylaminopyridine and triethylamine to yield the target compounds **11a,b–13a,b** (Table 1).

Biological activity of target compounds **11–13** as well as of intermediates **5–7** was assessed using the daunorubicin efflux protocol as described previously (see ESI‡). As negative charges are known to be detrimental for P-gp inhibitory activity the carboxylic acids **8–10** were not measured. In the daunorubicin efflux assay the effect of different modulators on

Table 1 Chemical structure and biological activity of enantiomerically pure benzopyrano[3,4-b][1,4]oxazines

#	R	log P	EC₅₀/μM
5a	CH₃	2.84	29.85
5b	CH₃	2.84	14.55
6a	CH(CH₃)₂	3.82	2.40
6b	CH(CH₃)₂	3.82	2.70
7a	CH₂(C₆H₅)	4.38	0.55
7b	CH₂(C₆H₅)	4.38	0.77
11a	CH₃	1.98	1241.65
11b	CH₃	1.98	76.89
12a	CH(CH₃)₂	2.95	15.32
12b	CH(CH₃)₂	2.95	59.33
13a	CH₂(C₆H₅)	3.51	2.69
13b	CH₂(C₆H₅)	3.51	259.78

the transport rate is measured in a direct functional assay. Furthermore, EC₅₀ values obtained correlate well with those from cytotoxicity assays and rhodamine 123 efflux studies.[13,14] Values are given in Table 1 and are the mean of at least three independently performed experiments. Generally, interexperimental variation was below 20%.

EC₅₀ values cover a range of more than three orders of magnitude with the two phenylalanine esters **7a** and **7b** being the most active compounds (**7a**: 0.55 μM; **7b**: 0.77 μM), followed by the valine analogues **6a** (2.40 μM) and **6b** (2.70 μM). Least active compounds in the series of esters were the alanine derivatives with 29.85 μM (**5a**) and 14.55 μM (**5b**), respectively. It has to be noted that for all three diastereoisomeric pairs almost no differences in biological activity was observed. This pattern changes remarkably upon ring closure to the benzopyrano[3,4-b][1,4]oxazines. Whereas the valine analogues **12a,b** are still within one order of magnitude, both the alanine and phenylalanine derivatives exhibit remarkable differences in their potential to inhibit P-gp. Most strikingly, in the case of alanine the **4aS,10bR**-isomer **11a** is by a factor of 15 less active than the diastereomeric **4aR,10bS** analogue **11b**, whereas in the case of the phenylalanine derivatives this behaviour reverses with the **4aS,10bR**-isomer **13a** being by two orders of magnitude more active than **13b**.

It is widely accepted that access of substrates/inhibitors to the binding cavity of P-gp occurs directly from the membrane bilayer rather than from the aqueous intracellular medium. Thus, QSAR studies very often show a correlation between lipophilicity of the compounds under investigation and their P-gp inhibitory activity.

In this case, differences in activity more likely reflect the ability of the compounds to enter the membrane bilayer rather than differences in their interaction pattern with P-gp. Calculating the log P values of all target compounds with the software package MOE and correlating the values with the log(1/EC₅₀) values exhibit remarkable differences between the (a) and (b) series of compounds. Compounds derived from the (S,S)-epoxide (a series, showing a (3S,4R)-configuration at the benzopyrane ring after aminolysis of the epoxide) showed an excellent correlation ($r^2 = 0.96$, $n = 6$). This indicates that within this series of compounds differences in their P-gp inhibitory potency are mainly due to their capability to permeate into the membrane bilayer rather than to protein–ligand interactions at P-gp.

Compounds from the (b) series yielded a significantly lower r^2 value (0.42). Considering the remarkable drop of activity for the benzyl-derivative **13b** strongly indicates steric constraints for this series of diastereoisomers and thus maybe leading to different binding modes at P-gp. This is further supported by results of docking studies performed on a homology model of human P-gp based on the X-ray structure of mouse P-gp co-crystallised with the cyclic peptide QZ59 (PDB ID: 3G5U; for details see ESI‡).

Both esters **5a,b–7a,b** as well as lactones **11a,b–13a,b** were docked into the homology model of human P-gp. Agglomerative Hierarchical Cluster analysis of the consensus RMSD matrix of esters **5a,b–7a,b** based on the common scaffold identifies 10 clusters. Five clusters contain all compounds of configuration (3S,4R), while four clusters contain all compounds having configuration (3R,4S). However, we

also identified one cluster which contains five (**5a,b**; **6a,b** and **7b**) out of six docked esters. Analysis of the ligand protein interaction profile of eight of these clusters showed mainly interactions with amino acid residues of TM5 and TM6. The dominant interacting amino acids for **5a,b–7a,b** include Tyr307, Tyr310, Phe343, Phe336 and Gln347 (Fig. S1, ESI‡). However, two clusters (one of series (a), one of series (b)) showed different interaction patterns. Compounds of (3*S*,4*R*)-configuration additionally interact with amino acid residues of TM11, including Phe951, Ser952, Cys956 and Met 69, while compounds of (3*R*,4*S*)-configuration showed interaction with TM1, TM2 and TM11, including Tyr117, Ser952, Phe72 and Met69. Using an identical clustering approach for the tricycles **11a,b–13a,b** identified 15 different clusters. Seven of them contain only compounds with (4a*S*,10b*R*)-configuration (**11a–13a**), and are located close to the potential entry pathway (Fig. 1A). Analysis of the protein–ligand interaction pattern showed mainly interactions with TM 4, 5, and 6, in particular with amino acid residues Tyr307, Phe343, Ala342, and Phe303. Eight clusters contain all compounds with (4a*R*,10b*S*)-configuration (**11b–13b**). These clusters are located in two different positions. One position is identical with those of **11a–13a**, the second position is located close to TM 7, 8, 9 and 12, surrounded by amino acid residues Ala985, Ile765 and Leu724 (Fig. 1B). Similar results are obtained when performing the agglomerative hierarchical clustering on the whole set of poses obtained (**5a–13b**; Table S1, ESI‡).

Comparing the main positioning of the benzopyrano-[3,4-*b*][1,4]oxazines with those of QZ59 some overlap could be observed. Especially interaction with Tyr307, Phe343, Phe336, Ala985, Ala342, Met69 and Phe728 was observed for all ligands (Fig. S2, ESI‡). Interestingly, almost all clusters observed are located near TM 4, 5, and 6, which are forming one of the two rings. This is consistent with our recent observation of two pseudosymmetric drug translocation pathways.[15]

Furthermore, it is also interesting to note that compounds of series (a), which show excellent correlation between log *P* values and P-gp inhibitory activity, are predominantly positioned at the potential entry gate, whereas compounds

of the series (b), which show a structure–activity pattern independent of log *P* values, are populating both the entry gate and positions deeper inside the protein. This might provide first insights into the entry path for the ligands.

A closer look of ligand–protein interaction profiles of compounds **13a,b** and **7a,b** identified 4 poses of **13b** showing a steric constraint of the benzyl moiety of **13b**, which is about 2 Å apart from Tyr307 and about 2.5 Å apart from Phe343. All these poses are located at the entry gate. No such steric constraint has been observed for **13a** or for **7a,b**. In the case of **7b** this is most probably due to its conformational flexibility, which allows adopting a conformation to minimize the steric interactions. This indicates that the differences observed for the biological activities of phenylalanine derivatives **13a** and **13b** might be due to steric constraints at the entry path rather than differences in drug/transporter binding. Of course, at the current stage this has to be taken very cautiously, as P-gp undergoes major conformational changes during the transport cycle and docking experiments represent only a single snapshot of this complex movement.

Within this manuscript we present a series of stereoisomers, which, upon rigidisation, show significant differences in their inhibitory potency of the drug efflux pump P-glycoprotein. Ligand docking studies into a homology model of P-gp could provide first evidence for different binding areas of the two diastereomeric compound series. Thus, benzopyrano-[3,4-*b*][1,4]oxazines are versatile tools for exploring the stereoselectivity of drug/P-glycoprotein interaction.

Fig. 1 (A) Shows the three main clusters obtained on the basis of a common scaffold clustering; blue: (4a*S*,10b*R*)-isomers **11a–13a**; green: (4a*R*,10b*S*)-isomers **11b–13b**; brown: **5a–7b**, (B) a docking pose of **13a** (blue) and **13b** (green) near the entry gate showing steric constraints for **13b**, as well as a pose of **13b** deeper inside the membrane (green); brown: a docking pose of the ester **6b** viewed from outside into the TM region.

Notes and references

1 M. M. Gottesman, T. Fojo and S. E. Bates, *Nat. Rev. Cancer*, 2002, **2**, 48–58.
2 M. Kuhnle, M. Egger, C. Muller, A. Mahringer, G. Bernhardt, G. Fricker, B. Konig and A. Buschauer, *J. Med. Chem.*, 2009, **52**, 1190–1197.
3 M. F. Fromm, *Trends Pharmacol. Sci.*, 2004, **25**, 423–429.
4 *Transporters as Drug Carriers*, ed. G. F. Ecker and P. Chiba, Wiley-VCH, Weinheim, 2009.
5 P. Bhatia, M. Kolinski, R. Moaddel, K. Jozwiak and I. W. Wainer, *Xenobiotica*, 2008, **38**, 656–675.
6 S. S. Carey, M. Gleason-Guzman, V. Gokhale and L. H. Hurley, *Mol. Cancer Ther.*, 2008, **7**, 3617–3623.
7 S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch and G. Chang, *Science*, 2009, **323**, 1718–1722.
8 I. K. Pajeva, C. Globisch and M. Wiese, *FEBS J.*, 2009, **276**, 7016–7026.
9 R. Hiessbock, C. Wolf, E. Richter, M. Hitzler, P. Chiba, M. Kratzel and G. Ecker, *J. Med. Chem.*, 1999, **42**, 1921–1926.
10 J. D. Godfrey, R. H. Mueller, T. C. Sedergran, N. Soundararajan and V. J. Colandrea, *Tetrahedron Lett.*, 1994, **35**, 6405–6408.
11 N. H. Lee, A. R. Muci and E. N. Jacobsen, *Tetrahedron Lett.*, 1991, **32**, 5055–5058.
12 F. M. Callahan, G. W. Anderson, R. Paul and J. Zimmerman, *J. Am. Chem. Soc.*, 1963, **85**, 201–207.
13 G. Ecker, M. Huber, D. Schmid and P. Chiba, *Mol. Pharmacol.*, 1999, **56**, 791–796.
14 G. Ecker, P. Chiba and K. J. Schaper, *J. Pharm. Pharmacol.*, 1997, **49**, 305–309.
15 Z. Parveen, C. Bentele, T. Stockner, S. Pferschy, M. Kraupp, M. Freissmuth, G. Ecker and P. Chiba, *Mol Pharmacol.*, 2011, under revision.

75

Exhaustive Sampling of Docking Poses Reveals Binding Hypotheses for Propafenone Type Inhibitors of P-Glycoprotein

Freya Klepsch[1], Peter Chiba[2], Gerhard F. Ecker[1]*

1 Department of Medicinal Chemistry, University of Vienna, Vienna, Austria, **2** Institute of Medical Chemistry, Medical University of Vienna, Vienna, Austria

Abstract

Overexpression of the xenotoxin transporter P-glycoprotein (P-gp) represents one major reason for the development of multidrug resistance (MDR), leading to the failure of antibiotic and cancer therapies. Inhibitors of P-gp have thus been advocated as promising candidates for overcoming the problem of MDR. However, due to lack of a high-resolution structure the concrete mode of interaction of both substrates and inhibitors is still not known. Therefore, structure-based design studies have to rely on protein homology models. In order to identify binding hypotheses for propafenone-type P-gp inhibitors, five different propafenone derivatives with known structure-activity relationship (SAR) pattern were docked into homology models of the apo and the nucleotide-bound conformation of the transporter. To circumvent the uncertainty of scoring functions, we exhaustively sampled the pose space and analyzed the poses by combining information retrieved from SAR studies with common scaffold clustering. The results suggest propafenone binding at the transmembrane helices 5, 6, 7 and 8 in both models, with the amino acid residue Y307 playing a crucial role. The identified binding site in the non-energized state is overlapping with, but not identical to, known binding areas of cyclic P-gp inhibitors and verapamil. These findings support the idea of several small binding sites forming one large binding cavity. Furthermore, the binding hypotheses for both catalytic states were analyzed and showed only small differences in their protein-ligand interaction fingerprints, which indicates only small movements of the ligand during the catalytic cycle.

Introduction

The development of multidrug resistance (MDR) is one major impediment in cancer and antibiotic therapies [1–3]. In 1976 Juliano and Ling were able to associate the occurrence of MDR with the presence of P-glycoprotein (P-gp), the most prominent member of the adenosine triphosphate (ATP) binding cassette (ABC) transporter superfamily [4–6]. ABC proteins are energy dependent transporters with P-gp (ABCB1), multidrug resistance protein 1 (MRP1, ABCC1) and breast cancer resistance protein (BCRP, ABCG2) playing an important role in the protection of cells from harmful xenotoxins. Additionally, ABC proteins are known for modulating the pharmacokinetic profile of drugs and therefore the food and drug administration (FDA) suggested that new drug candidates should be routinely screened for P-gp interaction [7]. In this respect reliable *in silico* methods to characterize P-gp interaction would be of great benefit and help to render the drug discovery process more efficient [8]. However, the polyspecificity of the transporter poses a remarkable challenge concerning this task [9]. A number of ligand based studies have been conducted and provide some insights into the molecular basis of ligand/transporter interaction [10,11]. With the help of biochemical studies like cysteine-cross linking, arginine scanning or photoaffinity labeling, amino acids contributing to binding of selected substrates were identified. On grounds of these experi-ments interaction sites for verapamil, rhodamine (R-site), Hoechst (H-site) and of cyclic peptide P-gp inhibitors (CPPI's) in the transmembrane (TM) domains (TMDs) of P-gp have been postulated [12–16]. Following the ABC transporter topology, P-gp possesses two TMDs, each consisting of 6 TM helices (TMHs), and two nucleotide binding domains (NBDs). While the TMDs are generally responsible for ligand interaction, ATP binding and hydrolysis takes place at the highly conserved nucleotide binding domains (NBDs) [17]. In case of propafenone type ligands photoaffinity labeling studies proposed two symmetrical binding regions at the interfaces of TMHs 5/8 and TMHs 2/11, respectively [18,19]. Nevertheless, due to the small number and the low resolution of crystal structures of ABC-exporters, concrete binding hypotheses remain to be elucidated [20]. The lack of high resolution structures can be explained by the fact that ABC efflux pumps are located in the membrane and that they are rather flexible proteins. As energy dependent transporters they undergo large structural changes during one catalytic cycle, comprising ligand and ATP binding, ligand release and nucleotide hydrolysis [17,21,22]. Up to now the structure of human P-gp could not be resolved, for which reason homology models relying on bacterial homologues had to be utilized. With respect to this the bacterial transporters Sav1866 and MsbA structures, representing different catalytic states of the transport cycle, were generally used as templates [20]. In 2009 the crystal structure of mouse P-gp [12] in

Author Summary

A major reason for the failure of cancer, antibiotic and antiviral therapies is the development of multidrug resistance (MDR). P-glycoprotein (P-gp), an ATP-dependent transport protein located in the membrane of epithelial cells of the kidney, liver, pancreas, colon and the blood-brain barrier, has been linked to the export of a broad variety of xenotoxins. Overexpression of P-gp leads to extrusion of therapeutic drugs and therefore triggers MDR. Thus, identification of potential P-gp inhibitors represents a promising concept for treatment of multiresistant tumours. However, due to lack of high resolution structural information and the polyspecific ligand recognition pattern only very limited information is available on the molecular basis of ligand/transporter interaction. Within this study we characterized the propafenone binding site of P-gp by docking a set of derivatives with known SAR into homology models of P-gp which represent both the apo and the nucleotide-bound state. Poses retrieved are in accordance with results from previous photoaffinity labeling studies and thus pave the way for structure-based in silico screening approaches.

complex with a cyclic tetrapeptide was resolved, thus representing a ligand binding competent conformation of the protein. With 88% sequence identity it is well suited for homology modeling of the human homologue and thus paves the way for structure-based approaches.

The present study aimed at elucidating the binding mode of propafenone type inhibitors of P-gp using a combined homology modeling/docking approach. Propafenones show a clear structure-activity relationship (SAR) pattern [11] and thus represent versatile tool compounds to pursue this task. The wealth of ligand-based information available allows judging the reliability of docking poses on basis of the SAR pattern rather than by use of energetic terms derived from scoring functions. The selected compounds were the piperidine analogue GPV005, the analogous des-hydroxy derivative GPV186, the arylpiperazine GPV019, the hydroxyphenylpiperidine GPV062, and the benzoylamide GPV366 (Figure 1). All compounds bear a carbonyl group, which has been shown to be important for high P-gp inhibitory activity [23].

There are numerous studies showing that there is a basic underlying correlation between P-gp inhibitory activity and lipophilicity of the compounds. This accounts for several compound classes and has also been shown for propafenone analogues.

However, propafenones which bear a 4-hydroxy-4-phenylpiperidine moiety are generally by a factor of 10 more active than equi-lipophilic derivatives without the hydroxy-group in 4-position of the piperidine moiety (Figure S1) [24]. This points at a distinct additional interaction mediated by the 4-hydroxy group, most probably in form of a H-bond. This distinct SAR pattern in combination with the recently described common scaffold clustering [25,26] was used to guide the prioritization of docking poses.

Results

Homology Modeling

In March 2009 Aller et al. published the crystal structure of mouse P-gp in the absence of a ligand (PDB ID: 3G5U) [12] and in complex with stereoisomeric CPPI's (PDB IDs: 3G60, 3G61) [12]. These structures represent the ligand binding competent

state and were therefore the first choices for investigating drug/P-gp binding.

As the structural difference between the apo protein and the co-crystallized structures was surprisingly low (0.61 Å of Cα atoms) the higher resolved 3G5U structure was utilized as homology modeling template (3G5U_Pgp). With the modeling program MODELLER 100 different homology models were created and refined. All models were assessed with the geometry check tool implemented in MOE, which was used as a selection criterion for the final model. As additional measure for model quality the GA341 method was used, which relies on sequence identity, compactness and the combined statistical z-score. All models obtained the highest possible GA341 value of 1. Furthermore, the final model was analyzed with the structure assessment program PROCHECK [27]. The Ramachandran plot showed that 84.6% of the residues lie in most favored, 12.5% in additionally allowed, 2.1% in generously allowed and 0.8% in disallowed regions. The 2.9% of residues in generously allowed or disallowed regions are located in the nucleotide binding domains (NBD) or extracellular loops (ECL) and are therefore not involved in drug binding (Figure S2). The QMEAN analysis [28] (Figure S3) showed that residues lining the binding pocket are of satisfactory quality.

In order to cover different catalytic states of the protein, a second homology model was generated on basis of the bacterial transporter Sav1866 in the nucleotide-bound state (PDB code: 2HYD) [29] (2HYD_Pgp). This crystal structure is the highest resolution ABC exporter structure and has therefore been frequently used as modeling template [20]. 100 different models were generated and refined with MODELLER, of which all obtained a GA341 score of 1. The final model was selected on basis of the geometry check function in MOE. The Ramachandran plot statistics provided by the evaluation tool PROCHECK showed that 92% of all residues lie in most favored regions, while 6.5% were found in additionally allowed, 0.2% in generously allowed and only 0.6% in disallowed regions (Figure S2). Most of the 0.8% residues that are located in generously allowed or disallowed regions can be found in the NDB. Although residue Y116 lies within the TMDs and could therefore be involved in drug binding, this residue is oriented outside the cavity. A Ramachandran analysis performed by MolProbity and MOE detected no outliers in the TM region. Furthermore, this model shows also good quality in the binding site region according to QMEAN analysis (Figure S3).

Docking

For the docking process five different propafenone derivatives were selected according to their differences in lipophilic efficiency and fit quality [30], and were docked into both homology models.

With the genetic algorithm based docking program GOLD [31] 100 poses for each of the five ligands were generated. To determine the ASN, GLN and HIS flips the web application MolProbity was utilized [32]. In order to avoid any bias, the binding site was defined as the complete TM region. According to the binding site assessment tool implemented in the software suite Schrödinger (SiteMap), this region in 3G5U_Pg mainly shows highly hydrophobic characteristics, which prompted us to dock the ligands in their non-ionized state. This is also supported by previous findings of ligand-based QSAR studies which indicated that the nitrogen atom not necessarily interacts in its charged form. However, since there is evidence that the protein's pore is water filled, the ligands were also docked in their ionized state [33]. This is also in accordance with recently published data which show that mutation of two glutamine residues at the entry path of

77
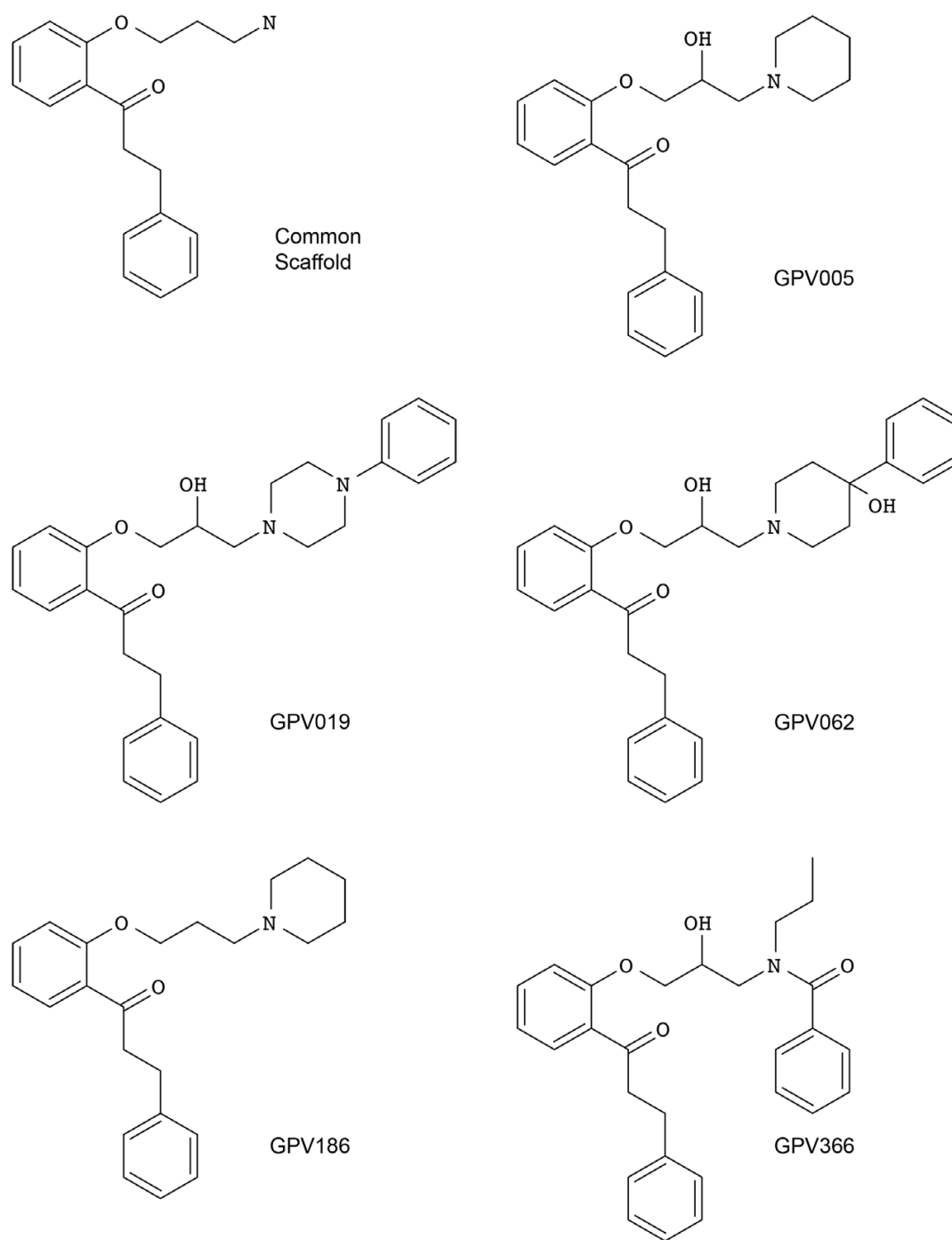
Figure 1. Ligand structures and codes that were used in this study. The common scaffold represents the largest common substructure and was used for root mean square deviation (RMSD) clustering.
doi:10.1371/journal.pcbi.1002036.g001

78

the transporter to positively charged arginines affected the inhibitory activity of an positively ionizable propafenone analog, whereas the activity of GPV366 remained unmodulated [34].

The resulting poses in both conformations were distributed largely within the TM region of P-gp (Figure 2), showing interactions with protein residues of multiple TM helices, located throughout the binding region. The calculation of protein ligand interaction fingerprints (PLIF) with MOE showed that in case of 3G5U_Pgp residues primarily located on TM helices 1, 5, 6, 7, 8, 11 and 12 were involved in binding (Figure 3). According to this tool, residues involved either show direct interactions with docking poses or are located within 4.5 Å distance to the ligand.

The unprocessed complexes were energetically minimized using LigX, a minimization tool implemented in MOE for further evaluation.

The minimized poses were clustered according to the root-mean-square deviation (RMSD) of the heavy atoms of the common scaffold (Figure 1) [35]. To follow the idea of a common binding mode only those clusters were kept that comprehend at least four out of the five compounds used (common scaffold clusters, CSCs). Clustering the poses of the docking run with 3G5U_Pgp resulted in 114 clusters, which were subsequently reduced to 12 CSC. As can be seen in Figures 2a and b some clusters protrude into the central cavity, but most of the CSCs are found in the vicinity of helices 5 and 8 (called the 5/8 interface). Previous photo-affinity labeling experiments suggested this region to be in involved in propafenone binding [36]. The position of the CSCs close to the 5/8 interface was also reflected in the PLIF pattern, as the involvement of residues L304 and Y307 located in TM helix 5, F343 of TM helix 6, L724 and I731 in TM helix 7, A761 in TM helix 8 and V981 in TM helix 12 was increased (Figure 3).

In case of 2HYD_Pgp, the RMSD clustering process resulted in 78 clusters, which were reduced to nine common scaffold clusters, containing 264 poses (Table 1). As can be seen in Figure 2c and d, also docking into the nucleotide-bound homology model results in CSCs that tend to accumulate closer to the 5/8 interface and thus in vicinity of the photo-affinity labeled residues (Figure S4). The clustering process did not change the general PLIF pattern. TM helices 5, 6, 7 and 8 show similar contributions before and after scaffold clustering, but more frequently interactions were observed with individual residues, like Y307 (TM helix 5), Y310 (TM helix 5), L724 (TM helix 7) and T769 (TM helix 8) (Figure 4).

The model based on the murine 3G5U structure represents the binding competent state, whereas the model based on the nucleotide-bound 2HYD structure likely represents the off-state of P-gp ligands [37]. Since propafenones might show different affinities towards these two structures, final pose evaluation was carried out in different ways.

In the hit-to-lead decision process as well as in lead optimization different efficiency metrics are applied to prioritise lead candidates. Briefly, in case of equi-potent compounds these parameters select for the smaller, more hydrophilic ones. As high lipophilicity correlates with promiscuity, poor solubility and poor metabolic clearance [38], candidates with high lipophilic efficiency (LLE = log(potency) - logP) are preferred. Ligand based studies clearly demonstrate a correlation between lipophilicity of P-gp inhibitors and their biological activity. However, as P-gp is extracting its ligands directly out of the membrane bilayer, this is most probably a consequence of concentration in the membrane rather than of direct protein interaction. Calculating the LLE normalizes for this effect and aids in identifying ligands with increased activity as a result of direct interaction with the protein rather than higher biomembrane distribution. The 4-hydroxy-4-phenylpiperidine

GPV062 shows by far the highest LLE (Table 2) suggesting that in contrast to the other ligands, the higher activity of GPV062 is not due to a high logP value. While LLE normalizes for the lipophilic bias in potency description, LE simply corrects for the size of a molecule by dividing the activity of a compound by its heavy atom count. This approach is extensively used in fragment based drug design to select those fragments, which are worth being further investigated. As Reynolds et al. [30] concluded that LE generally is biased towards smaller molecules, the normalized size-independent fit quality (FQ) was assessed. Both, LE and FQ, clearly highlight the hydroxyphenylpiperidine GPV062 as being the most efficient compound (Table 2). The explanation for the increased LLE and FQ values seems to be the 4-hydroxy-group of GPV062. As this group clearly reduces the lipophilicity of a molecule, the increase in activity was interpreted as a result of hydrogen bonding. Thus, those CSCs were prioritized in which GPV062 is able to form a hydrogen bond between the hydroxyl-group of the 4-hydroxy-4-phenyl moiety and the protein.

With 3G5U_Pgp only one quarter of all twelve common scaffold clusters showed a hydrogen bond between the hydroxyl-group of GPV062 and the protein (Table 1) (GPV062-OH Clusters). These three clusters (CSCs I, II, III) are located very close to each other at the 5/8 interface (Figure 5a), with an increased number of interactions formed by residues L304, Y310, L724, A761 and V981. Furthermore, the PLIF pattern showed that interactions with TM helices 1 and 11 are no longer present.

The positions of CSCs I and III are very similar, since both are forming a hydrogen bond with Y310 and a π/π-interaction with F336. In CSC II, on the contrary, a hydrogen bond interaction with A761 was observed.

For further evaluation of the poses a pharmacophore search was performed, utilizing a model published by Langer et al. that based on a set of propafenone type P-gp inhibitors [39]. Only those two clusters that formed a hydrogen bond with Y310 matched this pharmacophore query. As depicted in Figure 5b, both clusters perfectly fit the photolabeling pattern observed in this half of the protein.

Evaluation of the docking results with 2HYD_Pgp could not be based on ligand affinity data, since this structure represents the nucleotide-bound off-state and therefore is considered as the low-affinity state for substrates. This rules out prioritization on basis of SAR-information. All common scaffold clusters of 2HYD_Pgp are in close vicinity of the 3G5U_Pgp GPV062-OH poses (Figure 6).

Discussion

Homology Modeling

The homology models generated in this study resemble two different states of P-gp: the open-inward or apo state and the open-outward or nucleotide-bound state. Since the publication of mouse P-gp in the absence (PDB ID: 3G5U) or in complex with ligands (PDB Ids: 3G60, 3G61) only a few homology models of the human homologue were published on the basis of these structures. Pajeva et al. presented two homology models that were based on the structure of 3G61, chain A, which is complexed with QZ59-SSS [40,41]. The advantage of selecting this template for homology modeling is the presence of the complexed ligands. On the other hand, 3G5U is resolved at higher resolution (3.80 Å) and shows only minor differences in the binding site (RMSD of all atoms of QZ59-SSS surrounding residues: 1.251 Å). However, the still relatively low resolution of the template certainly needs to be taken into account when it is used for docking experiments.

The open-outward model relied on the structure of the bacterial homologue Sav1866 (PDB ID: 2HYD), which possesses the same

79
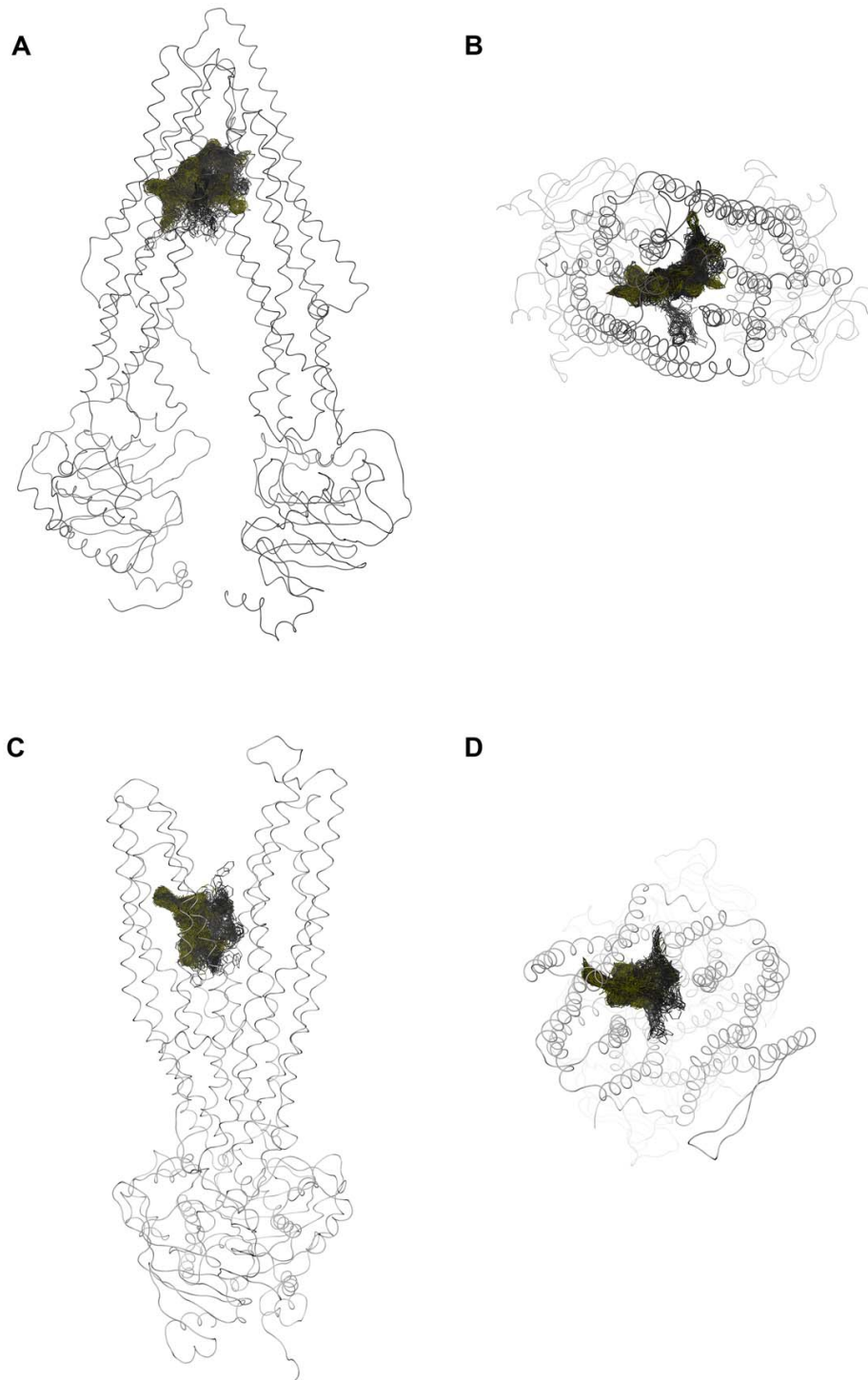
Figure 2. Distribution of all 500 poses in 3G5U_Pgp (A, B) and 2HYD_Pgp (C, D). Yellow: common scaffold cluster, grey: residual poses.
doi:10.1371/journal.pcbi.1002036.g002

80
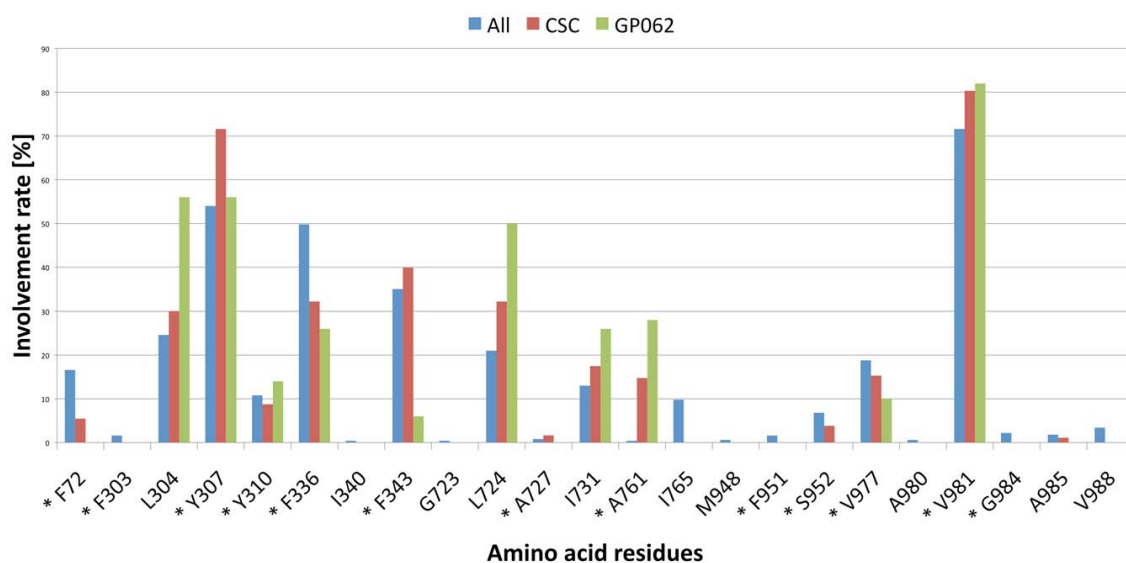
Figure 3. Protein ligand interaction fingerprint (PLIF) of the docking poses in 3G5U_Pgp, calculated with MOE. All: 500 poses after docking, CSC: common scaffold cluster, GPV062: cluster that showed an interaction between the OH-group of GPV062 and the protein. Residues marked with an asterisk show direct interaction with docking poses.
doi:10.1371/journal.pcbi.1002036.g003

domain architecture as P-gp [42] and therefore frequently served as modeling template. With a resolution of 3.0 Å it represents one of the best resolved full ABC transporters. The relevance of this nucleotide-bound structure is widely accepted, as experimental studies showed close association of the NBDs [43,44]. In contrast, the structures of mouse P-gp disagree with kinetic and FRET studies that report no complete dissociation of the NBDs [37,45]. In addition, a recent cross-linking study further strengthened this by showing that an M1M cross-link between L175C and N820C did not prevent verapamil and rhodamine B to be transported [46]. However, as P-gp is known to be highly flexible and to undergo large conformational changes during the catalytic cycle, the existence of a state with dissociated NBDs cannot be ruled out entirely. Additional evidence was presented by Sauna et al., who demonstrated that ATP binding reduces the affinity for propafenone analogues [37]. Finally, the fact that the mouse P-gp structure (3G5U) has been cocrystallized with two ligands strongly indicates that this structure represents a ligand-binding competent state of the protein. Thus it was considered as a versatile template for modeling the high-affinity state of the protein for subsequent docking studies.

Docking

Although ligand docking is a commonly used tool for the identification of ligand-protein interactions, in case of P-gp it bears a lot of challenges: (i) P-gp possesses a large binding cavity that consists of several binding sites, (ii) is highly flexible, and (iii) is probably able to harbor more than one ligand simultaneously [47,48]. Finally, there is no high resolution structure of human P-gp available, which requires to work with protein homology models. Considering the low resolution of the templates, this adds additional layers of uncertainty. Thus, results from ligand docking runs have to be interpreted very carefully. In an attempt to combat all these uncertainties we applied an exhaustive docking protocol avoiding to a maximum possible extent the use of scoring functions and including all the knowledge present from SAR and QSAR studies.

In docking experiments, the definition of the binding site is a key parameter of the docking protocol. As only little information is available about binding of propafenones into P-gp, the whole TM region was selected as a potential interaction region. In order to avoid any bias introduced by scoring functions, a large amount of docking poses was generated. While placement algorithms of

Table 1. Cluster statistics of docking runs into different catalytic states.

	3G5U_Pgp (non-ionized)	3G5U_Pgp (ionized)	2HYD_Pgp (non-ionized)	2HYD_Pgp (ionized)
Total number of poses	500	500	500	500
Number of clusters after RMSD clustering (3 Å)	114	111	78	77
Number of common scaffold clusters (CSCs)	12 (184 poses)	11 (195 poses)	9 (264 poses)	7 (240 poses)
CSCs with interaction between GPV062-OH and protein	I, II, III	IV, V, VI	–	–
ph4-matching clusters	I, III	IV, V, VI	–	–

doi:10.1371/journal.pcbi.1002036.t001

81
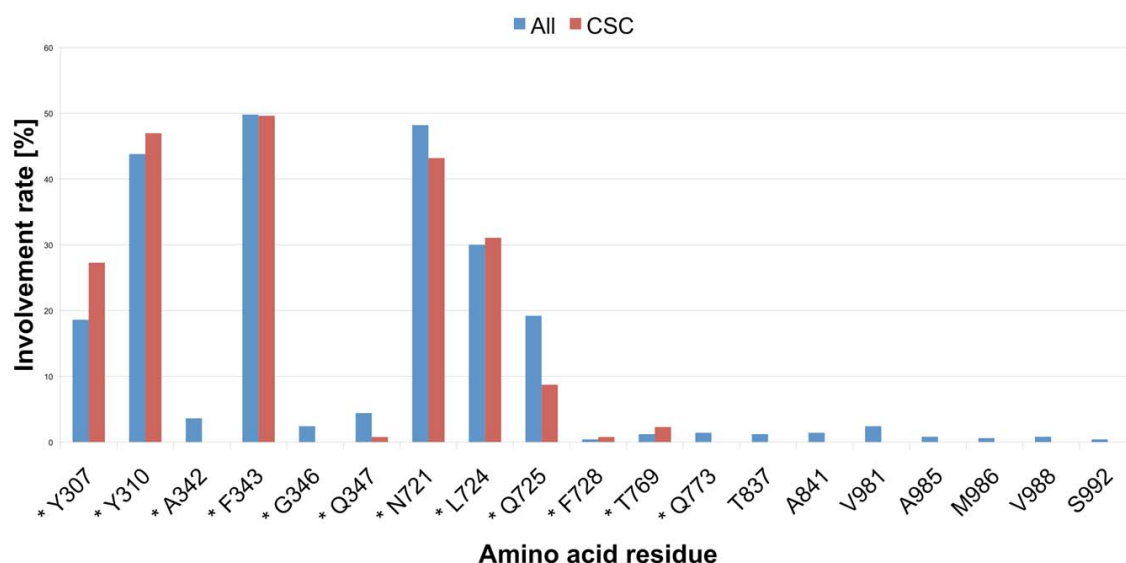
Figure 4. Protein ligand interaction fingerprint (PLIF) of the docking poses in 2HYD_Pgp, calculated with MOE. All: 500 poses after docking, CSC: common scaffold cluster. Residues marked with an asterisk show direct interaction with the docking poses.
doi:10.1371/journal.pcbi.1002036.g004

docking programs are most of the time able to find the native pose of a ligand in the binding pocket, the correct estimation of the binding energy leading to a correct ranking of the poses is still unsatisfying. To overcome this uncertainty of scoring functions, we recently implemented experimental data guided docking/scoring. In this approach prioritization of docking poses is performed on basis of mutagenesis data, biochemical data, and/or information from ligand based studies [25,26].

The interaction of propafenones with P-gp follows a clear structure-activity relationship pattern (for reviews see [11]). Based on these results and on calculation of lipophilic efficiency (LLE) and fit quality (FQ) we selected a small set of analogs for docking and subsequent common scaffold clustering. Both LLE and FQ as well as previously performed Hansch analysis stressed the importance of the hydroxyl-group of GPV062 for high activity. Clustering of all poses according to their common scaffold (Figure 1) combined with pose selection based on H-bonding interactions of the OH group allowed a considerable reduction of docking poses.

Although docking experiments have their limitations depending on the validity of the target structure, the results of docking into

Table 2. Activities of docked ligands.

Ligand	pIC50	HAC	ClogP	LLE	LE	FQ
GPV005	6.22	27	4.38	1.84	0.23	0.77
GPV019	6.21	33	5.15	1.06	0.19	0.75
GPV062	7.24	34	4.15	3.09	0.21	0.87
GPV186	6.19	26	5.54	0.65	0.24	0.77
GPV366	5.78	33	4.94	0.84	0.18	0.70

HAC = heavy atom count, LLE = lipophilic ligand efficiency, LE = ligand efficiency, FQ = fit quality.
doi:10.1371/journal.pcbi.1002036.t002

3G5U_Pgp are very consistent. As shown in Figure 5 the three final clusters are located in close vicinity. Especially CSCs I and III are very similar, showing strong H-bonding interactions with Y310 and thus supporting the importance of the hydroxyl group of GPV062. Both clusters also match the pharmacophore model of Langer et. al [39]. Due to previously performed ligand based studies also the importance of the carbonyl group of the propafenone scaffold became evident [49]. Although initial poses show no interaction with the carbonyl group, these become apparent after processing of data with the rotamer explorer implemented in MOE. When rotating amino acid residue Y307 towards the carbonyl group, an interaction can be generated (Figure 7). In a dynamic system H-bond formation thus might be observed. Interestingly, for CSC III a rotation of Y307 did not result in an interaction with the carbonyl group, most probably due to a small offset of the carbonyl group towards the cell interior. However, this assumption would need further investigations, since discussing possible interactions on atomistic detail has to be done with caution when working with a homology model, especially if the resolution is quite low. Nevertheless, the relevance of Y307 in ligand binding was also shown with cocrystallized CPPI's, where the R-stereoisomer forms an interaction with this residue [12]. Furthermore, this residue is in close vicinity to I306, which was shown to lead to permanent activation of ATPase activity when mutated to cysteine and covalently linked with the thiol-reactive drug substrate verapamil [15].

CSC II forms a weak H-bond between the hydroxyl-group of GPV062 and the backbone of A761. With respect to the ligand interaction tool in MOE the strength of this bond is only 1/10 compared to that in CSCs I and III. Applying the rotamer explorer results in either formation of a stronger hydrogen bond with the OH-group of GPV062 or formation of a new interaction with the carbonyl group (with these interactions not being coexistent). Finally, with respect to residues photoaffinity labelled by benzophenone analogous propafenones, CSCs I and III show a better match (Figure 5b), because the photoreactive carbonyl group is closer to the PAL region than in CSC I.

Figure 5. GPV062-OH interaction clusters in the binding pocket of 3G5U_Pgp. CSC I (green), CSC II (yellow), CSC III (cyan). A) Top view; the three interacting amino acids are colored according to their cluster-membership. B) side view; the blue surface indicates residues that are involved in propafenone binding, determined by photoaffinity labeling [14].
doi:10.1371/journal.pcbi.1002036.g005
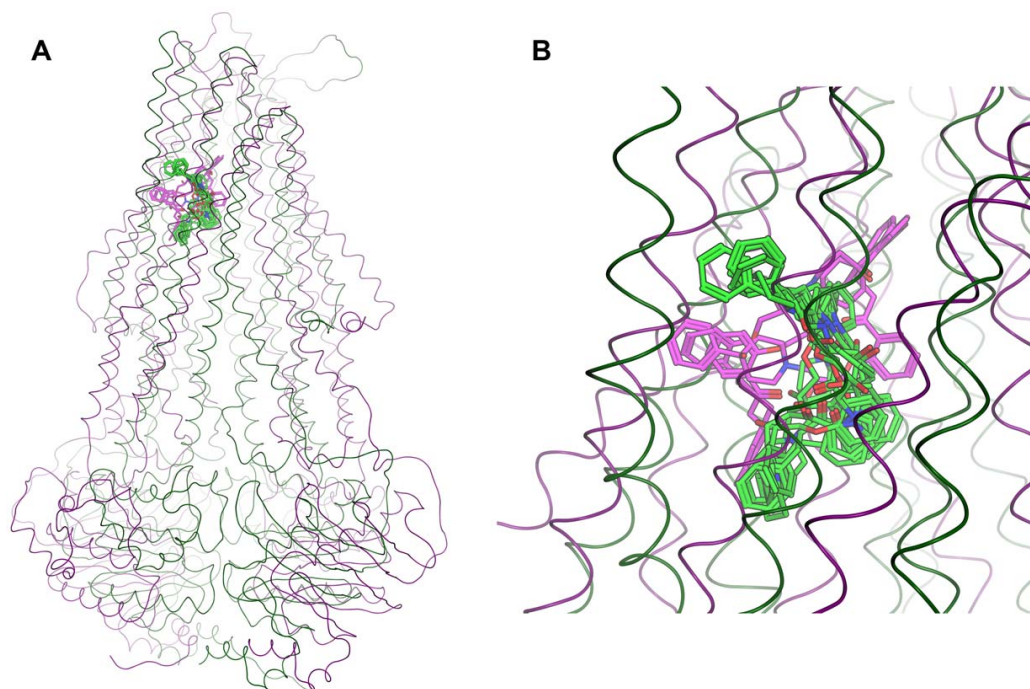
Figure 6. Comparison of docking poses in different stages of the catalytic cycle. Magenta: GPV062-OH clusters of docking into 3G5U_Pgp (high affinity), green: common scaffold clusters of docking into 2HYD_Pgp (low affinity). A) overview, B) close-up view.
doi:10.1371/journal.pcbi.1002036.g006
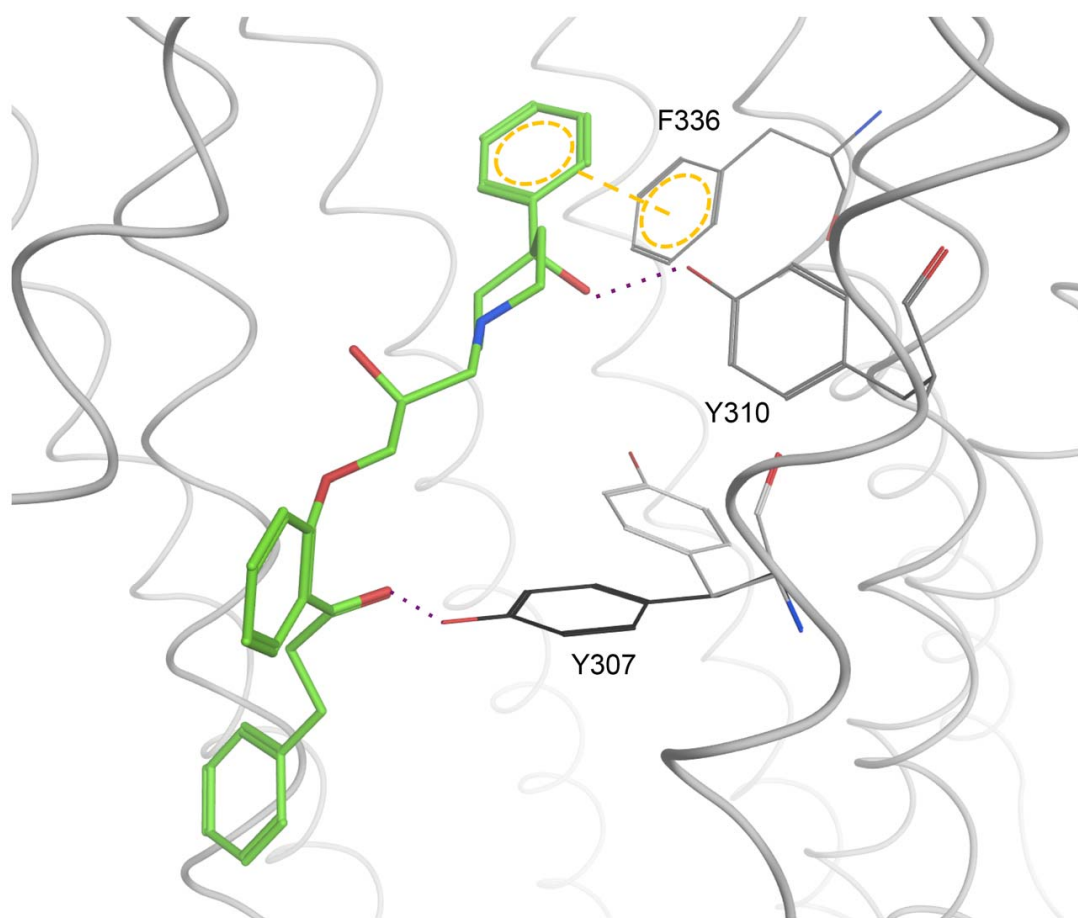
Figure 7. Interactions of CSC I with 3G5U_Pgp. By rotating the residue Y307 (grey:original, black: rotated) a new hydrogen bond between Y307 and the carbonyl group of the ligand was formed.
doi:10.1371/journal.pcbi.1002036.g007

In consideration of these findings the pose of CSC I was preferred over the other two clusters.

It is also known that binding of propafenones to P-gp meets steric constraints in the vicinity of the nitrogen atom, because diphenyl moieties in this position lead to a log order decrease in activity [49]. In all three clusters the introduction of a diphenyl substituted nitrogen results in steric clashes and subsequent minimization of the binding pocket leads to the loss of H-bond interactions.

Docking into 3G5U_Pgp with ionized ligands resulted in three different CSCs that show an interaction between the OH_group of GPV062 and the protein. While one is located very central in the pore (CSC IV) forming an H-bond between GP062-OH and A727, the other two (CSC V and VI) exactly match CSC I of the docking with neutral ligands. For the latter an H-bond between the hydroxyl-group and Y310 could be observed.

As can be seen in Figure 2, the different CSCs of 2HYD_Pgp are located in the same binding site at the 5/8 interface. Regarding their different orientation within this region, docking poses can be separated into two distinct groups. Docking poses belonging to group 1 (CSCs a, b, c and d) frequently form interactions between the carbonyl group and Y307. Furthermore,

H-bond interactions between the piperidine nitrogen or the hydroxyl-group and Y310 can be observed. This interaction pattern is similar to the one of CSCs I and III of the docking run performed with 3G5U_Pgp. Individual GPV062 poses show additional H-bond interactions between the 4-hydroxy-group and Y310, another frequently observed interaction in CSCs I and III. According to these observations the transformation of CSCs I and III in the apo state into CSCs of group 1 of the nucleotide-bound state seems possible.

In contrast, group 1 and group 2 are in an up-side-down orientation when compared to each other. In this case the carbonyl group is located near Y310 and thus closer to the extracellular portion of the protein. The nitrogen atom, as well as the hydroxyl group, is oriented towards Y307 and N721, which was also observed for CSC II of the 3G5U_Pgp docking run. Therefore, group 2, comprising clusters e, f and g, corresponds to the nucleotide-bound conformations of CSC II of the apo-conformation.

CSCs h and i cannot be clearly assigned to one of these groups and have to be regarded separately. The nitrogen atom of CSC h shows a similar location as the N of group 2, however, due to a shift of the central phenyl ring downwards, H-bond interactions

84

between the carbonyl oxygen and Y307 and the OH-group and N721 can be formed simultaneously.

CSC i shares its carbonyl group orientation with group 2, but the central phenyl ring lies in a perpendicular direction, which results in interactions between the ligand nitrogen and hydroxyl group with Q725.

Considering the docking run to 2HYD_Pgp with ionized ligands, group 1 could be clearly reproduced. Three out of seven CSCs form those characteristic H-bond interactions between the carbonyl oxygen and Y307 and the hydroxyl group and Y310. In contrast to the unprotonated ligands, the nitrogen atom and Y310 form a pi/cation interaction and occur at higher frequency. Overall the clusters belonging to group 1 show high homogeneity and strong interactions. In contrast to this the poses of each of the four other clusters share no consistent pattern and therefore the common binding was only reflected in geometrically similar positioning.

Interestingly, although the experimental data suggest two symmetrical binding sites, no common scaffold cluster and hardly any poses could be found at the second photoaffinity labeled site at the 2/11 interface. One possible explanation might be the asymmetry of the template crystal structure 3G5U. The region consisting of TM helices 4, 5, 7, 8, 9 and 12 in case of 3G5U_P-gp, and TM helices 3, 4, 5, 6, 7 and 8 in case of 2HYD_P-gp, in both cases showed larger sites when using the SiteFinder tool in MOE than their counterparts around the 2/11 interface. This demonstrates the limitations of docking experiments relying on one crystal structure that represents only a snapshot of a flexible protein. Thus, to rule out the possibility that every docked ligand will end up at the 5/8 interface just because of this asymmetry, a docking run with rhodamine 123 was conducted. In this case 21 of 39 clusters were found in vicinity of residues I340, L975 and V981, which are located on TM helices 6 and 12 and known to be involved in rhodamine binding [13].
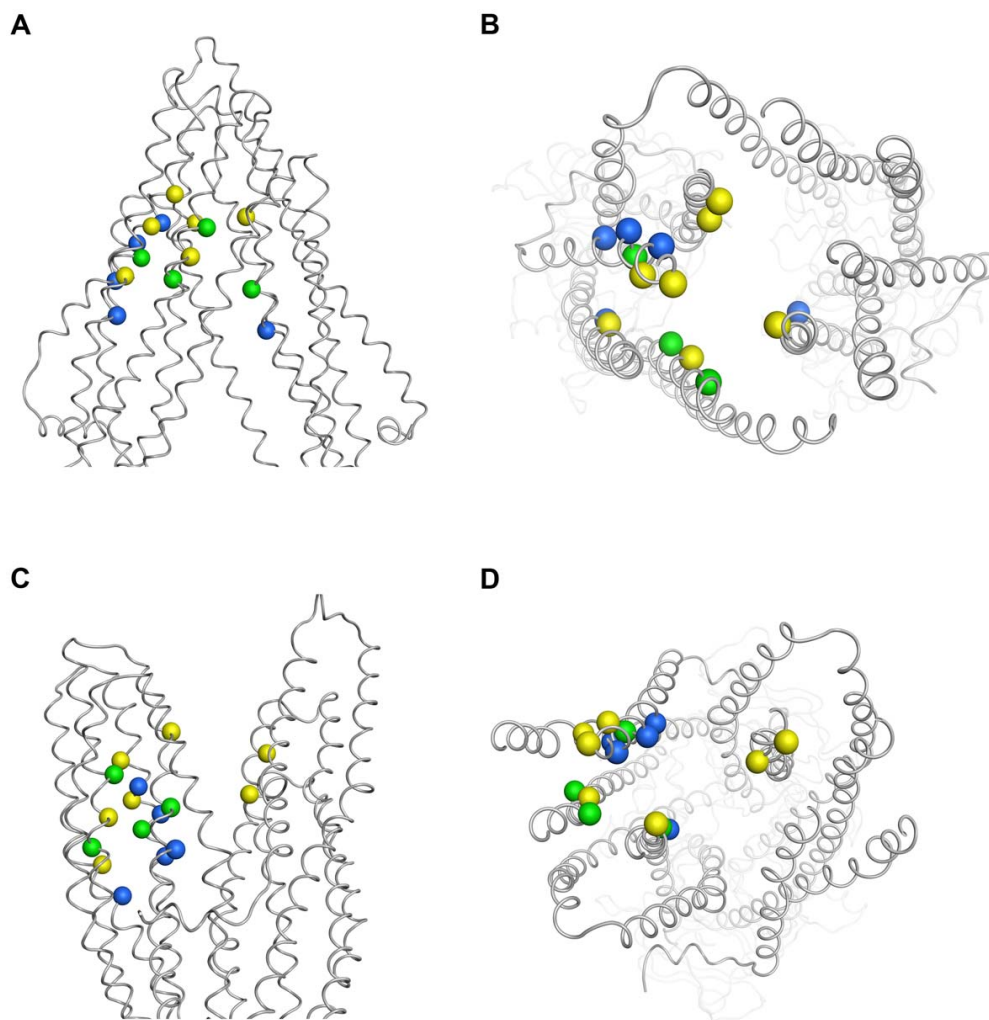
Figure 8. Comparison of main interacting residues. The spheres represent Cα-atoms of interacting residues of 3G5U_Pgp (panels A, B) and 2HYD_Pgp (panels C, D). Blue spheres: 2HYD_Pgp, green: 2HYD_Pgp and 3G5U_Pgp, yellow: 3G5U_Pgp. A) and b) 3G5U_Pgp in front and top view; c) and d) 2HYD_Pgp in front and top and view.
doi:10.1371/journal.pcbi.1002036.g008

Comparison of Open-Closed Binding Regions

In order to gain first insights into the potential ligand translocation pathways, the compounds were docked in two different catalytic states of P-gp. Interestingly, the docking results show similar interaction patterns. In both models, ligand poses are found in close vicinity (4,5 Å) of residues Y307 and Y310 of TM helix 5, F343 of TM helix 6 and L724 of TM helix 7, which suggests involvement of both TM domains in drug binding. This is in accordance with Loo et al., who showed that both TM domains are essential for drug translocation [50].

In Figure 8 the interacting amino acid residues of both docking approaches are depicted. In the 3G5U_Pgp structure the interactions seem to be very similar, concentrating on the 5/8 interface. Due to the conformational change and the resulting movement of TM helix 12, interactions between propafenones and V977 and V981 are lost. Top views of the models indicate that the corresponding interacting residues (3G5U_Pgp: yellow, 2HYD_Pgp: blue, both: green) face the central pore. It seems that the conformational change associated with nucleotide binding moves previously buried residues towards the binding pocket and therefore allows them to form new interactions with the ligands.

In Figure 9 a Venn diagram compares residues in binding sites of CPPIs and verapamil with that of propafenones. As TM helices 5, 6 and 7 are lining the central cavity in the murine P-gp structure, a considerable overlap of residues interacting with propafenones and that shown to interact with cocrystallized CPPIs can be found in this region. One residue of TM helix 7, F728, is suggested to interact with all four drugs and therefore plays a crucial role in ligand binding. This is in agreement with the finding of Loo et al. that TM helix 7 is part of the drug binding site [51]. Loo et al. also demonstrated that binding of vinblastine, cyclosporin A and rhodamine B could prevent the formation of a cross-link between L339C and F728C, suggesting that the ligands are at least partially located between these two residues [52]. This is also the case for the three docking clusters in 3G5U_Pgp, which are presented in this study.

Furthermore, the diagram is consistent with the notion that P-gp possesses a large binding cavity, which harbors different partially overlapping drug binding sites for different ligands [39,40]. In the cocrystallized structures 3G60 and 3G61 the cyclopeptides are located at the interface of the two TMDs, which explains the high overlap between these ligands and verapamil or propafenones, respectively.

Ligand docking into polyspecific antitargets such as the hERG potassium channel and the drug transporter P-glycoprotein requires thorough validation of the poses obtained. In this paper we describe the application of an SAR-guided docking protocol, which for the first time retrieves a binding hypothesis for propafenone-type inhibitors of P-gp. Although performing docking studies with homology models always bears a lot of risks the results are in agreement with experimental studies, which strengthens the applicability of the complex docking protocol we used for this study. This could pave the way for structure-based ligand design approaches.

Methods

Homology Modeling

Two homology models based on the bacterial homologue Sav1866 (PDB ID: 2HYD, resolution: 3.0 Å [29]) and murine P-gp (PDB ID: 3G5U, chain A, resolution: 3.8 Å [12]) were built. Both models were generated with the program MODELLER 9v7 using the automodel protocol [53]. In case of 3G5U_Pgp the alignment proposed by Aller et al. [8] was used (Figure S5). To correct the disruption in TM helix 12 of 3G5U a secondary structure constraint between residues 885 and 928 was applied. For 2HYD_Pgp the alignment was done according to Stockner et al. [54] (Figure S6). The linker region between the TM domains was modeled. Out of the 100 generated models those with the smallest number of outliers according to the geometry check function in MOE were selected for docking.

Docking

For the docking study five propafenone derivatives were selected on basis of known SAR and differences in LLE and FQ. LLE was calculated by subtracting ClogP from experimentally determined IC_{50} values and FQ was calculated as outlined in [30]. To examine the quality of the ClogP calculation, the values were compared with previously published experimentally defined logP data of propafenone analogs [23]. A correlation of $r = 0.92$ could be identified.

Minimization and protonation of the ligands was performed with MOE.

For the correct determination of ASN/GLN/HIS flips the web application MolProbity was utilized [32]. The docking process was performed using the Gold Suite 1.2.1 [31]. Hydrogens were added and the binding site was defined as the entire TM region of the homology model. All side chains were kept rigid and the ligand
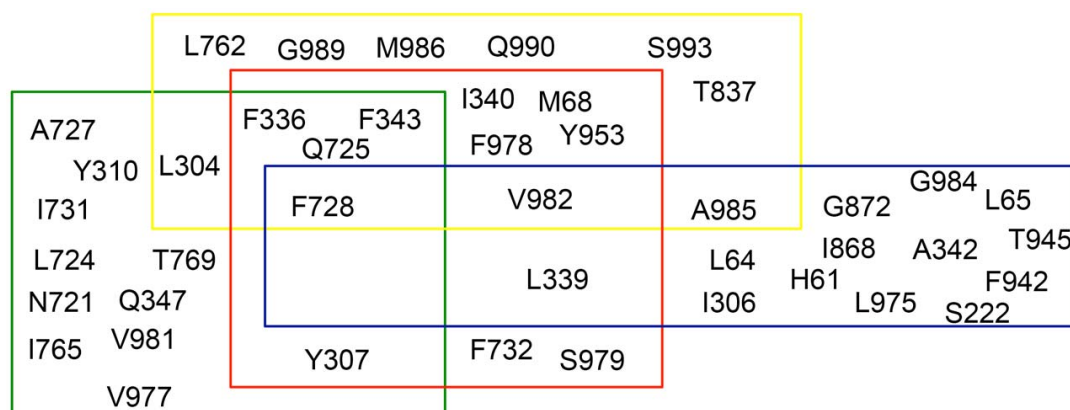
Figure 9. Venn diagram of drug binding sites in human P-gp.
doi:10.1371/journal.pcbi.1002036.g009

86

was treated flexible by performing 100 genetic algorithm runs per molecule. The implemented Gold scoring function GoldScore was used for evaluation of the complexes. The final poses and the surrounding protein amino acid residues were minimized using LigX implemented in the MOE software package. Rescoring was performed with the empirical scoring function XSCORE.

Cluster Analysis

On basis of the common scaffold an RMSD matrix of all five ligands was generated and used for clustering. The dissimilarity matrix was clustered with the program R [55], using complete linkage as clustering algorithm and a clustering height of 3 Å. Only those clusters were kept that inherited at least four out of the five ligands docked.

In case of 3G5U_Pgp those clusters were selected for final assessment that were able to form a hydrogen bond between the OH-group of GPV062 and the protein, detected by the ligand interaction tool of MOE.

Supporting Information

Figure S1 ClogP-pIC50 correlation of propafenone analogs. The ligands used for docking are highlighted. [24] (TIF)

Figure S2 Outliers defined by PROCHECK analysis. A) 3G5U_Pgp, B) 2HYD_Pgp. Grey: generously allowed residues, black: disallowed residues. (TIFF)

Figure S3 QMEAN analysis of the homology models generated with MODELLER. A) 3G5U_Pgp, B) 2HYD_Pgp. Blue: high quality regions, red: low quality regions. (TIFF)

Figure S4 Common scaffold clusters after docking into 2HYD_Pgp. The blue surface indicates residues that are involved in propafenone binding, determined by photoaffinity labeling [14]. (TIFF)

Figure S5 Sequence alignment used for the generation of the homology model 3G5U_Pgp. The sequences of human P-gp and of the X-ray structure of mouse P-gp have been aligned as suggested by Aller et al. [12]. (PDF)

Figure S6 Sequence alignment used for the generation of the homology model of 2HYD_Pgp. The sequences of human P-gp and the bacterial ABC-exporter SAV1866 have been aligned as suggested by Stockner et al. [54]. (PDF)

Author Contributions

References

1. Dean M, Fojo T, Bates S (2005) Tumour stem cells and drug resistance. Nat Rev Cancer 5: 275–284.
2. Globisch C, Pajeva IK, Wiese M (2008) Identification of putative binding sites of P-glycoprotein based on its homology model. ChemMedChem 3: 280–295.
3. Ramachandra M, Ambudkar SV, Gottesman MM, Pastan I, Hrycyna CA (1996) Functional characterization of a glycine 185-to-valine substitution in human P-glycoprotein by using a vaccinia-based transient expression system. Mol Biol Cell 7: 1485–1498.
4. Juliano R (1976) Drug-resistant mutants of Chinese hamster ovary cells possess an altered cell surface carbohydrate component. J Supramol Struct 4: 521–526.
5. Hrycyna CA (2001) Molecular genetic analysis and biochemical characterization of mammalian P-glycoproteins involved in multidrug resistance. Semin Cell Dev Biol 12: 247–256.
6. Ford RM, Kamis AB, Kerr ID, Callaghan R (2009) The ABC Transporters: Strucural Insights into Drug Transport. In: Ecker G, Chiba P, eds. Transporters as Drug Carriers. Weinheim: Wiley-VCH.
7. Food and Drug Administration (FDA) (2006) Drug Interaction Studies - Study Design, Data Analysis, and Implications for Dosing and Labeling. Available: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072101.pdf. Accessed 22 December 2010.
8. Schneider G Virtual screening: an endless staircase? Nat Rev Drug Discov 9: 273–276.
9. Klepsch F, Stockner T, Erker T, Muller M, Chiba P, et al. (2010) Using structural and mechanistic information to design novel inhibitors/substrates of P-glycoprotein. Curr Top Med Chem 10: 1769–1774.
10. Raub TJ (2006) P-glycoprotein recognition of substrates and circumvention through rational drug design. Mol Pharm 3: 3–25.
11. Pleban K, Ecker GF (2005) Inhibitors of p-glycoprotein–lead identification and optimisation. Mini Rev Med Chem 5: 153–163.
12. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722.
13. Loo TW, Clarke DM (2002) Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein. J Biol Chem 277: 44332–44338.
14. Loo TW, Clarke DM (1997) Identification of residues in the drug-binding site of human P-glycoprotein using a thiol-reactive substrate. J Biol Chem 272: 31945–31948.
15. Loo TW, Clarke DM (2008) Mutational analysis of ABC proteins. Arch Biochem Biophys 476: 51–64.
16. Qu Q, Sharom FJ (2002) Proximity of bound Hoechst 33342 to the ATPase catalytic sites places the drug binding site of P-glycoprotein within the cytoplasmic membrane leaflet. Biochemistry 41: 4744–4752.
17. Linton KJ, Higgins CF (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. Pflugers Arch 453: 555–567.
18. Pleban K, Kopp S, Csaszar E, Peer M, Hrebicek T, et al. (2005) P-glycoprotein substrate binding domains are located at the transmembrane domain/transmembrane domain interfaces: a combined photoaffinity labeling-protein homology modeling approach. Mol Pharmacol 67: 365–374.
19. Chiba P, Mihalek I, Ecker GF, Kopp S, Lichtarge O (2006) Role of transmembrane domain/transmembrane domain interfaces of P-glycoprotein (ABCB1) in solute transport. Convergent information from photoaffinity labeling, site directed mutagenesis and in silico importance prediction. Curr Med Chem 13: 793–805.
20. Klepsch F, Ecker G (2010) Impact of the Recent Mouse P-Glycoprotein Structure for Structure-Based Ligand Design. Mol Inf 29: 276–286.
21. Seeger MA, van Veen HW (2009) Molecular basis of multidrug transport by ABC transporters. Biochim Biophys Acta 1794: 725–737.
22. Callaghan R, Ford RC, Kerr ID (2006) The translocation mechanism of P-glycoprotein. FEBS Lett 580: 1056–1063.
23. Chiba P, Ecker G, Schmid D, Drach J, Tell B, et al. (1996) Structural requirements for activity of propafenone-type modulators in P-glycoprotein-mediated multidrug resistance. Mol Pharmacol 49: 1122–1130.
24. Chiba P, Hitzler M, Richter E, Huber M, Tmej C, et al. (1997) Studies on Propafenone-type Modulators of Multidrug Resistance III: Variations on the Nitrogen. Quant Struct-Act Relat 16: 361–366.
25. Sarker S, Weissensteiner R, Steiner I, Sitte HH, Ecker GF, et al. (2010) The high-affinity binding site for tricyclic antidepressants resides in the outer vestibule of the serotonin transporter. Mol Pharmacol 78: 1026–1035.
26. Richter L, Ernst M, Sieghart W, Ecker G (2010) Identification of binding modes of benzodiazepeine binding site ligands by a combined docking-pharmacophore modeling approach. Drugs Future 35: 173.
27. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291.
28. Benkert P, Kunzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res 37: W510–514.
29. Dawson RJ, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180–185.
30. Reynolds CH, Bembenek SD, Tounge BA (2007) The role of molecular size in ligand efficiency. Bioorg Med Chem Lett 17: 4258–4261.
31. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52: 609–623.
32. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21.
33. Gutmann DA, Ward A, Urbatsch IL, Chang G, van Veen HW (2010) Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. Trends Biochem Sci 35: 36–42.

87

34. Parveen Z, Stockner T, Bentele C, Pferschy S, Kraupp M, et al. (2011) Molecular Dissection of Dual Pseudosymmetric Solute Translocation Pathways in Human P-Glycoprotein. Mol Pharmacol 79: 443–452.
35. Chema D, Eren D, Yayon A, Goldblum A, Zaliani A (2004) Identifying the binding mode of a molecular scaffold. J Comput Aided Mol Des 18: 23–40.
36. Ecker GF, Csaszar E, Kopp S, Plagens B, Holzer W, et al. (2002) Identification of ligand-binding regions of P-glycoprotein by activated-pharmacophore photoaffinity labeling and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. Mol Pharmacol 61: 637–648.
37. Sauna ZE, Kim IW, Nandigama K, Kopp S, Chiba P, et al. (2007) Catalytic cycle of ATP hydrolysis by P-glycoprotein: evidence for formation of the E.S reaction intermediate with ATP-gamma-S, a nonhydrolyzable analogue of ATP. Biochemistry 46: 13787–13799.
38. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 6: 881–890.
39. Langer T, Eder M, Hoffmann RD, Chiba P, Ecker GF (2004) Lead identification for modulators of multidrug resistance based on in silico screening with a pharmacophoric feature model. Arch Pharm (Weinheim) 337: 317–327.
40. Pajeva IK, Globisch C, Wiese M (2009) Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. Chem Med Chem 4: 1883–1896.
41. Pajeva IK, Globisch C, Wiese M (2009) Comparison of the inward- and outward-open homology models and ligand binding of human P-glycoprotein. FEBS J 276: 7016–7026.
42. Zolnerciks JK, Wooding C, Linton KJ (2007) Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein. FASEB J 21: 3937–3948.
43. Lee JY, Urbatsch IL, Senior AE, Wilkens S (2008) Nucleotide-induced structural changes in P-glycoprotein observed by electron microscopy. J Biol Chem 283: 5769–5779.
44. Loo TW, Bartlett MC, Clarke DM (2002) The "LSGGQ" motif in each nucleotide-binding domain of human P-glycoprotein is adjacent to the opposing walker A sequence. J Biol Chem 277: 41303–41306.
45. Qu Q, Sharom FJ (2001) FRET analysis indicates that the two ATPase active sites of the P-glycoprotein multidrug transporter are closely associated. Biochemistry 40: 1413–1422.
46. Loo TW, Bartlett MC, Clarke DM (2010) Human P-glycoprotein is active when the two halves are clamped together in the closed conformation. Biochem Biophys Res Commun 395: 436–440.
47. Loo TW, Bartlett MC, Clarke DM (2003) Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein. J Biol Chem 278: 39706–39710.
48. Lugo MR, Sharom FJ (2005) Interaction of LDS-751 and rhodamine 123 with P-glycoprotein: evidence for simultaneous binding of both drugs. Biochemistry 44: 14020–14029.
49. Ecker G, Chiba P (2009) QSAR Studies on ABC Transporter - How to Deal with Polypspecificity. In: Ecker G, Chiba P, eds. Transporters as Drug Carriers. Weinheim: Wiley-VCH. pp 121–137.
50. Loo TW, Clarke DM (1998) Superfolding of the partially unfolded core-glycosylated intermediate of human P-glycoprotein into the mature enzyme is promoted by substrate-induced transmembrane domain interactions. J Biol Chem 273: 14671–14674.
51. Loo TW, Bartlett MC, Clarke DM (2006) Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket. Biochem J 399: 351–359.
52. Loo TW, Bartlett MC, Clarke DM (2007) Suppressor mutations in the transmembrane segments of P-glycoprotein promote maturation of processing mutants and disrupt a subset of drug-binding sites. J Biol Chem 282: 32043–32052.
53. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815.
54. Stockner T, de Vries SJ, Bonvin AM, Ecker GF, Chiba P (2009) Data-driven homology modelling of P-glycoprotein in the ATP-bound state indicates flexibility of the transmembrane domains. FEBS J 276: 964–972.
55. R Development Core Team (2010) R: A Language and Environment for Statistical Computing. ViennaAustria: R Foundation for Statistical Computing, Available: http://www.R-project.org.
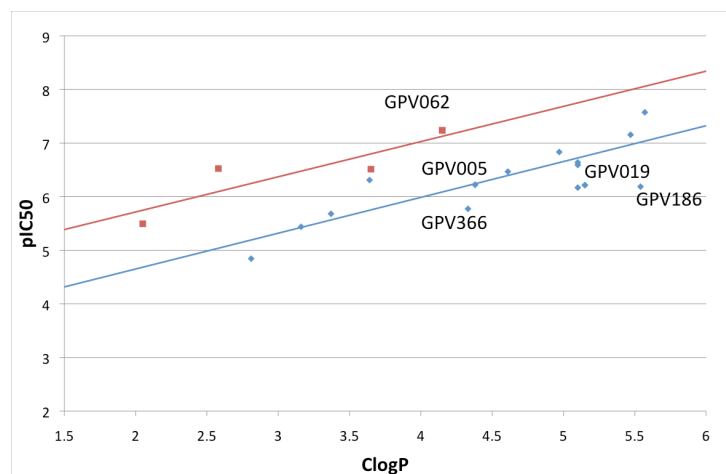
88

Supporting Information

Figure S1: ClogP-pIC50 correlation of propafenone analogs. The ligands used for docking are highlighted.
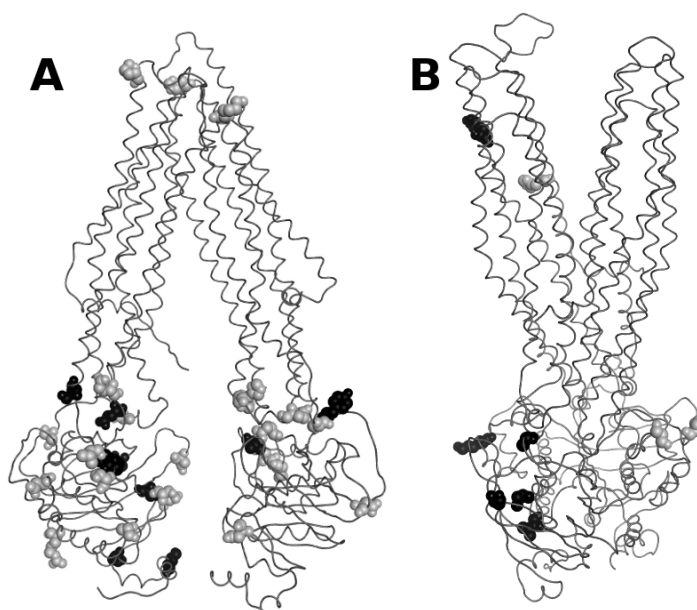
Figure S2: Outliers defined by PROCHECK analysis. A) 3G5U_Pgp, B) 2HYD_Pgp. Grey: generously allowed residues, black: disallowed residues.
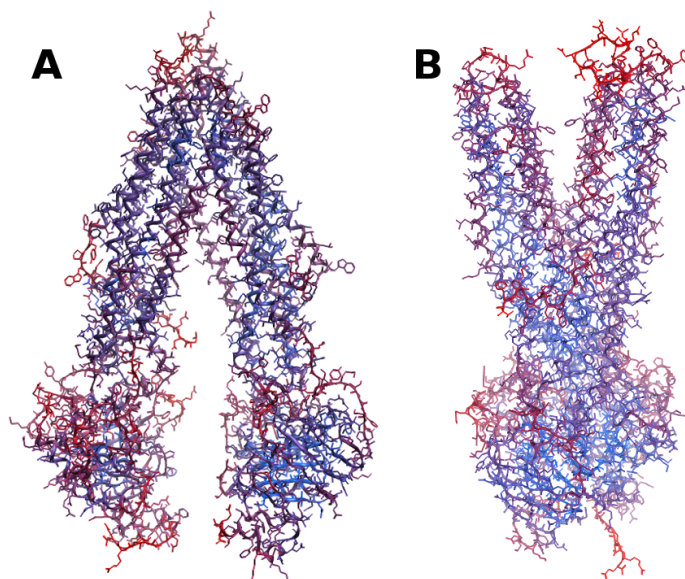
Figure S3: QMEAN analysis of the homology models generated with MOD-ELLER. A) 3G5U_Pgp, B) 2HYD_Pgp. Blue: high quality regions, red: low quality regions.
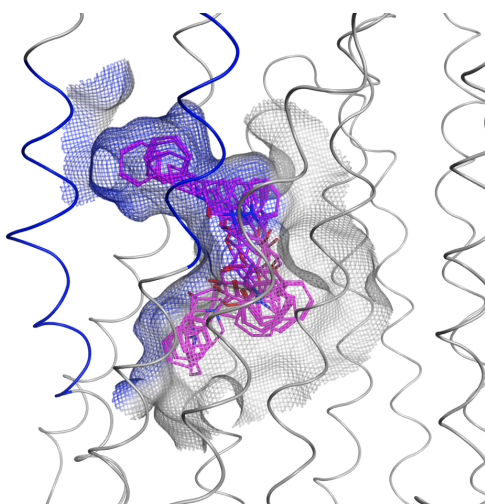
Figure 3.1: Figure S4: Common scaffold clusters after docking into 2HYD_Pgp. The blue surface indicates residues that are involved in propafenone binding, determined by photoaffinity labeling.

```
Alignment
Template: PDB code 3G5U (Chain A)
Target: human P-gp

Reference: Aller SG, Yu J, Ward A, Wenig Y, Chittaboina S, et al. (2009)
Structure of P-glycoprotein reveals a molecular basis for poly-specific drug
binding. Science 323: 1718-1722


3G5U_A.pdb    --------------------------------VSVLTMFRYAGWLDRLYMLVGTLAAII 27
human_P-gp    MDLEGDRNGGAKKKNFFKLNNKSEKDKKEKKPTVSVFSMFRYSNWLDKLYMVVGTLAAII 60
                                              ***::****:.***:***:*******


3G5U_A.pdb    HGVALPLMMLIFGDMTDSFASVGNVS---KNSTNMSEADKRAMFAKLEEEMTTYAYYYTG 84
human_P-gp    HGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRSDINDTGFFMNLEEDMTRYAYYYSG 120
              **..******:**:*** **..**:.    .* ** *: :. ..:* :***:** *****.*


3G5U_A.pdb    IGAGVLIVAYIQVSFWCLAAGRQIHKIRQKFFHAIMNQEIGWFDVHDVGELNTRLTDDVS 144
human_P-gp    IGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHDVGELNTRLTDDVS 180
              ******:.*.*******************::****** .*********************


3G5U_A.pdb    KINEGIGDKIGMFFQAMATFFGGFIIGFTRGWKLTLVILAISPVLGLSAGIWAKILSSFT 204
human_P-gp    KINEGIGDKIGMFFQSMATFFTGFIVGFTRGWKLTLVILAISPVLGLSAAVWAKILSSFT 240
              ***************:***** ****:************************.:*********


3G5U_A.pdb    DKELHAYAKAGAVAEEVLAAIRTVIAFGGQKKELERYNNNLEEAKRLGIKKAITANISMG 264
human_P-gp    DKELLAYAKAGAVAEEVLAAIRTVIAFGGQKKELERYNKNLEEAKRIGIKKAITANISIG 300
              **** ******************************:*******:************:*


3G5U_A.pdb    AAFLLIYASYALAFWYGTSLVISKEYSIGQVLTVFFSVLIGAFSVGQASPNIEAFANARG 324
human_P-gp    AAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARG 360
              ******************:**:* *********************:*********


3G5U_A.pdb    AAYEVFKIIDNKPSIDSFSKSGHKPDNIQGNLEFKNIHFSYPSRKEVQILKGLNLKVKSG 384
human_P-gp    AAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSG 420
              ****:************ *:***********:*****:*:*:***********.*********.**


3G5U_A.pdb    QTVALVGNSGCGKSTTVQLMQRLYDPLDGMVSIDGQDIRTINVRYLREIIGVVSQEPVLF 444
human_P-gp    QTVALVGNSGCGKSTTVQLMQRLYDPTEGMVSVDGQDIRTINVRFLREIIGVVSQEPVLF 480
              ************************** :****:***********:***************


3G5U_A.pdb    ATTIAENIRYGREDVTMDEIEKAVKEANAYDFIMKLPHQFDTLVGERGAQLSGGQKQRIA 504
human_P-gp    ATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKLPHKFDTLVGERGAQLSGGQKQRIA 540
              ************:*************************:***********************


3G5U_A.pdb    IARALVRNPKILLLDEATSALDTESEAVVQAALDKAREGRTTIVIAHRLSTVRNADVIAG 564
human_P-gp    IARALVRNPKILLLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAG 600
              *****************************.*****:********************


3G5U_A.pdb    FDGGVIVEQGNHDELMREKGIYFKLVMTQT----------------------------- 594
human_P-gp    FDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDSRS 660
              **.*****:*******:********* **


3G5U_A.pdb    --------------------------LDEDVPPASFWRILKLNSTEWPYFVVGIFCAII 627
human_P-gp    SLIRKRSTRRSVRGSQAQDRKLSTKEALDESIPPVSFWRIMKLNLTEWPYFVVGVFCAII 720
                                        ***.:**.*****:*** **********:*****


3G5U_A.pdb    NGGLQPAFSVIFSKVVGVFTNGGPPETQRQNSNLFSLLFLILGIISFITFFLQGFTFGKA 687
human_P-gp    NGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLLFLALGIISFITFFLQGFTFGKA 780
```

91

```
                    ********::****::****.  .  ***:*********** *****************

3G5U_A.pdb    GEILTKRLRYMVFKSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGATGSRLAVIFQNI 747
human_P-gp    GEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVITQNI 840
                    ***************:***************************************** ******* ***

3G5U_A.pdb    ANLGTGIIISLIYGWQLTLLLLAIVPIIAIAGVVEMKMLSGQALKDKKELEGSGKIATEA 807
human_P-gp    ANLGTGIIISFIYGWQLTLLLLAIVPIIAIAGVVEMKMLSGQALKDKKELEGAGKIATEA 900
                    **********:*************************************************:*******

3G5U_A.pdb    IENFRTVVSLTREQKFETMYAQSLQIPYRNAMKKAHVFGITFSFTQAMMYFSYAACFRFG 867
human_P-gp    IENFRTVVSLTQEQKFEHMYAQSLQVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFG 960
                    ***********:***** *******:****::***:*****************.*****

3G5U_A.pdb    AYLVTQQLMTFENVLLVFSAIVFGAMAVGQVSSFAPDYAKATVSASHIIRIIEKTPEIDS 927
human_P-gp    AYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDS 1020
                    ****::**:**:*********:****************** .:**:*** ****** ***

3G5U_A.pdb    YSTQGLKPNMLEGNVQFSGVVFNYPTRPSIPVLQGLSLEVKKGQTLALVGSSGCGKSTVV 987
human_P-gp    YSTEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVV 1080
                    ***:** ** ***** *. ********* *******************************

3G5U_A.pdb    QLLERFYDPMAGSVFLDGKEIKQLNVQWLRAQLGIVSQEPILFDCSIAENIAYGDNSRVV 1047
human_P-gp    QLLERFYDPLAGKVLLDGKEIKRLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNSRVV 1140
                    *********:**.*:*********:*******:***************************

3G5U_A.pdb    SYEEIVRAAKEANIHQFIDSLPDKYNTRVGDKGTQLSGGQKQRIAIARALVRQPHILLLD 1107
human_P-gp    SQEEIVRAAKEANIHAFIESLPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLD 1200
                    * *************  **:***:**.*:*********************************

3G5U_A.pdb    EATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVIQNGKVKEHGTHQQL 1167
human_P-gp    EATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQL 1260
                    *********************************************.***.**********

3G5U_A.pdb    LAQKGIYFSMVSVQA----- 1182
human_P-gp    LAQKGIYFSMVSVQAGTKRQ 1280
                    **************
```

```
Alignment
Template: PDB code 2HYD
Target: human P-gp

Reference: Stockner T, de Vries SJ, Bonvin AM, Ecker GF, Chiba P (2009)
Data-driven homology modelling of P-glycoprotine in the ATP-bound state
indicates fleibility of the transmembrane domains. FEBS J 276: 964-972

2HYD.pdb      --MIKRYLQFVKPYKYRIFATIIVGIIK-FGIPMLIPLLIKYAIDGVINNHA-------- 49
human_P-gp    VFSMFRYSNW--LDKLYMVVGTLAAIIHGAGLPLMMLVFGEMTDIFANAGNLEDLMSNIT 58
                : ** ::    *  :..  :..**:  *:*:::: :: : :   .  .:

2HYD.pdb      -------LTTDEKVHHLTIAIGIALFIFVIVRPPIEFIRQYLAQWTSNKILYDIRKKLYN 102
human_P-gp    NRSDINDTGFFMNLEEDMTRYAYYYSGIGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFH 118
                     ::..     .       :       :*:    ::.: ::.***:::

2HYD.pdb      HLQALSARFYA--NNQVGQVISRVINDVEQTKDFILTGLMNIWLDCITIIIALSIMFFLD 160
human_P-gp    AIMRQEIGWFD--VHDVGELNTRLTDDVSKINEGIGDKIGMFFQSMATFFTGFIVGFTRG 176
               :   .  ::     ::**:: :*: :**.: ::  *   :  :: .  *: .: : *  .

2HYD.pdb      VKLTLAALFIFPFYILTVYVFFGRLRKLTRERSQALAEVQGFLHERVQGISVVKSFAIED 220
human_P-gp    WKLTLVILAISPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAEEVLAAIRTVIAFGGQK 236
               ****. * * *. *:. *:   * .:* :. * *:. .. .* : .* .* :*. :.

2HYD.pdb      NEAKNFDKKNTNFLTRALKHTRWNAYSFAAINTVTDIGPIIVIGVGAYLAISGSITVGTL 280
human_P-gp    KELERYNKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGEYSIGQV 296
               :* :.::*: :     .:*::       *:.*     :   .  :. *: *.:**. ::* :

2HYD.pdb      AAFVGYLELLFGPLRRLVASFTTLTQSFASMDRVFQLIDEDYDIKNG-VGAQPIEIKQGR 339
human_P-gp    LTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGN 356
               :.. . : :   .: :  .*: ::::: .: .:*::**:. .*.. ..: : :*.

2HYD.pdb      IDIDHVSFQYNDN-EAPILKDINLSIEKGETVAFVGMSGGGKSTLINLIPRFYDVTSGQI 398
human_P-gp    LEFRNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQRLYDPTEGMV 416
               :::  :* *.* .. *. ***.:**.::.*:***:** ** **** ::*. *:** *.* :

2HYD.pdb      LIDGHNIKDFLTGSLRNQIGLVQQDNILFSDTVKENILLGRP--TATDEEVVEAAKMANA 456
human_P-gp    SVDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRE--NVTMDEIEKAVKEANA 474
                :**::*: : .  **: **:*.*: :**: *: *** **    ..* :*: :*.* ***

2HYD.pdb      HDFIMNLPQGYDTEVGERGVKLSGGQKQRLSIARIFLNNPPILILDEATSALDLESESII 516
human_P-gp    YDFIMKLPHKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEAVV 534
               :****:**: :** *****.:*********:.*** :** ::.** **:********* ***:::

2HYD.pdb      QEALDVLSKDRTTLIVAHRLSTITHADKIVVIENGHIVETGTHRELIAKQGAYEHLYSIQ 576
human_P-gp    QVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQ 594
               * ***   *.***:::*******: :** *. :::* ***.*.** ::* * :* * :* ::*

2HYD.pdb      NLMIKRYLQFVKPYKYRIFATIIVGIIK-FGIPMLIPLLIKYAIDGVINNHALTTDEKVH 635
human_P-gp    TA--WRIMKLNLTEWPYFVVVGVFCAIINGGLQPAFAIIFSKIIGVFTRID----DPETKR 648
               .    * :::  .    :.. :: .**:    * :  :: *    :        *. :

2HYD.pdb      HLTIAIGIALFIFVIVRPPIEFIRQYLAQWTSNKILYDIRKKLYNHLQALSARFYA--NN 693
human_P-gp    QNSNLFSLLFLALGIISFITFFLQGFTFGKAGEILTKRLRYMVFRSMLRQDVSWFDDPKN 708
               : :  :.: :: : *:     *:: .      .: :   :* ::. : ..  :.    :*

2HYD.pdb      QVGQVISRVINDVEQTKDFILTGLMNIWLDCITIIIALSIMFFLDVKLTLAALFIFPFYI 753
human_P-gp    TTGALTTRLANDAAQVKGAIGSRLAVITQNIANLGTGIIISFIYGWQLTLLLLAIVPIIA 768
```

93

```
                   .* : :*: **. *.*. * : *   *   :   .:  .: * *: . :*** * *.*:
2HYD.pdb     LTVYVFFGRLRKLTRERSQALAEVQGFLHERVQGISVVKSFAIEDNEAKNFDKKNTNFLT 813
human_P-gp   IAGVVEMKMLSGQALKDKKELEGAGKIATEAIENFRTVVSLTQEQKFEHMYAQSLQVPYR 828
                   ::  * : *    : : .: *   . :   * ::.: .* *:: *:: : : :.

2HYD.pdb     RALKHTRWNAYSFAAINTVTDIGPIIVIGVGAYLAISGSITVGTLAAFVGYLELLFGPLR 873
human_P-gp   NSLRKAHIFGITFSFTQAMMYFSYAGCFRFGAYLVAHKLMSFEDVLLVFSAVVFGAMAVG 888
                   .:*:::: . :*: :::  :.    : .****.    ::.  :  ...::   .:

2HYD.pdb     RLVASFTTLTQSFASMDRVFQLIDEDYDIKNG-VGAQPIEIKQGRIDIDHVSFQYNDN-E 931
human_P-gp   QVSSFAPDYAKAKISAAHIIMIIEKTPLIDSYSTEGLMPNTLEGNVTFGEVVFNYPTRPD 948
                   :: :  .  ::: *  ::: :*::   *..  .  . :   :*.: :..* *:* . :

2HYD.pdb     APILKDINLSIEKGETVAFVGMSGGGKSTLINLIPRFYDVTSGQILIDGHNIKDFLTGSL 991
human_P-gp   IPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGKVLLDGKEIKRLNVQWL 1008
                    *:*:...*.::**:*:*:** ** ****::*: **** .*:.:*.**:.** : .  *

2HYD.pdb     RNQIGLVQQDNILFSDTVKENILLGRP--TATDEEVVEAAKMANAHDFIMNLPQGYDTEV 1049
human_P-gp   RAHLGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVRAAKEANIHAFIESLPNKYSTKV 1068
                   * ::*:*.*: ***. :: *** *    ..:**:*.*** ** * ** .**: *.*:*

2HYD.pdb     GERGVKLSGGQKQRLSIARIFLNNPPILILDEATSALDLESESIIQEALDVLSKDRTTLI 1109
human_P-gp   GDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKAREGRTCIV 1128
                   *:*.:********:*** :::.* **:********* ***.::****** :.** ::

2HYD.pdb     VAHRLSTITHADKIVVIENGHIVETGTHRELIAKQGAYEHLYSIQNL---- 1156
human_P-gp   IAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSVQAGTKRQ 1179
                   :******* :** ***::**:: * ***::*:*::* *  : *:*
```

94

3.3 Post-docking analyses

The previous section showed, how information-based docking can lead to binding hypotheses that are consistent with SAR and experimental data. Finally, two binding hypotheses (binding mode I and II) could be identified that fulfill these criteria. However, binding mode I turned out to be the preferred one, as it could theoretically not only explain the higher affinity for propafenone derivatives bearing a hydroxyphenylpiperidine moiety by forming a hydrogen bond to TYR310, but also the importance of the carbonyl group at the phenylpropanone part could be explained by forming a hydrogen bond with TYR307.

Although binding mode II could not explain the importance of the carbonyl group it might be risky to choose between two very probable binding modes on the basis of docking results. Since the protein is only represented by a snapshot, interactions between ligand and transporter, as well as the complete orientation of the ligand might change in a dynamic system. In that sense, molecular dynamics simulations were performed in order to analyze the stability of the two selected binding modes in the protein binding site. In addition, corresponding binding modes in the posthydrolytic state have been simulated to investigate the affinity-decrease of the ligand due to conformational change of the protein.

Furthermore, the more stable binding mode should be able to undergo a prospective validation. With the assumption that this binding mode is correct it should be able to predict new active compounds. In that sense, the binding mode complex served as starting structure for generating a structure-based pharmacophore model, which was subsequently used for screening a large vendor database in order to identify new P-gp inhibitors.

3.3.1 Molecular dynamics simulation

The aim of performing molecular dynamics simulations was to be able to select the right binding mode from the two hypotheses obtained from the docking studies. In summary following questions should be answered:

Table 3.1: MD simulation runs.

Protein conformation	Binding mode	Ligand
apo	I	GPV062
		GPV019
	II	GPV062
		GPV019
nb	1	GPV062
		GPV019
	2	GPV062
		GPV019

- Which binding mode is more stable in the protein's binding pocket? Therefore the interactions throughout the simulation have been observed.

- Which binding mode could explain the activity difference between the hydroxyphenylpiperidine derivative GPV062 and the phenylpiperazine analog GPV019? Thus, for each binding mode a simulation for each of the propafenone compounds was performed.

- Which binding mode could explain the affinity decrease between the nucleotide-free (high affinity) and the nucleotide-bound (low affinity) complex? In that sense, each binding mode was simulated both in the nucleotide-free and -bound protein conformation.

In total, eight different simulations have been performed and are listed in Table 3.1.

Setup

A detailed description of the MD simulation setup can found in the Appendix. The next paragraph should just briefly explain the workflow of the simulations.
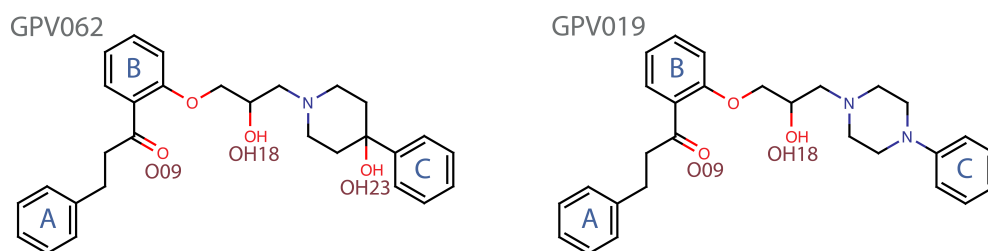
Figure 3.2: Ligands GPV062 and GPV019 that were used in the MD simulations.

For each ligand and binding mode, a ligand-protein complex has been generated and inserted into the membrane. For the membrane a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer was chosen, as POPC lipids are well studied and are naturally present in eukaryotic cells membranes. The complex was minimized and the system was solvated.
Using the software package GROMACS,[1] the whole system has been equilibrated for 1 ns before starting the 10 ns production run. Due to the instability of the NB domains of the protein structure, those parts were position restraints throughout the simulations. Beside that no restraints have been applied.

Results and Discussion

For the analysis mainly the interaction patterns of the ligands to the protein and solvent molecules were investigated. In that sense, the hydrogen bond formation was detected by the program *hbond*, implemented in GROMACS, and a Perl-script provided by Justin Lemkul of the Bevan-Lab.[2] Additionally, possible π-π-interactions could be detected by applying the program *g_ sgangle* of GROMACS. Furthermore in this section the ligands' functional groups will be referred to as depicted in Figure 3.2.

Binding mode stability When analyzing the RMSD (root mean square deviation) of the ligands in the binding site over the trajectory, one can easily

Figure 3.3: RMSD of ligands GPV062 and GPV019 in the binding site of the homology model representing the nucleotide-free P-gp conformation in binding modes I (green) and II (blue).

notice that both ligands, GPV062 and GPV019, stay comparably stable in binding mode I, by showing a maximum RMSD of 5 Å (Figure 3.3).

On the contrary, considering binding mode II large movements of GPV062 (up to 13 Å) could be observed. Also GPV019 showed some fluctuations, as especially after 1 ns the RMSD started to increase considerably. Based on that information binding mode I was considered more stable than binding mode II, and could thus be interpreted as the preferred binding mode.

Activity difference between ligands   Although binding mode I showed clearly more stability in the binding pocket than binding mode II, there still remained the question if one of the two binding modes is able to explain the activity difference between GPV062 and GPV019.

The shape of both ligands is rather similar, as both possess a six-membered

98

heterocycle with an additional phenyl-moiety in para position.

However, GPV062 belongs to the group of derivatives that additionally bear a hyroxyl group in para position on the piperidine ring. These derivatives are one magnitude more active than analogs lacking this group.[3] According to this information, the presence of the OH-group might be responsible for the activity-increase of hydroxyphenylpiperidine derivatives.

In binding mode I, GPV062 formed H-bonds with OH23 to TYR310 and TYR307. In addition, an H-bond between TYR307 and O09 of the ligand could be observed. This is in accordance with the results of the docking study. However, these interactions have been mainly detected in ns 1-4. More stable interactions were found in the form of water mediated H-bonds between OH23 and TYR310 or TRP315 respectively (Table 3.2).

Furthermore, the analysis showed pi-stacking interactions between ring B and TYR307, as well as between ring C and PHE978 (Table 3.3.

With GPV019 no H-bond interactions directly with the protein could be observed. In contrast, very stable H-bonds were formed to water molecules in the binding pocket. A water mediated H-bond was detected between O09 and LEU304 and ALA308 (Table 3.2). Furthermore the residue TYR310 was found to be involved in pi-stacking with the ring C (Table 3.3.

Comparing the interaction patterns of both ligands, GPV062 seemed to form more interactions, which could be a hint for its higher activity. Especially the number of direct interactions with the protein are an indicator for that.

In case of binding mode II, the OH23 group of GPV062 was found to be interacting directly with TYR307 and indirectly (water mediated) with TRP315 and TYR310 (Table 3.2), although quite weakly. π-π-interactions could be observed between ring A and B and PHE72, as well as between ring B and PHE336 (Table 3.3.

GPV019 formed weak interactions with the OH18 group and TYR307 and TYR310 (Table 3.2). With ring B of GPV019 a pi-pi interaction with PHE314 could be detected (Table 3.3).

Also in this case GPV062 showed more direct and indirect interactions throughout the simulation. Furthermore, the interactions observed with GPV062 were primarily formed with OH23. This indicates that also binding

Table 3.2: Detected H-bonds throughout the simulations with the nucleotide-free protein conformation.

		Ligand - Protein		Ligand - H_2O - Protein	
I	GPV062	OH23 - Y310	$+^c$	OH23 - H_2O - W315, Y310	$++^b$
	GPV019	*no stable H-bonds*	-	O09 - H_2O - L304, A308	+
II	GPV062	OH23 - Y307	$\sim^d$	OH23 - H_2O - W315, Y310	+
	GPV019	OH18 - Y307, Y310	$\sim$	*no stable H-bonds*	-

[a] strong interaction: present $> 50\%$ of simulation time
[b] medium interaction: present $> 30\%$ of simulation time
[c] interaction: present $> 10\%$ of simulation time
[d] low interaction: present $< 10\%$ of simulation time

mode II would be able to explain the increased activity of GPV062 compared to GPV019. However all direct interactions turned out to be rather short-lived and thus weak.

Table 3.3: π-π-interaction detected during the simulation.

		nucleotide-free			nucleotide-bound
I	GPV062	ring B - Y307	1	GPV062	ring B - F303
		ring C - F978			
	GPV019	ring C - Y310		GPV019	-
II	GPV062	ring B, C - F72	2	GPV062	ring B - F303, Y310
		ring B - F336			ring C - F335
	GPV019	ring B - F314		GPV019	-

Activity decrease upon conformational change of the protein? The catalytic cycle of P-gp starts with its initial state, that is free of a ligand or ATP with a conformation that is open to the inside of the cell. To this inverted V-shape conformation of the protein, the ligand should bind with
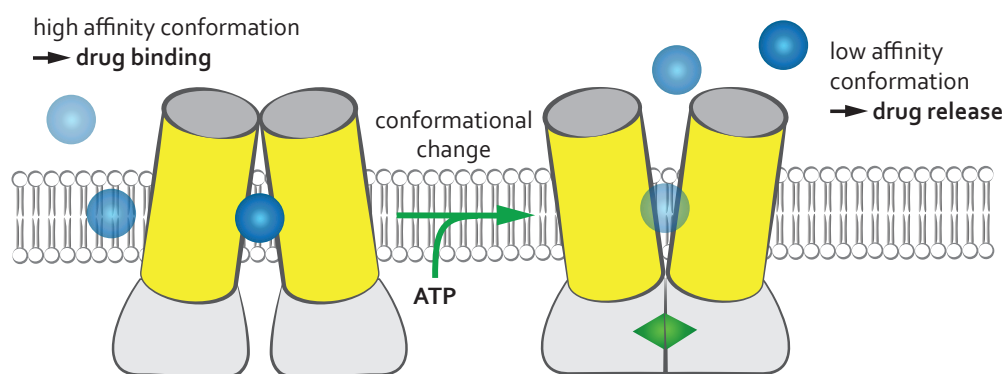
Figure 3.4: Schematic representation of drug transport by P-gp.

high affinity. Ligand-binding eventually stimulates the binding of ATP, which leads to the association of both NBDs and subsequently to a conformational change of the protein. In that sense, the transmembrane (TM) domains get pushed apart and thus assume a wing-like structure, which is open to the extracellular space. Due to the affinity-loss between protein and ligand, the latter is released (Figure 3.4). The hydrolysis of ATP is furthermore needed to restore the initial state of P-gp.[4,5] Based on the theory of the catalytic cycle, the nucleotide-free conformation can be referred to as the high-affinity state for the ligand, whereas the nucleotide-bound conformation represents the low-affinity state.

To investigate the possible affinity decrease, MD simulations using the nucleotide-bound conformation have been performed as well. The poses were selected on basis of similar interaction patterns, as explained in the paper in Section 3.2. According to that, binding mode I of the nucleotide-free conformation corresponds to binding mode 1 when ATP is bound, and similarly binding mode II corresponds to binding mode 2.

Surprisingly, the H-bond analysis revealed more stable direct interactions between ligand and protein in the nucleotide-bound conformation for both binding modes. Almost all interactions were present throughout the 10 ns simulation, with TYR307 and TYR310 being the most prominent interaction partners (Table 3.4). Much less π-π interactions could be detected. These

Table 3.4: Detected H-bonds throughout the simulations with the nucleotide-bound protein conformation.

		Ligand - Protein		Ligand - H_2O - Protein	
1	GPV062	OH18 - Y307	$+++^a$	O09 - H_2O - Y307	$++^b$
	GPV019	OH18 - Y307 O09 - Y307	$+++$ $+++$	*no stable H-bonds*	-
2	GPV062	OH23 - Y307	$+++$	*no stable H-bonds*	-
	GPV019	O09 - Y307	$+++$	*no stable H-bonds*	-

[a] strong interaction: present $> 50\%$ of simulation time
[b] medium interaction: present $> 30\%$ of simulation time
[c] interaction: present $> 10\%$ of simulation time
[d] low interaction: present $< 10\%$ of simulation time

were only observed with GPV062, mainly between ring B and PHE303 and TYR310, as well as between ring C and PHE335.

Interestingly, no water-mediated H-bonds were found during the simulation for both ligands. In fact, the binding area in the open-outward conformation enclosed much more water molecules than in the open-inward state. This resulted in a high number of H-bonds to various water molecules. However, none of these interactions remained stable or could be linked to the protein.

Additionally the movement of ligands in the binding site was analyzed. As can be seen in Figures 3.5 and 3.6 there was more fluctuating movement of all the ligands throughout the simulation, indicating a certain instability of the poses. Especially ligand GPV019 in binding mode 1 showed a displacement of about 7 Å after 4 ns compared to the initial coordinates. By taking a look at the trajectories it could be noticed that in all four simulations the ligand was flipping around in the binding site, with some showing quite a shift at the beginning.

However, the interaction patterns suggested that the ligands have been fixed at certain positions by constant H-bonds with the protein.

According to this information, one might assume that propafenone derivatives are more affine towards the nucleotide-bound conformation of P-gp,

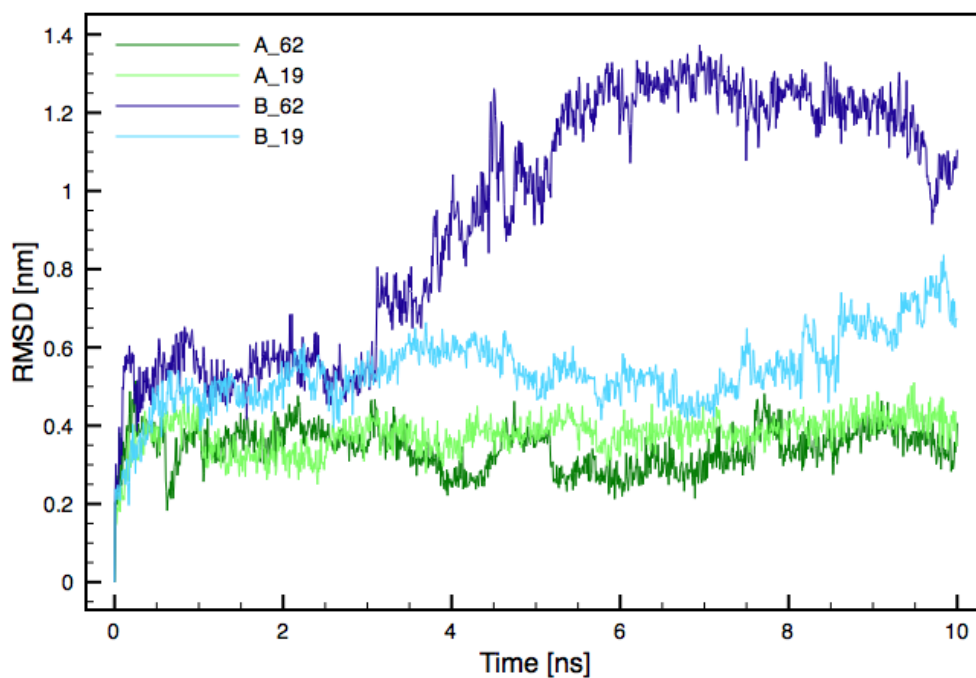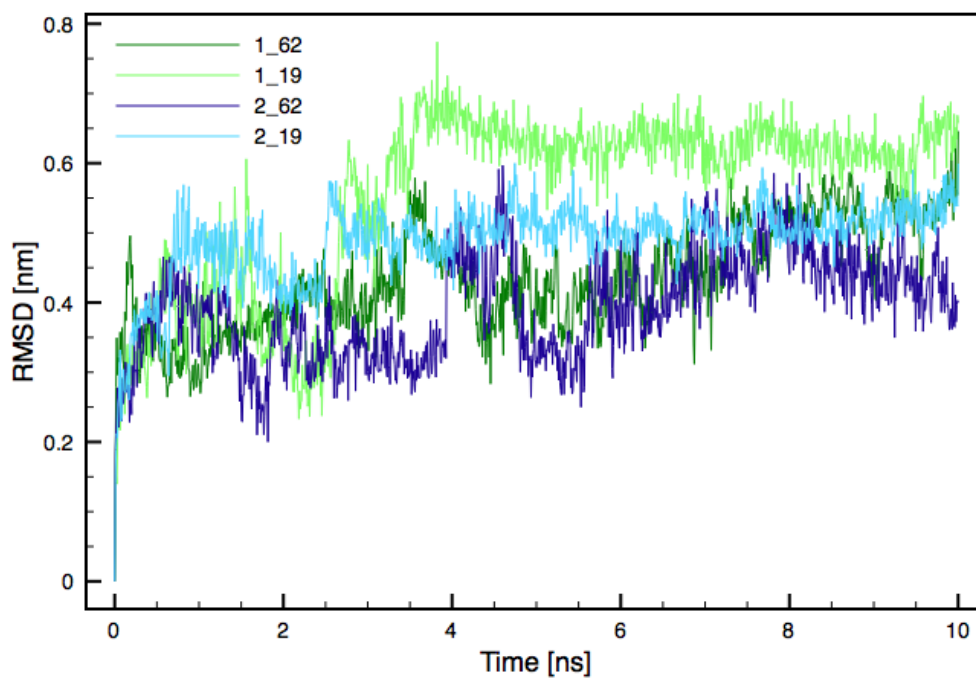**Figure 3.5:** RMSD of ligands GPV062 and GPV019 in the binding site of the homology model representing the nucleotide-bound P-gp conformation in binding modes 1 (green) and 2 (blue).
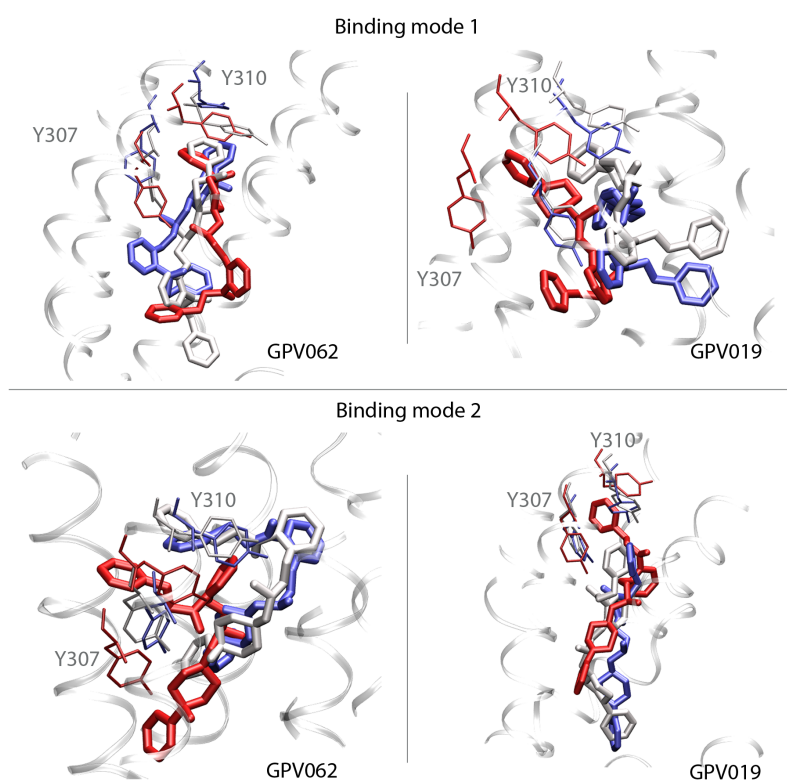
Figure 3.6: Ligand movement within the binding sites for binding mode 1 and 2 in the nucleotide-bound conformation of the protein. The color coding represents the time step (red: start, white: intermediate, blue: end).

which might explain the inhibitory effect of these ligands. If the ligands bind to P-gp, but remain in the binding site, the transport is not completed and the protein would be unable to bind another ligand.

On the other hand, the high amount of H-bonds to random water molecules in the nucleotide-bound conformation could result in an attenuation of the ligand-protein interactions, which would further result in the release of the ligand. To confirm this, much longer simulations would be needed.

Additionally, there still remains the question, if the corresponding binding modes selected in the nucleotide-bound conformation really represent the end-point of the ligand translocation. The final proof for the affinity difference could only be demonstrated by performing a targeted MD of the bound ligand of the whole transport cycle. A similar experiment was conducted by John Wise,[6] who ran a targeted MD of P-gp from the open-inward to the open-outward conformation of the protein. By applying multidimensional QR fractorization,[7,8] 26 non-redundant protein conformations have been extracted from the whole trajectory, which have further been used for docking studies. It could be shown, that the ligand binding poses in the open-outward conformation could preferably be found in the cytoplasmic site of the transmembrane area. This lead to the assumption that binding to the extracellular half of the TM domain would result in destabilization of the ligand. In contrast, docking into the open-inward conformation showed no binding preference for any of the halves in the TM domain. This might be in contrast to the results presented in this study. However, Wise et al. did not analyze the stability of the poses in a dynamic environment, the results of which might have shown the opposite. Furthermore no differences between substrates and inhibitors could be detected, leaving the question of how P-gp could be inhibited, unanswered.

As stated before, the simulation of the whole transport cycle of a ligand by P-gp should be investigated.

3.3.2 Pharmacophore modeling and screening[†]

With the application of an exhaustive docking protocol, described in Chapter 3.2, and subsequent MD simulations a binding mode could be identified that is able to explain SAR and experimental data, but is also stable in a dynamic setting.

Further validation of this binding mode hypothesis should involve testing the ability to predict new P-gp inhibitors. In that sense the docking pose of GPV062 in binding mode I was used as starting geometry for a structure-based virtual screening approach, which in contrast to ligand-based screenings integrates information of the protein binding site into the compound search.

Pharmacophore modeling is the method of choice to screen for new active compounds having a distinct chemical scaffold. Instead of taking the configuration of a ligand into account, it tries to define important pharmacophoric features in the molecule.[10,11] In that sense, a hydroxyl group is no longer recognized as OH, but rather as a hydrogen bond acceptor and donor. Whereas simple similarity screenings remain in the same chemical space, pharmacophore modeling is able to perform scaffold hopping.[12,13]

There are several software packages that are able to generate ligand-based pharmacophore models.[14–17] However, the program LigandScout[18] provides a tool for generating pharmacophore models on basis of a drug-receptor complex.

Pharmacophore model generation

In this study several structure based pharmacophore models have been built, using the docking pose of GPV062 in binding mode I as starting geometry. These models shared aromatic features at the phenyl-rings of the propafenone-scaffold, but varied regarding the treatment of the tertiary nitrogen atom and the hydroxy and carbonyl groups. Previous studies showed that the importance of the nitrogen atom in propafenone derivatives is due to its hydrogen bond acceptor strength rather than to its positive charge.[19] In

---

[†]The work was done in cooperation with Katharina Prokes.[9]

contrast, recent charge repulsion experiments suggest the tertiary nitrogen atom to be protonated.[20] In LigandScout both theories could be implemented either by defining the nitrogen as an H-bond acceptor or by generating a positive ionizable feature around it.

Furthermore, the docking pose selection procedure based on the fact that propafenone derivatives with a 4-hydroxy-4-phenyl moiety at the piperidine ring are more active than expected from their logP. This exception of the activity-logP correlation of P-gp inhibitors, could indicate the presence of an H-bond between the 4-hydroxy group and the protein. This information was also implemented in the pharmacophore model. Depending on the model either H-bond acceptor, H-bond donor or both features have been assigned to this atom.

For selecting the best performing pharmacophore model, our in-house database, comprising 374 propafenone derivatives, was screened. As this database contains activity data from P-gp efflux assays, it provides a quantitative measure for the predictive quality of the models. Only one model was able to predict the highly active 4-hydroxy-4-phenyl piperidine analog GPV062 at top. This model comprised in total 8 features: three hydrophobic features at the phenyl rings of the molecule, a positive ionizable property around the nitrogen atom and H-bond donor and acceptor features at the 4-hydroxy group. As the hydroxy group is able to capture two directions, in each direction a donor and an acceptor feature was placed. Additionally, 19 exclusion volumes were included, defining the shape of the binding site (Figure 3.7). They represent a steric limitation for the screening compounds which would not be considered in a ligand-based pharmacophore model.

The confusion matrix parameters have been calculated by defining all compounds with an activity $< 100\,\text{nM}$ as inhibitors and all others as non-inhibitors. According to this threshold, the model could correctly classify $77\,\%$ of the database compounds into inhibitors or non-inhibitors.
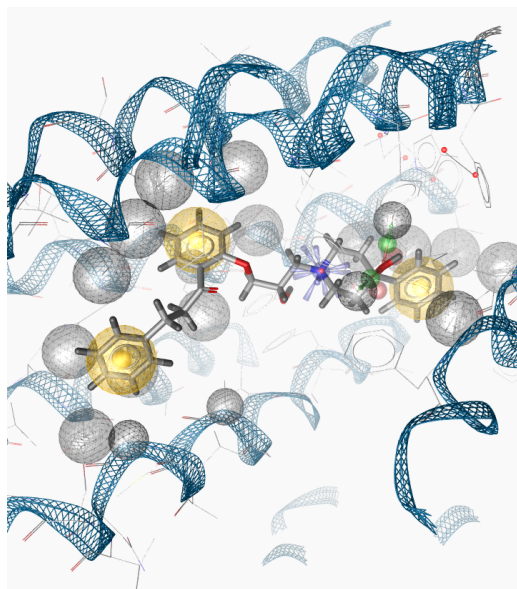
Figure 3.7: Selected pharmacophore model. The yellow spheres represent hydrophobic features, the positive ionizable feature around the tertiary nitrogen is marked in blue and the hydrogen bond donor and acceptor features describe the interaction of the 4-hydroxy group. The shape of the binding site is considered by using exclusion spheres, depicted as grey spheres.

Pharmacophore model validation

As the in-house database mostly contained propafenone type P-gp inhibitors, an additional evaluation of the model's predictive quality with different compound classes was performed. Therefore two P-gp inhibitor sets, published by Broccatelli et al.[21] and Chen et al.,[22] were combined, curated and prepared for screening with the selected pharmacophore model. After all duplicates have been removed this literature dataset contained 1954 compounds, comprising 1208 actives and 746 inactives, indicated with 1 or 0 respectively. Up to 500 conformations per ligand have been generated by applying the omega-best protocol within LigandScout. For screening all features were defined as obligatory, however one feature was allowed to be omitted by each compound. To ensure that the best conformational fit of each molecule was found, the first-match mode option of LigandScout was inactivated.

The pharmacophore model turned out to be highly specific, excluding 98 % of

the non-inhibitors in the database. However, the relatively large number of false negatives could be traced back to the polyspecific nature of P-gp. The protein is assumed to possess one large binding pocket that is comprised of several small binding sites.[5] So far three interaction sites have been characterized, referred to as the R-site, which binds Rhodamine,[23] the H-site known to interact with Hoechst 33342[24] and a regulatory site where prazosin/progesterone are binding.[25][26] As the pharmacophore model represents only one of several distinct binding sites of P-gp, one has to bear in mind that some active compounds might bind to different regions in the binding pocket.

Further evaluation involved virtual screening of a spiked DUD set. Therefore all 1208 P-gp inhibitors of the merged database and 40 compounds of our in-house database showing activities below 100 nm were included into the high number of decoy set. Together with the decoys, the screening database comprised in total 94888 entries. Again, the pharmacophore model performed well, exhibiting an accuracy of about 98 %. Furthermore an enrichment factor of 9.2 was retrieved after screening 0.5 % of the database. Because of these results the pharmacophore model was considered as a qualified tool for the prediction of P-gp-inhibitors.

Pharmacophore screening

In order to test the predictive quality of the pharmacophore model, a commercial database should be screened and selected hits should be tested experimentally for their P-gp inhibiting activity. In that sense the LifeChemicals database, comprising 308038 compounds has been screened and all compounds that fulfilled 7 out of 8 pharmacophoric features of the model were considered as hits. After visual inspection 7 compounds have been identified that were further tested via a Rhodamine123 efflux assay. As described by Parveen et al.,[20] HEK293/EBNA cells were incubated in the presence of different inhibitor concentrations and transport of the known P-gp substrate Rhodamine123 was detected by fluorescent emission at 534 nm wavelength.

Additionally, the LifeChemicals database was used to perform a similarity screen by calculating SHannon Entropy Descriptors (SHED).[27] SHED

Table 3.5: Statistics of pharmacophore model evaluation.

	in-house	literature[a]	spiked DUD
N	374	1954	9488
Actives	40	1208	1321
Inactives	334	746	93567
Hits	87	76	400
TP	20	62	61
TN	267	732	93228
FP	67	14	339
FN	20	1146	1260
Sensitivity	0.50	0.05	0.05
Specificity	0.80	0.98	0.99
Accuracy	0.77	0.41	0.98
MCC	0.22	0.08	0.08
EF 0.5 %[b]	9.35	1.29	9.24
EF 1 %[c]	4.67	0.96	4.61

[a] literature database (Cruciani et al. and Chen et al.)
[b] enrichment factor for 0.5 % of the screening database
[c] enrichment factor for 1 % of the screening database
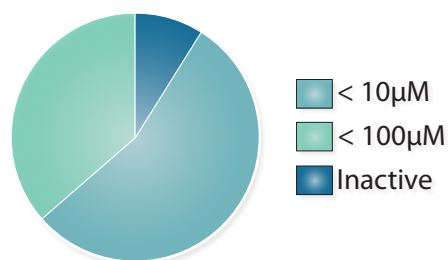
Figure 3.8: Screening results. Out of 11 experimentally tested compounds, 4 showed activity below $10\,\mu$M, 6 below $100\,\mu$M and only 1 molecule was inactive.

make use of the information-theoretical concept of Shannon Entropy and represent a means to quantify the variability displayed by topological distributions of atom-centered feature pairs in molecules. In that sense, a similarity search using SHED should provide compounds with similar arranged structural features while exhibiting distinct scaffolds to the highly active propafenone derivative GPV062.

Out of the 308038 compounds of the LifeChemicals database, 410 hits have been identified by the SHED similarity screen as being 82 % similar to GPV062. This number of diverse hits were again screened with the structure-based pharmacophore model. Only eighteen substances contained structures that fitted at minimium 6 out of 8 pharmacophoric features of the model. The inhibiting activity of five selected compounds of those was as well determined experimentally.

Finally, in total eleven compounds, identified by pharmacophore screening alone and in combination with the SHED similarity search, have been tested using the Rhodamine efflux assay. Among those, six compounds demonstrated activities in the double digit, and three even in the single digit μmolar range. Only one compound was considered as inactive, showing an IC50 value of $132.34\,\mu$M (Figure 3.8).

With this finding not only the retrospective quality of the docking pose, but also its prospective capability of identifying new inhibitors, could be shown.

Bibliography

[1] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation 4: 435–447.

[2] Lemkul J (Accessed 31 July 2012). plot_hbmap.pl. http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/Scripts/ plot_hbmap.txt.

[3] Chiba P, Hitzler M, Richter E, Huber M, Tmej C, et al. (1997) Studies on propafenone-type modulators of multidrug resistance iii: Variations on the nitrogen. Quant Struct-Act Rel 16: 361–366.

[4] Sauna ZE, Kim IW, Nandigama K, Kopp S, Chiba P, et al. (2007) Catalytic cycle of atp hydrolysis by p-glycoprotein: evidence for formation of the e.s reaction intermediate with atp-gamma-s, a nonhydrolyzable analogue of atp. Biochemistry 46: 13787–99.

[5] Seeger MA, van Veen HW (2009) Molecular basis of multidrug transport by abc transporters. Biochim Biophys Acta 1794: 725-737.

[6] Wise JG (2012) Catalytic transitions in the human mdr1 p-glycoprotein drug binding sites. Biochemistry 51: 5125-41.

[7] O'Donoghue P, Luthey-Schulten Z (2005) Evolutionary profiles derived from the qr factorization of multiple structural alignments gives an economy of information. J Mol Biol 346: 875-94.

[8] Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) Multiseq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics 7: 382.

[9] Prokes K (2012) Development of in silico models for Identification of New Ligands Acting as Pharmacochaperones for P-glycoprotein. Master's thesis, University of Vienna.

[10] Wermuth CG (2006) Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In: T Langer RDH, editor, Pharmacophores and Pharmacophore Searches, Wiley-VCH, Weinheim, volume 32 of *Methods and Principles in Medicinal Chemistry*. pp. 3-13.

[11] Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15: 444-50.

[12] Kirchmair J, Distinto S, Schuster D, Spitzer G, Langer T, et al. (2008) Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates. Curr Med Chem 15: 2040-53.

[13] Schuster D, Wolber G (2010) Identification of bioactive natural products by pharmacophore-based virtual screening. Curr Pharm Des 16: 1666-81.

[14] Chemical Computing Group Inc, Montreal, QC (2011). Molecular operating environment (MOE), 2011.10.

[15] Accelrys Software Inc San Diego, CA (2007). Discovery studio modeling environment.

[16] Schrödinger, LLC, New York, NY (2012). Phase.

[17] Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) Phase: a new engine for pharmacophore perception, 3d qsar model development, and 3d database screening: 1. methodology and preliminary results. J Comput Aided Mol Des 20: 647-71.

[18] Wolber G, Langer T (2005) Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45: 160-9.

[19] Ecker G, Huber M, Schmid D, Chiba P (1999) The importance of a nitrogen atom in modulators of multidrug resistance. Mol Pharmacol 56: 791–6.

[20] Parveen Z, Stockner T, Bentele C, Pferschy S, Kraupp M, et al. (2011) Molecular dissection of dual pseudosymmetric solute translocation pathways in human p-glycoprotein. Mol Pharmacol 79: 443–52.

[21] Broccatelli F, Carosati E, Neri A, Frosini M, Goracci L, et al. (2011) A novel approach for predicting p-glycoprotein (abcb1) inhibition using molecular interaction fields. J Med Chem 54: 1740–51.

[22] Chen L, Li Y, Zhao Q, Peng H, Hou T (2011) Adme evaluation in drug discovery. 10. predictions of p-glycoprotein inhibitors using recursive partitioning and naive bayesian classification techniques. Mol Pharm 8: 889–900.

[23] Loo TW, Clarke DM (2002) Location of the rhodamine-binding site in the human multidrug resistance p-glycoprotein. The Journal of Biological Chemistry 277: 44332–8.

[24] Qu Q, Sharom FJ (2002) Proximity of bound hoechst 33342 to the atpase catalytic sites places the drug binding site of p-glycoprotein within the cytoplasmic membrane leaflet. Biochemistry 41: 4744-4752.

[25] Shapiro AB, Fox K, Lam P, Ling V (1999) Stimulation of p-glycoprotein-mediated drug transport by prazosin and progesterone. evidence for a third drug-binding site. Eur J Biochem 259: 841-50.

[26] Klepsch F, Ecker GF (2010) Impact of the recent mouse p-glycoprotein structure for structure-based ligand design. Molecular Informatics 29: 276–286.

[27] Gregori-Puigjané E, Mestres J (2006) Shed: Shannon entropy descriptors from topological feature distributions. J Chem Inf Model 46: 1615–22.

Ligand and structure-based classification models for Prediction of P-glycoprotein inhibitors

Freya Klepsch†, Poongavanam Vasanthanathan†, Gerhard F Ecker *

Department of Medicinal Chemistry, University of Vienna, Althanstraße 14, A-1090,

Vienna, Austria.

†These authors contributed equally to this work.

*Email: gerhard.f.ecker@univie.ac.at

ABSTRACT

The ABC transporter P-glycoprotein (P-gp) actively transports a wide range of drugs and toxins out of cells, and is therefore related to multidrug resistance and the pharmacokinetic profile of therapeutics. Thus, the development of suitable prediction models for the identification of P-gp inhibitors is essential for the improvement of drug therapy. So far *in-silico* P-gp inhibitor prediction was dominated by ligand-based approaches, due to the lack of high-quality structural information about P-gp. The present study aims at comparing the P-gp inhibitor/non-inhibitor classification performance of docking into a homology model of P-gp, to supervised machine learning methods, such as Kappa nearest neighbor, support vector machine, random forest and binary QSAR, by using a large, structurally diverse data set. In addition, the applicability domain of the models was assessed using an algorithm based on Euclidean distance. From the applied ligand- and structure-based classification models, the ligand-based methods random forest and SVM performed best for classification of P-gp inhibitors and non-inhibitors, correctly predicting 73/75 % of the external test set compounds. Classification based on the docking experiments using the scoring function Chemscore resulted in the correct prediction of 61 % of the external test set. This demonstrates that ligand-based models are the methods of choice for accurately predicting P-gp inhibitors. However, structure-based classification offers information about possible drug/protein interactions, which helps in understanding the process of inhibition and could furthermore be applied in lead optimization.

INTRODUCTION

The ABC transporter (<u>A</u>TP <u>B</u>inding <u>C</u>assette) family is one of the largest protein families comprising a group of functionally distinct proteins that are mainly involved in actively transporting chemicals across cellular membranes. Depending on the subtype, transported substrates range from endogenous amino acids and lipids, up to hydrophobic or charged small molecules.[1] In total more than 80 genes for ABC transporters have been characterized across all animal families, among which fifty-seven genes were reported for vertebrates. Human ABC transporters comprise 48 different proteins that can be divided into seven different sub-families; ABCA, ABCB, ABCC, ABCD, ABCE, ABCF and ABCG.[2] The correct function of ABC transporters is of high importance, as mutations or deficiency of these membrane proteins lead to various diseases such as immune deficiency (ABCB2), cystic fibrosis (ABCC7), progressive familial intra-hepatic cholestasis-2 (ABCB11), and Dubin-Johnson syndrome (ABCC2). Moreover, some highly polyspecific ABC transporters are known for their ability to export a wide variety of chemical compounds out of the cell. Overexpression of these so-called multidrug transporters, which include P-glycoprotein (P-gp, ABCB1), ABCC1 and ABCG2, might lead to the acquisition of multidrug resistance (MDR), which is one major reason for the failure of anti-cancer and antibiotic treatment.[3]

Furthermore, P-gp plays an essential role in determining the ADMET (<u>A</u>bsorption, <u>D</u>istribution, <u>M</u>etabolism, <u>E</u>xcretion and <u>T</u>oxicity) properties of many compounds. Drugs that are substrates of P-gp are subject to low intestinal absorption, low blood-brain barrier permeability, and face the risk of increased metabolism in intestinal cells.[4] Moreover, P-gp modulating compounds are capable of influencing the pharmacokinetic profiles of co-administered drugs that are either substrates or inhibitors of P-gp[5, 6], thus giving rise to drug-drug interactions. This urges on the

development of suitable *in silico* models for the prediction of P-gp inhibitors in the early stage of the drug discovery process to identify potential risks. So far the focus of prediction models was lying on ligand-based approaches such as QSAR[7], rule-based models[8] and pharmacophore models[9-11]. Very recently, also machine-learning methods have been successfully used for the prediction of P-gp substrates and inhibitors.[12, 13] For example, grid-based methods (FLAP) have been applied to a set of 1200 P-gp inhibitors and non-inhibitors with a success rate of 86 % for an external test set.[14] Also unsupervised machine learning methods (Kohonen self-organizing map) were used to predict substrates and non-substrates from a dataset formed by 206 compounds. In this study the best model was able to correctly predict 83 % of substrates and 81 % of inhibitors.[13] Recently, Chen et al. reported recursive partitioning and naïve Bayes based classification to a set of 1273 compounds. In this case, the best model predicted accurately 81 % of the compounds of the test set.[15] Due to the lack of structural information, developing prediction models using structure-based approaches has not been actively pursued. However, in the recent years the number of available 3D structures of ABC proteins[16,17] and the performance of experimental approaches[18] has paved the way for the application of structure-based methods to predict drug/transporter interaction. In that sense, a small number of structure-based prediction models have been developed in the last two years. Bikadi et al. built a free web-server for online prediction of P-gp substrates based on a SVM classification model.[19, 20] Molecular docking into the crystal structure and a homology model of mouse P-gp were used to additionally generate possible protein-ligand complexes. Dolghih et al. used induced fit docking into the crystal structure of mouse P-gp to separate P-gp binders from non-binders on the basis of their docking score.[21] Although the datasets were considerably small (126 and 64 compounds), an AUC of 0.93 and 0.90,

respectively, could be observed. Very recently also Chen et al. used a set of 245 P-gp substrates and non-substrates to assess the prediction capability of docking.[22] Nevertheless, based on the Glide docking scores SP and XP, no clear separation of the two classes could be observed.

However, the above mentioned machine-learning and structure-based studies only used data sets of relatively small size, which might not be sufficient for the correct prediction of P-gp, which is known for its high polyspecificity. Thus, in the present study, we applied supervised machine-learning and structure-based techniques to predict P-gp inhibitors and non-inhibitors, using a large and structurally diverse dataset, comprising 1079 compounds. The methods applied comprised 1. ligand-based supervised machine learning (ML) methods, including random forest (RF), decision tree (DT), support vector machine (SVM), Kappa nearest neighbor (kNN) and binary QSAR (BQSAR), and 2. structure-based docking studies using five different scoring functions (ChemScore, GoldScore, ASP, ChemPLP and XScore).

5

COMPUTATIONAL METHODS

P-gp Inhibitors Data Source. The publications from Broccatelli et al.[14] and Chen et al.[15] served as starting point for the present classification studies. A set of 2548 compounds reported as P-gp inhibitors and non-inhibitors was compiled from both literature sources. In brief, Broccatelli et al. compiled a dataset of 1275 compounds from more than 60 literature references. Threshold values for inhibitors and non-inhibitors were assigned based on the IC50 values and on the percentage of inhibition as suggested by Rautio et al.[23] Compounds with an IC50 ≤15 μM, or > 25-30 % of inhibition were considered as inhibitors. Conversely, compounds possessing IC50 and % of inhibition values of ≥ 100 μM or < 10-12 % were classified as non-inhibitors. In addition, Tingjun Hou kindly provided us with the 3D structures of 797 inhibitors and 476 non-inhibitors from their data set, as recently published by Chen et al.[15] They used MDRR (Multi-drug-resistance ratio) values measured in adriamycin-resistant P388 murine leukemia cells for classification. Compounds with MDRR values greater than 0.5 were assigned inhibitors, whereas molecules with lower or equal to 0.4 MDRR values were considered as non-inhibitors.

Both datasets were analyzed in order to eliminate duplicated structures using 2D SMILES representations of each compound. In that sense, the dataset was also cleaned from stereoismers that have been equally assigned (inhibitor or non-inhibitor). While for the Chen database no duplicates have been found, from the Broccatelli dataset 53 compounds have been removed due to identical 2D structures. Additionally 429 compounds were found to be present in both datasets. Among those, 33 compounds were differently annotated in the two datasets, and 17 possessed a permanent charged. Those molecules have been removed. The residual 346 compounds (132 inhibitors and 214 non-inhibitors) were stored as external test set. Finally, the fused dataset

comprised 1699 unique compounds, from which 91 permanently charged molecules have been removed. This lead to a dataset of 1608 compounds, comprising 1076 inhbitors and 532 non-inhibitors.

Molecular Descriptors and Fingerprint Calculation. The 3D structures of the dataset were imported into the modeling software MOE (Version 2010.10)[24] and subsequently energy minimized using the MMFF94x force field. The energy-minimized molecules were used to compute 62 2D descriptors implemented in MOE. The 2D molecular descriptors calculated comprised physicochemical properties, atom and bond counts, and pharmacophoric features. In addition, a set of 166 MACCS fingerprints and a set of 307 substructure fingerprints were calculated using the freely available software PaDEL[25]

Principle Component Analysis (PCA). A PCA of the whole dataset was conducted using the software SIMCA-p (version 10.5). The descriptors included for PCA have been selected based according the variable importance (VIP) calculated in SIMCA. A complete list of descriptors is provided in the Supplementary Material, SI-Table 1.

Selection of Training and Test set. The activity of the compounds was represented by the introduction of a binary variable (1 for inhibitor, 0 for non-inhibitor). Subsequently, for assessing the internal predictivity of the models, the dataset was divided into training and test set using D-optimal onion design (DOOD) as implemented in the MODDE software (Version 7.0).[26] DOOD is a multivariate method, used for selecting training and test sets of reasonable size, which are representatives for the chemical property space defined by the molecular structures. The general principle of DOOD can be found elsewhere.[27, 28] In the present study we used the scores from principal component analysis (R2=0.99, 25 principal components) calculated by SIMCA-P.

Machine Learning Methods and Attribute Selection. In the present study, a set of representative classification machine learning methods such as SVM, kNN, DT, RF and BQSAR was used. These classifiers are primarily used for ADMET property prediction, since they are efficient to handle large compounds. The principles of these methods have been described in detail elsewhere.[29, 30] SVM, RF, kNN and DT classification experiments were performed using the WEKA data mining software (Version 3.6.4)[31], which provides a set of classifications, regressions, attribute (variable) selections and clustering methods. BQSAR was performed using the tool "QuaSAR-model" implemented in MOE. For descriptor selection, an automatic attribute selection procedure called BestFirst search algorithm, as implemented in WEKA was used.[29] The BestFirst attribute selection has been shown to be a better attribution selection method as compared to GeneticSearch or the use of all the descriptors. [32]

Docking and Scoring Functions. The atomic coordinates of human P-glycoprotein required for docking have been obtained by homology modelling. The model has been built on basis of the X-ray structure of murine P-gp (PDB ID: 3G5U, 3.8 Å) as described in our previous work.[33] The template structure was chosen because of its high sequence identity and the fact that it represented the binding competent state of the transporter.

The protein was prepared using the Protein preparation wizard, implemented in the Schrödinger Suite (2011).[34] During the process, hydrogen atoms were added, optimal protonation states and ASN/GLN/HIS flips were determined. The 3D coordinates of the ligands were built with CORINA and energy minimized with MOE, using the MMFF94x force field.

According to our previous study,[35] the inhibitory activity of tertiary amines is due to the H-bond acceptor strength of the nitrogen rather than its positive charge, which

8

suggested that the ligands might bind in an unprotonated way. On the other hand charge-repulsion experiments indicate that P-gp ligands probably possess a positive charge.[36] Thus, separate docking runs were performed considering neutral and charged molecules. The correct protonation state was calculated using the program LigPrep, implemented in the Schrödinger Suite. Two cyclopeptides could not be processed by LigPrep and thus have been excluded from the dataset.

The remaining 1606 molecules, comprising 1073 inhibitors and 533 non-inhibitors, were used for docking with the genetic algorithm-based GOLD suite (Version 5.1.0).[37] The active site was specified as the entire transmembrane (TM) region of the protein, thus taking 20 Å around the coordinates of the center point (21.07, 57.95, -2.31) into consideration. All the docking runs were performed in high throughput mode as implemented in GOLD. Concerning the fitness function used during docking, either ChemScore (CS) or GoldScore (GS) has been chosen. Together with the different protonation settings of the ligand database this resulted in a total of four docking runs (Table 1). The resulting docking poses were subsequently rescored with five scoring functions implemented in GOLD, which comprised ChemScore, GoldScore, Astex Statistical Potential (ASP)[38] and Piecewise Linear Potential (ChemPLP)[39], as well as the external scoring function XScore.[40] Altogether there were four different docking runs, each of which was scored with five different fitness functions, resulting in 20 scoring models for which the prediction capabilities have been investigated (Table 1).

Table 1. Summary of docking runs performed and scoring functions used in this study.

Docking run	Ligand protonation state	Main scoring function	Rescoring functions
1	Neutral	ChemScore	GoldScore, ASP, ChemPLP, XScore
2		GoldScore	ChemScore, ASP, ChemPLP, XScore
3	Protonated	ChemScore	GoldScore, ASP, ChemPLP, XScore
4		GoldScore	ChemScore, ASP, ChemPLP, XScore

In order to get deeper insights into the binding modes of P-gp inhibitors and non-inhibitors, the protein-ligand interaction fingerprints (PLIF) of the resultant complexes have been analyzed. As the standard PLIF tool in MOE does not support π-π-interactions, a customized svl script has been used that calculated fingerprints from interactions provided by the ligand interaction module in MOE. The types of intermolecular interaction provided comprised ionic, hydrogen bond and pi-pi.

Model Evaluation. The quality of the classification models was evaluated in terms of standard parameters derived from the confusion matrix (true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN)). The predictive abilities of inhibitor and non-inhibitor classification were calculated from sensitivity (Eqn.1) and specificity (Eqn.2) terms respectively. The G-mean value (Eqn.3) was used to measure the balanced prediction of each of the two classes.

$$Sensitivity = \frac{TP}{TP + FN} \qquad\qquad \text{Eqn. 1}$$

$$Specif\,icity = \frac{TN}{TN + FP} \qquad\qquad \text{Eqn. 2}$$

$$G - Mean = \sqrt{Sensitivity \times Specificity} \qquad\qquad \text{Eqn. 3}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(FP + FN)(TN + FP)(TN + FN)}} \qquad \text{Eqn. 4}$$

The quality of the overall binary classification model was estimated using Matthews's correlation coefficient (MCC, Eqn.4).

N-Fold Cross Validation (N-FCV). In addition to the internal and external test set prediction, the model quality was estimated via n-fold cross validation of the training set. In N-FCV, The original dataset is divided into n subsets 10 fold in the case of this study. Out of 10 subsets, 9 subsets (n-1) were used as training set, and the remaining

single subset was retained as validation data for testing the trained model. This process is repeated 10 times and each one of the 10 subsets was used exactly once for validation. In the present study all the N-FCV were carried out as implemented in WEKA.

RESULTS AND DISCUSSION

Characterization of the dataset. An initial set of 1608 P-gp ligands was divided into training and internal test set using D-optimal onion design (DOOD). Thus, the DOOD analysis resulted in 1201 training (841 inhibitors, 360 non-inhibitors) and 407 test compounds (235 inhibitors, 172 non-inhibitors) (internal test set). Principal component analysis (PCA) was performed as explained in the methods section, to inspect potential clusters and the coverage of the chemical space of the P-gp ligands. The first two principal components explained 71.7% of the variance in the dataset (Figure 1A). In the score plot a distinct cluster of inhibitors at the right top corner could be observed, which mainly comprised cyclopeptolide derivatives. Moreover, there was quite a good separation between inhibitors and non-inhibitors, which urged for the development of classification models. In order to understand the influence of the descriptors on the first two PCs, the loading plot was analyzed (Figure 1B). It can be seen from the loading plot that the majority of the inhibitors are highly influenced by the descriptors that provide hydrophobic information, e.g. the number of aromatic bonds (b_aro) or the partition coefficient (logP(o/w)). Furthermore, the high contribution of LogS to non-inhibitors indicates that non-inhibitors are considerably more hydrophilic than inhibitors. The hydrophobic requisite of P-gp inhibitors can be explained by the need of diffusing through the cell membrane in order to effectively bind to the hydrophobic active site of the protein.[41]
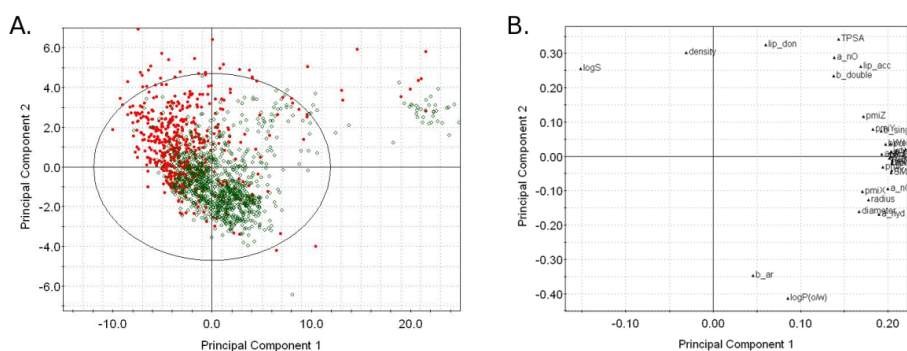
Figure 1. A. Score plot from principal component analysis (first two principal components shown). Inhibitors are shown in green circles and non-inhibitors are shown in red dots, B: Loading Plot of descriptors used for PCA analysis.

In addition, our previous docking study on propafenone-type ligands revealed that the active site of P-gp is primarily formed by the hydrophobic residues Tyr307, Tyr310, Phe343, Leu724, Phe336, Ile731, Ala761 and Val981 (Figure 2).[33]
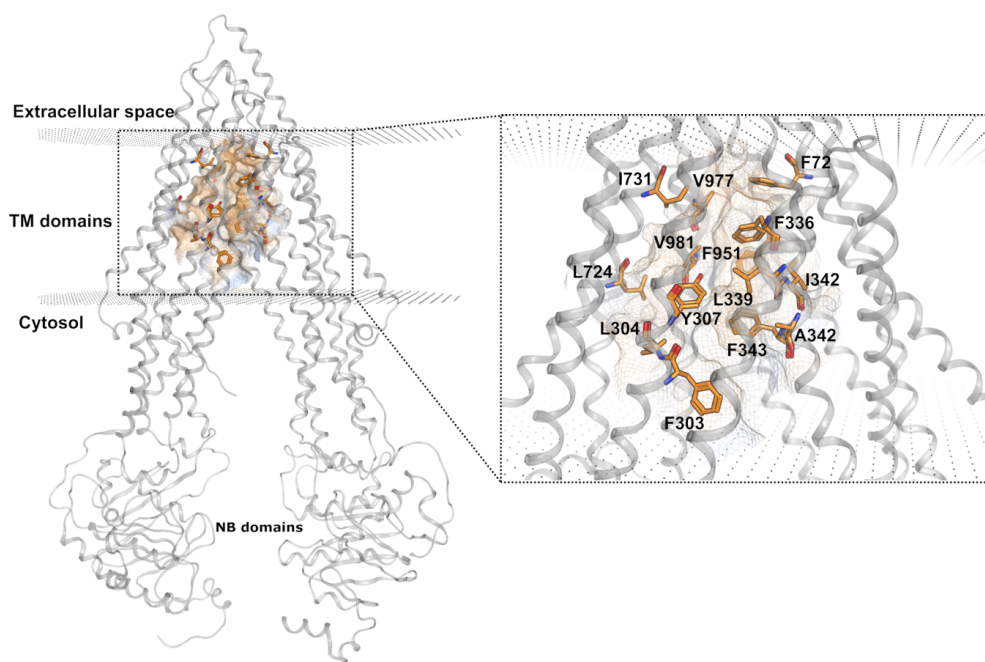
Figure 2. Hydrophobic binding site, formed by non-polar residues of both TM domains.

12

Furthermore, the distribution of inhibitors and non-inhibitors (n=1608) in the dataset on the basis of some common molecular properties was studied. The examination showed that inhibitors and non-inhibitors could be quite well differentiated according to the logP, molecular weight, logS or molar refractivity (the distribution plots are provided in the Supporting Information, SI-Figure 1). From the intersection point of the inhibitor and non-inhibitor distribution curve, the true classification (TP, TN) and misclassification (FP, FN) rates were measured, which were needed for calculating statistical parameters, such as MCC, sensitivity, specificity and overall accuracy of the classification. The summary of the results is given in Table 2. The results show that molecular weight, logS, logP and molar refractivity (MR) lead to a good discrimination between inhibitors and non-inhibitors (MCC > 0.4, overall accuracy ≥ 69 %). In particular, molecular weight and MR correctly discriminated 78 and 79% of the compounds at the intersection of 300 and 10, respectively. The majority of P-gp inhibitors are of relatively bulky and hydrophobic nature compared to compounds that do not inhibit the protein. Imbalanced and hence poor separation was observed with the models derived from the number of hydrogen bond donors (Sensitivity: 94 %, specificity: 20 %), hydrogen bond acceptors (Sensitivity: 85 % and Specificity: 29 %) and oxygen and nitrogen atoms (Sensitivity: 83 % and Specificity: 31 %). Drug-likeness of the dataset was analyzed using Lipinski's rule-of-five properties.[42] It was inferred that 80.9 % of the compounds in the dataset follow the drug-likeness rules, violating at maximum only one of the Lipinski rules.

Table 2. Models obtained from common molecular descriptors distribution.

Property	Intersection Point	Confusion matrix				Sens.	Spec.	MCC	Accuracy
		TP	TN	FP	FN				
H-Acc.	2.5	902	156	385	165	0.85	0.29	0.16	0.66
H-Don	3.5	1005	108	433	62	0.94	0.20	0.22	0.69
LogP	3	886	334	208	180	0.83	0.62	0.45	0.76
LogS	-4	896	355	186	171	0.84	0.66	0.50	0.78
MR	10	894	373	168	173	0.84	0.69	0.53	0.79
MolWt	300	1013	238	303	54	0.95	0.44	0.48	0.78
N+O	3.5	883	168	373	184	0.83	0.31	0.16	0.65

H-Acc.: Number of hydrogen bond acceptor, H-Don.: Number of hydrogen bond donors, LogP: Logarithm of partition coefficient (octonal/water), LogS: Logarithm of water solubility, MR: Molar refractivity, MolWt: Molecular weight, N+O: Number of nitrogen and oxygen, Sens.: Sensitivity, Spec.: Specificity.

Development of Machine Learning Models. Different machine learning methods were used to build P-gp inhibitor and non-inhibitor classification models using a set of 1201 training compounds encoded by physicochemical descriptors and fingerprints. Models have been either built from all descriptors or on the basis of descriptors selected by the BestFirst algorithm. These two scenarios were applied to three sets of X-variables such as 2D physicochemical properties (n=62), MACCS fingerprints (n=166) and substructure fingerprints (n=307). In general, BestFirst algorithm descriptor based models performed better than the models obtained using all descriptors. According to the principle of parsimony, we discuss only models, which used the BestFirst algorithm for variable selection (Table 3 and 4; data describing the performance of the training set and 10 fold cross-validation of the training set are provided in the Supporting Information (SI-Table 2)). The BestFirst algorithm selected a set of descriptors for model generation as follows: 11 of 62 MOE 2D properties, 16 of 166 MACCS fingerprints and 19 of 307 substructure fingerprints. With each descriptor set four different classification models were developed using the ML techniques RF, SVM, kNN and BQSAR.

14

Table 3. Summary of machine-learning models based on BestFirst feature selection method with the internal test set.

Descriptors	Models	Confusion matrix				Sensitivity	Specificity	Accuracy	G-Mean
		TP	TN	FP	FN				
MOE[A]	RF	215	112	60	20	0.91	0.65	0.80	0.77
	SVM	219	109	63	16	0.93	0.63	0.81	0.77
	KNN	215	114	58	20	0.91	0.66	0.81	0.78
	BQSAR	196	120	52	39	0.83	0.70	0.78	0.76
MACCS[B]	RF	207	96	76	28	0.88	0.56	0.74	0.70
	SVM	199	75	97	36	0.85	0.44	0.67	0.61
	KNN	215	79	93	20	0.91	0.46	0.72	0.65
	BQSAR	158	117	55	77	0.67	0.68	0.68	0.68
SS-FP[C]	RF	215	73	99	20	0.91	0.42	0.71	0.62
	SVM	220	66	106	15	0.94	0.38	0.70	0.60
	KNN	220	67	105	15	0.94	0.39	0.71	0.60
	BQSAR	188	86	86	47	0.80	0.50	0.67	0.63
Combined[D]	RF	215	118	54	20	0.91	0.69	0.82	0.79
	SVM	219	106	66	16	0.93	0.62	0.80	0.76
	KNN	207	124	48	28	0.88	0.72	0.81	0.80
	BQSAR	193	118	54	42	0.82	0.69	0.76	0.75

Note: RF: Random Forest, SVM: Support vector machine, KNN: Kappa nearest neighbor, BQSAR: Binary QSAR. [A] Bestfirst descriptors from 2D-MOE, [B] Bestfirst descriptors from MACCS fingerprints, [C] Substructure fingerprints, [D] Bestfirst descriptors from all the calculated descriptors.

For the MOE 2D descriptors, all four models were able to correctly predict > 75 % of the compounds of the test set, whereupon the best model was obtained with kNN (MCC=0.61, Accuracy=81 %). Also regarding a balanced prediction, kNN performed best achieving a G-Mean value of 0.78. However, highly similar performance was observed with the methods random forest (MCC=0.60, Accuracy=0.80 %, G-Mean=0.77) and SVM (MCC=0.61, Accuracy=0.81 %, G-Mean=0.77). With 10-fold CV random forest outperformed kNN and SVM, showing an MCC value of 0.64 compared to 0.55 (SVM) and 0.61 (kNN) (Table 4).

Table 4. Matthews correlation coefficient of the Models for the internal test set predictions (10-fold cross-validations are provided in the parentheses).

Classification Methods	BestFirst Descriptor Models			
	MOE	MACCS	SS-FP	Combined
RF	0.60 (0.64)	0.47 (0.55)	0.40 (0.43)	0.63 (0.66)
SVM	0.61 (0.55)	0.31 (0.38)	0.40 (0.38)	0.59 (0.59)
KNN	0.61 (0.61)	0.43 (0.46)	0.40 (0.41)	0.61 (0.59)
BQSAR	0.54 (0.63)	0.35 (0.41)	0.32 (0.41)	0.51 (0.57)

Note: RF: Random Forest, SVM: Support vector machine, KNN: Kappa nearest neighbor, BQSAR: Binary QSAR

The models created using MACCS fingerprints showed that random forest performed better than the other machine learning methods, correctly predicting 74 % of the internal test set (MCC=0.47). Also the kNN model correctly predicted more than 70 % of the test compounds. However, the model suffers from a high false positive rate of more than 0.5. The MACCS fingerprints selected by the BestFirst algorithm comprised *bin-8* (QAA@1), *bin-17* (CTC), *bin-50* (C=C(C)), *bin-54* (QHAAQH), *bin-69* (QQH), *bin-75* (A!N$A), *bin-76* (C=C(A)A) and *bin-84* (NH2), *bin-86* (CH2QCH2), *bin-102* (QO), *bin-112* (AA(A)(A)A), *bin-125* (aromatic ring > 1), *bin-129* (ACH2AACH2A), *bin-139* (OH), *bin-145* (6M ring > 1), *bin-155* (A!CH2!A), *bin-162* (aromatic). In order to get insights into the distribution of the structural keys, the frequency of the fingerprint bins selected by BestFirst for inhibitors and non-inhibitors were compared. For instance, bins 50, 75, 86, 125, 129, 145, 155 and 162 are represented more often in the group of inhibitors, whereas bin 45, 84 and bin 139 are more prevalent in the set of non-inhibitors (SI-Table 3). The bins more prevalent in inhibitors were mainly of hydrophobic nature, as e.g. aromatic or ring substructures. On the other hand, bins more often hit by non-inhibitors represent hydrophilic substructures, comprising the number of heteroatoms, hydroxylic groups and primary amines. In Figure 3 an example of a phenylpyrazolon-type P-gp inhibitor with the matched MACCS fingerprints is depicted.
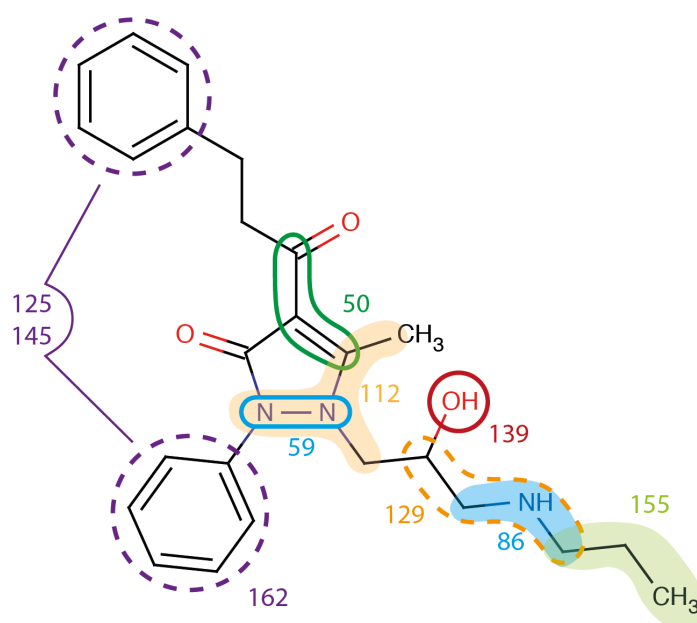
Figure 3. Schematic representation of occurrence of MACCS fingerprints in phenylpyrazolon-type P-gp inhibitor

Substructure/functional group fingerprints based models generally showed similar performance compared to the models developed from MACCS fingerprints. Analogous to the models generated on the basis of the MOE 2D descriptors, the methods RF, kNN and SVM showed extremely similar statistics. All three were able to correctly predict more than 70 % of the internal test set compounds, which resulted in an MCC of 0.4. However, the RF model showed a slightly better G-mean value due to a more balanced prediction. In a previous study using a similar dataset (the databases share over 80 % similarity) the application of an FP growth algorithm revealed that alkyl-aryl ethers (59 % of inhibitors in the dataset), tertiary aliphatic amines (51 %) and aromatic groups were significantly more present in inhibitors compared to non-inhibitors in

their dataset. Moreover, phenols (18 % of non-inhibitors in the dataset), carboxylic acids (11%) and primary amines (14%) were quite prevalent in non-inhibitors.[43]

The same groups have been selected by the BestFirst algorithm for model generation in this study, leading to a set of 16 fingerprints in total. Again inhibitors were mainly described by possessing alkylarylether and aromatic groups, as well as amines and secondary carbons. While also a high number of non-inhibitors bear aromatic groups, they stand out for their relatively high occurrence of carboxylic acids and phenols (Supplementary material, SI-Table 2).

Although on basis of overall accuracy the models developed from three different descriptor sets (2D, MACCS and SSFP) performed relatively good, the majority suffered from a high number of false positives. Thus, additional models have been built by using a BestFirst selected set of descriptors derived from combining all descriptors. A set of 15 descriptors was selected, which comprised 4 MACCS keys, 2 substructure fingerprints, and 9 MOE 2D descriptors (Table 6).

Table 6. Descriptors selected by the BestFirst algorithm for the combined descriptor set.

Descriptor set		Descripotr set	
MACCS	17 50 54 125	MOE 2D	a_hyd a_nC b_single logP(o/w)
Substructure fingerprint	84 90		b_rotR density
			logS vdw_area vsa_hyd

Considering the four subsequently generated models, the overall quality did not change significantly compared to the models developed using only BestFirst 2D descriptors, which showed the best statistics among the three descriptor sets. However,

18

the combined descriptor models exhibited slight improvements in the prediction of non-inhibitors, as can be noticed from increased G-Mean values (all ≥ 75 %). Additionally, the random forest model on the basis of the combined descriptor showed the best performance of all models developed in this study. Thus, it correctly predicted 215/235 inhibitors and 118/172 non-inhibitors from the internal test set, resulting in an overall accuracy and MCC of 82 % and 0.63, respectively.

Validation of external compounds. As mentioned in the method section, considering the 2D constitution, 429 compounds were present in both the Chen et al. and the Broccatelli et al. dataset, out of which 346 could be used as external test set. The models derived from the combined BestFirst descriptors were used to predict these 346 compounds. As can be seen from Table 7, SVM performed reasonably well, correctly predicting 97% of the inhibitors and 62% of the non-inhibitors, resulting in an overall accuracy of 77%. Also with RF a similarly good accuracy of 75% could be observed. Interestingly, the BQSAR and the kNN model showed better prediction of non-inhibitors, as can be seen from the comparably high specificity values of > 70.

Table 7. External test set predictions.

Descriptors	Models	Sensitivity	Specificity	Accuracy	G-Mean	MCC
MOE	RF	0.98	0.52	0.70	0.71	0.52
	SVM	0.99	0.51	0.69	0.71	0.52
	kNN	0.87	0.56	0.68	0.70	0.43
	B-QSAR	0.86	0.65	0.73	0.75	0.50
MACCS	RF	0.63	0.71	0.68	0.67	0.34
	SVM	0.27	0.93	0.68	0.50	0.28
	kNN	0.79	0.67	0.71	0.72	0.44
	B-QSAR	0.77	0.16	0.39	0.35	-0.09
SS-FP	RF	0.79	0.42	0.56	0.57	0.21
	SVM	0.94	0.24	0.51	0.48	0.23
	kNN	0.91	0.27	0.51	0.50	0.22
	B-QSAR	0.49	0.82	0.69	0.63	0.33
Combined	RF	0.99	0.57	0.73	0.75	0.57
	SVM	0.97	0.62	0.75	0.77	0.59
	kNN	0.64	0.72	0.69	0.68	0.36
	B-QSAR	0.58	0.74	0.68	0.66	0.32

Although the models developed from different sets of descriptors and fingerprints performed quite good, there remained always the question whether the classification of P-gp inhibitors and non-inhibitor can also be done using simple drug-likeness descriptors (molecular weight, hydrogen bond acceptor, hydrogen bond donor and LogP).[42] Thus, four models were developed using decision tree (DT), kNN, SVM and random forest as classifiers. All models were able to correctly identify > 85 % of inhibitors, but non-inhibitors were predicted relatively poor (Table 8). The DT method provides easy interpretation of the model, however, it often lacks predictability. Interestingly, in this case the DT approach showed good predictability, by exhibiting an overall accuracy of 75% with an MCC of 0.57 for the external test set. According to this

model, non-inhibitors possessed lower lipophilicity and lower molecular weight than inhibitors (see SI-Figure 2). This observation is in agreement with observations derived from the molecular property distribution plot and PCA analysis (Figure 1).

Table 8. Simplified Classification Models using Rule of Five descriptors. Statistics describe the classification performance on the external test set.

Models	Confusion matrix				Sensitivity	Specificity	MCC	Accuracy
	TP	TN	FP	FN				
RF	127	117	97	5	0.96	0.55	0.52	0.71
SVM	131	90	124	1	0.99	0.42	0.46	0.64
KNN	117	102	112	15	0.89	0.48	0.37	0.63
DT	126	132	82	6	0.95	0.62	0.57	0.75

Note: RF: Random Forest, SVM: Support vector machine, KNN: Kappa nearest neighbor, Dt: Decision tree, MCC: Matthews's correlation coefficient.

In order to understand the reason for a certain model output, selected compounds that were incorrectly classified by the combined descriptor-set models were analyzed. The results showed that 32 compounds (8 % of the internal test set) were misclassified by all four methods (SVM, kNN, BQSAR and RF), and more than 63 % of the internal test set compounds were correctly classified by all methods. Among the 32 compounds, which were consistently false classified, 7 were misclassified as non-inhibitors (FN), and 25 as inhibitors (FP). Representative examples are given in Figure 4. Cefotaxime was misclassified as non-inhibitor by all for methods with a probability score of > 0.9. Analyzing the molecular properties of Cefotaxime showed that the values for logP (-0.6), logS (-3.7) and MR (10.7) lie in the region for non-inhibitors (Table 2). Interestingly, with this polar compound a strong inhibition of 82 % (normalized to the inhibitory activity of cyclosporine A) in a Calcein-AM efflux assay was observed.[44] Similarly, Clindamycin, the physicochemical properties of which are rather located in the range of non-inhibitors, showed a > 80 % in the same assay.[44] On the other hand the laxative Bisacodyl was misclassified as inhibitor by all methods. Regarding its physicochemical

properties, the hydrophobic drug (logP = 4.2) exhibits values similar to P-gp inhibitors (low logs, high MR and molecular weight). As a matter of fact, taking a look at the assay data, Bisacodyl was not clearly identified as P-gp non-inhibitor, but rather showed inconclusive data.[44] This might suggest a weak P-gp inhibitory activity under certain circumstances. Digoxin also exhibits inhibitor-like physicochemical properties, although it is not known for inhibiting the efflux pump. On the contrary, digoxin was identified as P-gp substrate and thus might still act as a competitive inhibitor in e.g. a rhodamine efflux assay. As has recently been demonstrated, the respective assay and assay conditions are of utmost importance for the assessment of P-gp inhibitory activity (REF Zdrazil et al, Mol Inform). Thus, a more detailed analysis of the respective biological data might reveal other examples where the prediction might be in line with biological data.
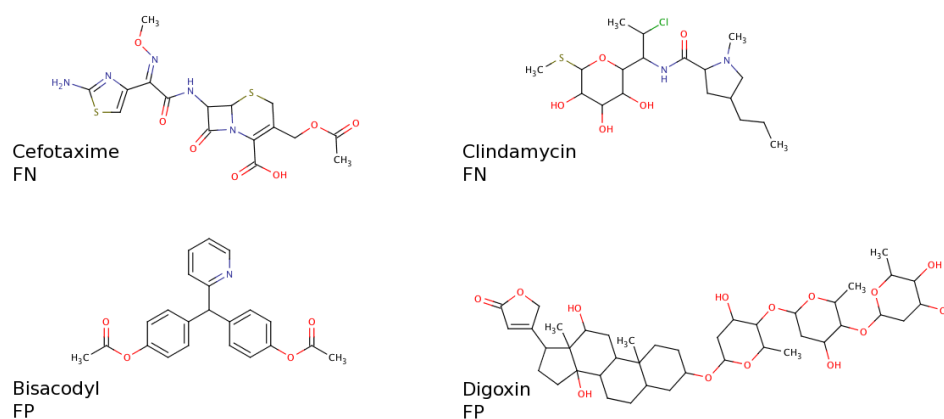
Figure 4. Examples of misclassified compounds in the test set

Probability of prediction of inhibitors and non-inhibitors. Each classification model gives the probability of a compound for belonging to a certain class. In that sense, inhibitors would be ideally characterized by a probability of 1, whereas non-inhibitors should be assigned 0. Thus, a ranking of the database compounds according to their

22

probability for being an inhibitor was performed. In Figure 5, the number of TP, FP, TN and FN were plotted against their probabilities based on the combined RF model. As can be seen in the figure, compounds with an inhibitor probability of > 0.6 are more likely to be true positives. On the other hand, it is highly probable to identify a true negative if its probability value for being an inhibitor is < 0.4. Based on that analysis, statements on compounds with probabilities ranging from 0.4 to 0.6 should be treated carefully.
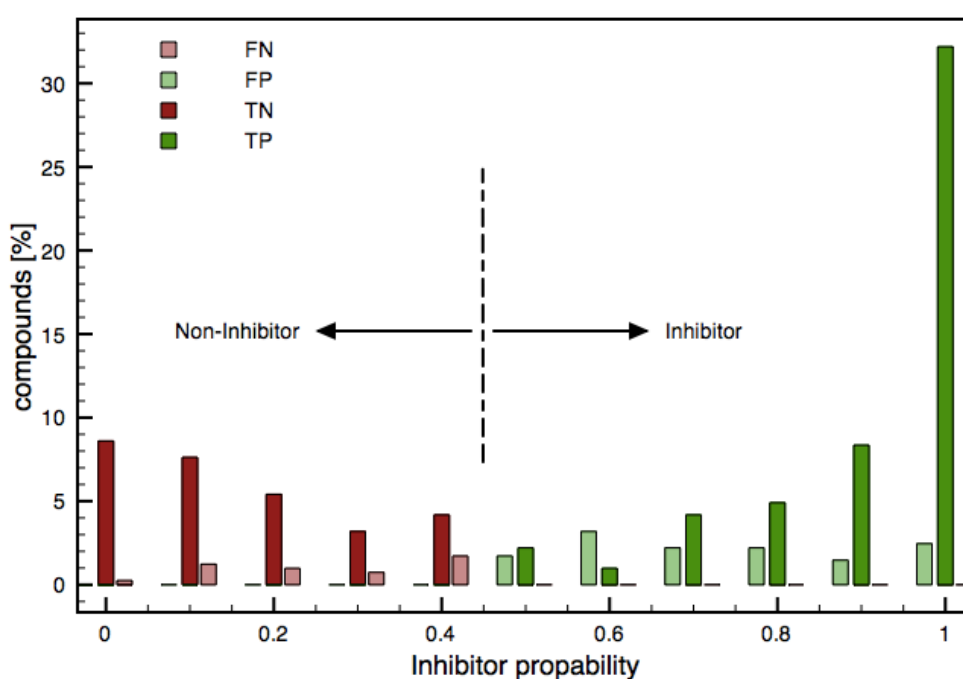
Figure 5. Probability score of P-gp inhibitors and non-inhibitors in the test set

   Applicability Domain. In addition to the classification models, an applicability domain (AD) experiment was performed to check the reliability of the developed compound prediction models. AD provides a first structural alert on the dataset and is primarily used to check whether a new molecular entity (NME) is within the chemical space of the training set or not. There have been many AD approaches proposed and each had their own pros and cons. Some of the well-known AD approaches are

23

descriptor "ranges", "Euclidean distances (ED)" and "probability density (PD)".[45, 46] In the present study, AD analysis was performed on the basis of the ED approach using the Ambit Discovery software (Version 0.04).[47] Ambit Discovery preprocesses the given dataset by principal component analysis in order to eliminate collinearities among the descriptors. Subsequently, the AD is estimated based on the ED approach for the internal and the external test set. The results showed all compounds from both datasets were found to be inside the chemical domain of the training compounds. An additional AD experiment was performed using a set of 986 FDA approved drugs from DrugBank. It had been observed that 973 compounds were predicted to be in the domain of the training set and only a small amount (13 compounds) of the FDA drugs were located outside. Some of these compounds were found to be peptides, e.g. bacitracin. colistimethate and degarelix, while others contained transition metals, as cisplatin or organoplatin. Thus, the models should be valid over a broad range of the druglike chemical space. The scoring plot of the first two principal components obtained by PCA of the FDA approved drugs, as well as the training and both test sets is provided in the Supporting Information (SI-Figure 3).

   Structure-based classification. The publication of the mouse P-gp structure paved the way for structure-based studies. We recently showed that docking into a homology model of human P-gp based on this X-ray structure lead to poses consistent with QSAR data[33] and that these poses can successfully be exploited for identification of new P-gp inhibitors.[48] We thus used our homology model in a wider setting and docked a large set of P-gp inhibitors and non-inhibitors into P-gp in order to investigate the possibility to use docking for classification purposes.

   The scoring values of the different fitness functions were binned into ranges of 0.5 or 5, respectively, and plotted against the occurrence of inhibitors and non-inhibitors. This

24

resulted in two distinct curves for each scoring function (Figure 6a). The intersection point of those curves was used as classification criteria, and compounds scored higher than this point were considered as inhibitors. Vice versa, compounds showing lower scores were classified as non-inhibitors. According to this criterion, the confusion matrix parameters, as well as other performance measures could be calculated and are summarized in Table 8. As can be seen from the results, the CS docking runs showed highest MCC values, especially when also using ChemScore as discrimination criteria. On the other hand, GS docking runs performed best when being rescored with the external fitness function Xscore. Rescoring CS docking results with GoldScore resulted in a dramatic decrease of performance. The different ligand protonation settings did not considerably affect the outcome. However ChemScore docking runs showed a slight preference for neutral ligands, while ChemPLP and ASP performed better with protonated ligands.
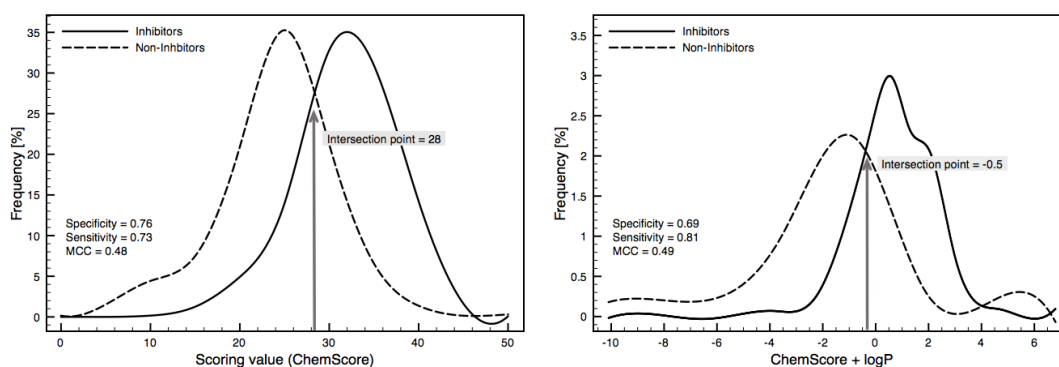
Figure 6.   Distribution of P-gp Inhibitors and Non-inhibitors based on ChemScore scoring. Sensitivity, specificity and MCC were calculated from true and misclassification rate at intersection point of two curves. A) Distribution based on ChemScore alone; B) distribution based on a combined ChemScore-logP score.

25

With an MCC of 0.5 the CS docking run using neutral ligands performed best. Overall, this model accurately predicted 75 % of the dataset compounds. According to this model, all compounds that had a ChemScore value above 28 were predicted as inhibitors. The obtained ChemScore values for non-inhibitors ranged from 0 to 40, whereas in the case of inhibitors the values varied from 10 to 50.

Table 8. Summary of Models obtained using different scoring functions

	Ligand Protonation	Scoring Function	Intersection Point	Sensitivity	Specificity	Accuracy	G-Mean	MCC
CS docking run	Neutral	Chemscore	28	0.76	0.73	0.75	0.75	0.48
		Goldscore	25	0.36	0.66	0.46	0.49	0.02
		ASP	25	0.76	0.62	0.71	0.68	0.36
		ChemPLP	50	0.66	0.69	0.67	0.68	0.34
		XScore	6	0.63	0.78	0.68	0.70	0.38
	Charged	Chemscore	30	0.68	0.79	0.71	0.73	0.44
		Goldscore	18	0.50	0.42	0.47	0.46	-0.08
		ASP	29	0.60	0.79	0.67	0.69	0.38
		ChemPLP	50	0.70	0.68	0.70	0.69	0.37
		XScore	6	0.68	0.73	0.70	0.71	0.39
GS docking run	Neutral	Chemscore	22	0.73	0.58	0.68	0.65	0.31
		Goldscore	45	0.61	0.75	0.66	0.67	0.34
		ASP	25	0.72	0.56	0.67	0.64	0.28
		ChemPLP	50	0.65	0.66	0.65	0.65	0.29
		XScore	6	0.65	0.73	0.68	0.69	0.36
	Charged	Chemscore	25	0.59	0.75	0.64	0.66	0.32
		Goldscore	45	0.71	0.63	0.68	0.67	0.33
		ASP	25	0.74	0.57	0.68	0.65	0.31
		ChemPLP	50	0.70	0.65	0.68	0.67	0.33
		XScore	6	0.68	0.71	0.69	0.69	0.37
combined Score (ChemScore + logP)			-0.5	0.81	0.69	0.77	0.75	0.49

In terms of MCC and balanced prediction, the overall classification based on docking (ChemScore) was less accurate compared to models developed using SVM or random forest. Nevertheless, by correctly predicting 75 % of the training and internal test set the docking model proved to be suited for classification studies.

To improve the statistics of the structure-based classification, we tried to implement the fact that the compounds access the protein's binding site via the lipid bilayer. In that

sense a score was calculated, combining ChemScore and the descriptor logP(o/w) (each normalized) (Figure 6b). This resulted in slightly improved statistics, as with this combined score 77 % of the compounds could be predicted correctly, resulting in an MCC value of 0.49.

Subsequently, the ChemScore and the combined ChemScore-logP scoring models were applied on the external test set. As expected, the performance of the external validation was lower than for the dataset consisting of training and internal test set. Both models showed high sensitivity values (< 0.9), indicating that there were hardly any FN classified. However, high FP rates could be observed, which resulted in low specificity (< 0.4) for both models. The model only relying on the ChemScore value was able to correctly classify 54 % of the external test set compounds, while the model that was also taking the logP into account predicted 61 % of the compounds correctly. This indicates that implementing information about the entry path of the molecules clearly increases a structure-based prediction model

However, beside statistical performance measures like MCC, docking poses can also be evaluated by comparing them with experimental data. For the P-gp substrate verapamil, the following interacting residues on TM helices 1, 4, 5, 6, 7, 10, 11 and 12 are known: His61, Ala64, Leu65, Ser222, Ile306, Leu339, Ala342, Phe728, Ile868, Gly872, Phe942, Thr945, Leu975, Gly984, Val982.[33] Figure 7 shows the top view of the P-gp binding pocket with four different docking poses of verapamil, generated by the docking runs presented in Table 8. As can be noticed, all four poses are in vicinity of the experimentally derived interacting amino acid residues, which are rendered blue. Interestingly, the pose, which resulted from the GS docking run with neutral ligands (Figure 6-B), is located closest to most of the residues, while the corresponding CS

docking run is placed only in vicinity to the residues on TM helices 4, 5 and 6. This suggests that the GS docking run could better reproduce the experimental data.
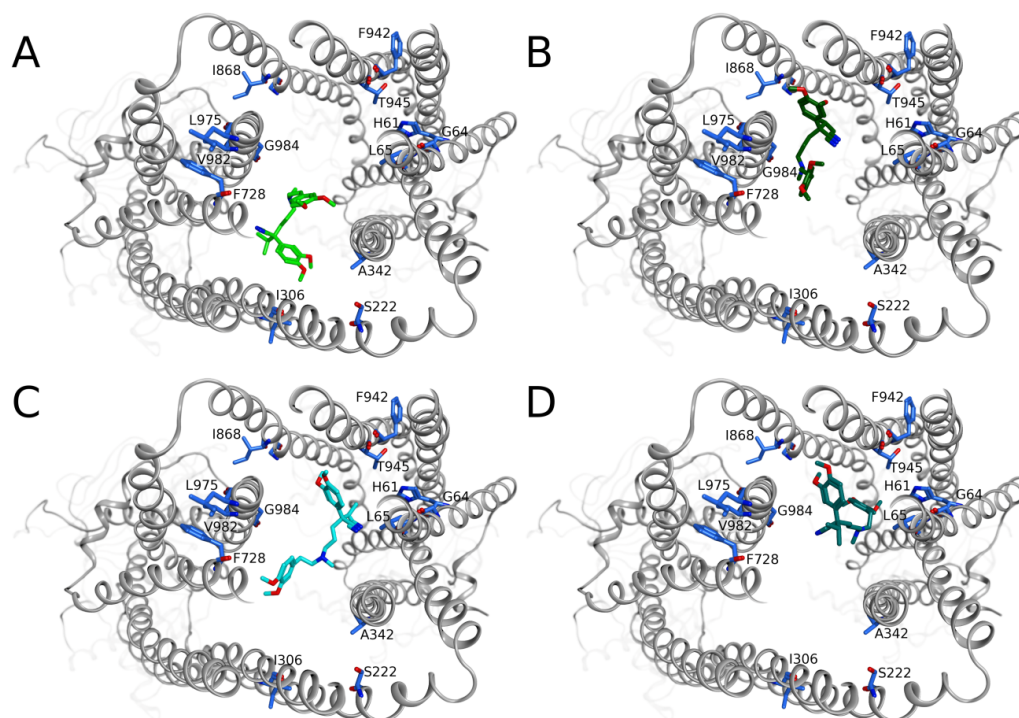
Figure 7. Verapamil docking poses generated by 4 different docking runs. A) CS docking with neutral ligand, B) CS docking run with positively charged ligand, C) GS docking run with neutral ligand, D) GS docking run with positively charged ligand.

Protein-Ligand Interactions. Predicting protein-ligand interactions is a highly valuable analysis tool that helps to examine energetically favourable conformations or orientations of ligands in the protein active site. The P̲rotein L̲igand I̲nteraction F̲ingerprints (PLIF) tool, implemented in MOE, computes different molecular interactions between residues of the binding site and the corresponding ligand conformation. In this study, PLIF analysis was performed on the basis of the docking poses generated by the CS docking run. As can be seen in Figure 8, there was no

significant difference in interaction occurrences between inhibitors and non-inhibitors. However, non-inhibitors showed a higher number of H-bond interactions than inhibitors. This is in agreement with the results of the ML techniques, as they described non-inhibitors being more hydrophilic than inhibitors. This also reflects the mode of action of P-gp, as its substrates have to cross the lipid bilayer in order to reach the protein's active site. The main interacting amino acid residues are the same for inhibitors are Phe303, Tyr307, Phe336 and Phe343, which are located on helices 5 and 6 of TM domain 1 (Figure 2). The low involvement of the corresponding helices in TM domain 2 might be traced back to a certain asymmetry of the crystal structure template, which was already pointed out in our previous study.[33]
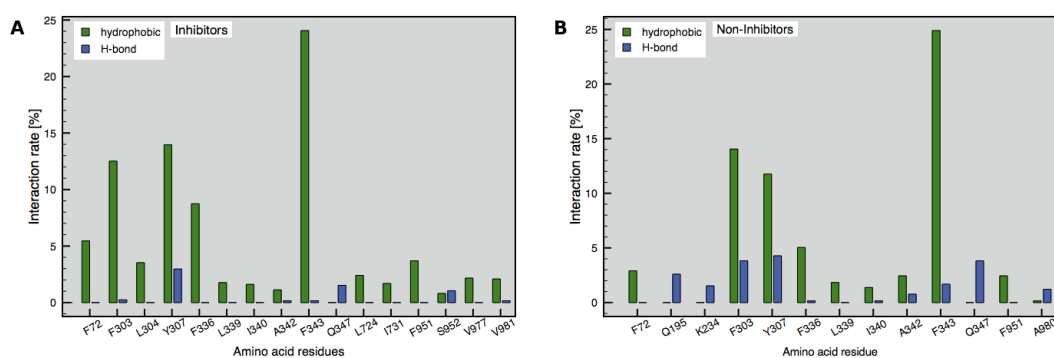
Figure 8. PLIF analysis. Important residues are shown with their hydrophobic and hydrogen bonding interactions. A) inhibitors, B) non-inhibitors

CONCLUSIONS

In the present study, we developed structure and ligand-based classification models from a set of 1608 P-gp inhibitors and non-inhibitors. It could be observed that molecular properties, which are overwhelmingly involved in P-glycoprotein inhibition, are hydrophobic parameters, such as logP and molar refractivity. Various descriptor contributions, as well as the PLIF analysis, point towards distinct membrane

permeabilities of inhibitors and non-inhibitors, as the former are for instance considerably less water-soluble and bulkier. Models obtained by support vector machine and random forest in combination with BestFirst descriptors performed better than other ML models. Both models were performing good in discriminating inhibitors (> 90 % of the internal and external test sets) from non-inhibitors, with overall accuracies of SVM and RF of 83/75 and 86/73 %, respectively. The structure-assisted docking model that used the GOLD implemented ChemScore scoring function predicted reasonably well P-gp inhibition (accuracy=0.75, MCC=0.48). The model was able to correctly predict 76% of P-gp inhibitors and 73% of non-inhibitors. Adding the logP-value of the compounds to their docking score showed that implementing information about membrane diffusion of P-gp inhibitors could slightly improve the prediction (accuracy =0.77, MCC (0.49). However, structure-based classification on the external test set was less satisfying, showing an overall accuracy of only 0.61 (ChemScore + logP). Nevertheless, also in this case the method was highly predictive for inhibitors (97 % correctly classified), but lacked specificity.

Furthermore, an AD experiment using the BestFirst selection of the combined descriptor sets suggested that these models are applicable to predict drug-like compound libraries.

The models built in this study from ligand-based methods were efficient and precise and thus could be in future used for the identification of P-gp inhibitors or non-inhibitors from virtual screenings of large compound libraries. As the analysis of simple properties demonstrated good sensitivity values, these descriptors could be used as pre-filter for screening, which reduces the number of compounds to be considered for random forest or SVM. Structure-based classification lacked overall accuracy for the external test set, but might be useful in combination with ML techniques. Positively

classified compounds of the latter could thus be processed by docking methods, e.g., ChemScore, which would further narrow the screening process for the identification of potent P-gp inhibitors. In addition, the PLIF analysis provided molecular protein-ligand interaction information, which may play a key role in optimization of ligands for P-gp inhibition. In conclusion a workflow comprising pre-screening with simple descriptors, classification by ML techniques and post-processing by structure-based methods provides accurate prediction with information for further drug development.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

All materials available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1.   Higgins, C. F. ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* **1992,** *8*, 67-113.

2.   Dean, M.; Hamon, Y.; Chimini, G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* **2001,** *42* (7), 1007-17.

3.   Borst, P.; Elferink, R. O. Mammalian ABC transporters in health and disease. *Annu Rev Biochem* **2002,** *71*, 537-92.

4.   Cummins, C. L.; Salphati, L.; Reid, M. J.; Benet, L. Z. In vivo modulation of intestinal CYP3A metabolism by P-glycoprotein: studies using the rat single-pass intestinal perfusion model. *J Pharmacol Exp Ther* **2003,** *305* (1), 306-14.

5.   Sugano, K.; Kansy, M.; Artursson, P.; Avdeef, A.; Bendels, S.; Di, L.; Ecker, G. F.; Faller, B.; Fischer, H.; Gerebtzoff, G.; Lennernaes, H.; Senner, F. Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov* **2010,** *9* (8), 597-614.

6.   Szakacs, G.; Paterson, J. K.; Ludwig, J. A.; Booth-Genthe, C.; Gottesman, M. M. Targeting multidrug resistance in cancer. *Nat Rev Drug Discov* **2006,** *5* (3), 219-34.

7.   Ecker, G. F.; Stockner, T.; Chiba, P. Computational models for prediction of interactions with ABC-transporters. *Drug Discov Today* **2008,** *13* (7-8), 311-7.

8.   Demel, M. A.; Kraemer, O.; Ettmayer, P.; Haaksma, E.; Ecker, G. F. Ensemble Rule-Based Classification of Substrates of the Human ABC-Transporter ABCB1 Using Simple Physicochemical Descriptors. *Molecular Informatics* **2010,** *29* (3), 233-242.

9.   Cianchetta, G.; Singleton, R. W.; Zhang, M.; Wildgoose, M.; Giesing, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. A pharmacophore hypothesis for P-glycoprotein substrate recognition using GRIND-based 3D-QSAR. *J Med Chem* **2005,** *48* (8), 2927-35.

10.   Langer, T.; Eder, M.; Hoffmann, R. D.; Chiba, P.; Ecker, G. F. Lead identification for modulators of multidrug resistance based on in silico screening with a pharmacophoric feature model. *Arch Pharm (Weinheim)* **2004,** *337* (6), 317-27.

11.   Pearce, H. L.; Safa, A. R.; Bach, N. J.; Winter, M. A.; Cirtain, M. C.; Beck, W. T. Essential features of the P-glycoprotein pharmacophore as defined by a series of reserpine analogs that modulate multidrug resistance. *Proc Natl Acad Sci U S A* **1989,** *86* (13), 5128-32.

12.   Sakiyama, Y. The use of machine learning and nonlinear statistical tools for ADME prediction. *Expert Opin Drug Metab Toxicol* **2009,** *5* (2), 149-69.

13.   Wang, Y. H.; Li, Y.; Yang, S. L.; Yang, L. Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. *J Chem Inf Model* **2005,** *45* (3), 750-7.

14.   Broccatelli, F.; Carosati, E.; Neri, A.; Frosini, M.; Goracci, L.; Oprea, T. I.; Cruciani, G. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. *J Med Chem* **2011,** *54* (6), 1740-51.

15.   Chen, L.; Li, Y.; Zhao, Q.; Peng, H.; Hou, T. ADME Evaluation in Drug Discovery. 10. Predictions of P-Glycoprotein Inhibitors Using Recursive Partitioning and Naive Bayesian Classification Techniques. *Mol Pharm* **2011,** *8* (3), 889-900.

16.   Aller, S. G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I. L.; Chang, G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **2009,** *323* (5922), 1718-22.

17.   Klepsch, F.; Ecker, G. Impact of the recent mouse P-glycoprotein structure for structure-based ligand design. *Molecular Informatics* **2010,** *29*, 276-286.

18.   Winter, S. S.; Lovato, D. M.; Khawaja, H. M.; Edwards, B. S.; Steele, I. D.; Young, S. M.; Oprea, T. I.; Sklar, L. A.; Larson, R. S. High-throughput screening for daunorubicin-mediated drug resistance identifies mometasone furoate as a novel ABCB1-reversal agent. *Journal of Biomolecular Screening* **2008,** *13* (3), 185-193.

19.   Bikadi, Z.; Hazai, I.; Malik, D.; Jemnitz, K.; Veres, Z.; Hari, P.; Ni, Z.; Loo, T. W.; Clarke, D. M.; Hazai, E.; Mao, Q. Predicting P-glycoprotein-mediated drug transport based on support vector machine and three-dimensional crystal structure of P-glycoprotein. *PLoS One* **2011,** *6* (10), e25815.

20.   Blower, P. E.; Yang, C.; Fligner, M. A.; Verducci, J. S.; Yu, L.; Richman, S.; Weinstein, J. N. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J* **2002,** *2* (4), 259-71.

21.   Dolghih, E.; Bryant, C.; Renslo, A. R.; Jacobson, M. P. Predicting binding to p-glycoprotein by flexible receptor docking. *PLoS Comput Biol* **2011,** *7* (6), e1002083.

22.   Chen, L.; Li, Y.; Yu, H.; Zhang, L.; Hou, T. Computational models for predicting substrates or inhibitors of P-glycoprotein. *Drug Discov Today* **2012,** *17* (7-8), 343-51.

23.   Rautio, J.; Humphreys, J. E.; Webster, L. O.; Balakrishnan, A.; Keogh, J. P.; Kunta, J. R.; Serabjit-Singh, C. J.; Polli, J. W. In vitro p-glycoprotein inhibition assays for assessment of clinical drug interaction potential of new drug candidates: a recommendation for probe substrates. *Drug Metab Dispos* **2006,** *34* (5), 786-92.

24.   MOE (Molecular Operating Environment), Version 2009.10, Chemical Computing Group Inc, 1010 Sherbrooke Street West, Suite 910, Montreal, Canada. http://www.chemcomp.com/.

25.   Yap, C. W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* **2011,** *32* (7), 1466-74.

26.   *SIMCA-P+ (version 10.5) and MODDE (Version 7.0), Umetrics, Umeå, Sweden (http://www.umetrics.com).*

27.   Olsson, I. M.; Gottfries, J.; Wold, S. D-optimal onion designs in statistical molecular design. *Chemometrics and Intelligent Laboratory Systems* **2004,** *73* (1), 37-46.

28.   Kriegl, J. M.; Eriksson, L.; Arnhold, T.; Beck, B.; Johansson, E.; Fox, T. Multivariate modeling of cytochrome P450 3A4 inhibition. *European Journal of Pharmaceutical Sciences* **2005,** *24* (5), 451-463.

29.   Witten, I.; Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques*. 2 ed.; Morgan Kaufmann: San Francisco, 2005.

30.   Fox, T.; Kriegl, J. M. Machine learning techniques for in silico modeling of drug metabolism. *Curr Top Med Chem* **2006,** *6* (15), 1579-91.

31.   Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. The WEKA Data Mining Software: An Update: SIGKDD Explorations. **2009,** *11* (1).

32.   Vasanthanathan, P.; Taboureau, O.; Oostenbrink, C.; Vermeulen, N. P.; Olsen, L.; Jorgensen, F. S. Classification of cytochrome P450 1A2 inhibitors and noninhibitors by machine learning techniques. *Drug Metab Dispos* **2009,** *37* (3), 658-64.

33.   Klepsch, F.; Chiba, P.; Ecker, G. F. Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of p-glycoprotein. *PLoS Comput Biol* **2011,** *7* (5), e1002036.

34.   Schrödinger Suite, L., New York, NY, 2011.

35.   Ecker, G.; Huber, M.; Schmid, D.; Chiba, P. The importance of a nitrogen atom in modulators of multidrug resistance. *Mol Pharmacol* **1999,** *56* (4), 791-6.

36.   Parveen, Z.; Stockner, T.; Bentele, C.; Pferschy, S.; Kraupp, M.; Freissmuth, M.; Ecker, G. F.; Chiba, P. Molecular dissection of dual pseudosymmetric solute translocation pathways in human P-glycoprotein. *Mol Pharmacol* **2011,** *79* (3), 443-52.

37.   Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **2005,** *48* (20), 6504-15.

38.   Mooij, W. T.; Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins* **2005,** *61* (2), 272-87.

39.   Korb, O.; Stutzle, T.; Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **2009,** *49* (1), 84-96.

40.   Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **2002,** *16* (1), 11-26.

41.   Gatlik-Landwojtowicz, E.; Aanismaa, P.; Seelig, A. Quantification and characterization of P-glycoprotein-substrate interactions. *Biochemistry* **2006,** *45* (9), 3020-3032.

42.   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 3-26.

43.   Poongavanam, V.; Haider, N.; Ecker, G. F. Fingerprint-based in silico models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg Med Chem* **2012**.

44.   PDSP pdsp.med.unc.edu/indexR.html (accessed 23 July 2012).

45.   Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *Atla-Alternatives to Laboratory Animals* **2005,** *33* (5), 445-459.

46.   Weaver, S.; Gleeson, M. P. The importance of the domain of applicability in QSAR modeling. *J Mol Graph Model* **2008,** *26* (8), 1315-26.

47.   Ambit   Discovery,   Ideaconsult   Ltd.,   Sofia,   Bulgaria (http://ambit.sourceforge.net/download_ambitdiscovery.html).

48.   Klepsch, F.; Prokes, K.; Parveen, Z.; Chiba, P.; Ecker, G. F. Structure-based pharmacophore screening for new P-gp inhibitors. *Abstr Pap Am Chem S* **2012,** *243*.

Ligand and structure-based classification models for Prediction of P-glycoprotein inhibitors

Poongavanam Vasanthanathan, Freya Klepsch, Gerhard F Ecker *

Department of Medicinal Chemistry, University of Vienna, Althanstraße 14, A-1090, Vienna, Austria.

Email: gerhard.f.ecker@univie.ac.at

Supplementary material

1

SI-Table 1:   List of descriptors used for PCA.

SI-Table 2:   Results of training set prediction and 10 fold cross-validation.

SI-Table 3:   List of MACCS fingerprints that contributed to the P-gp Inhibitor models.

SI-Table 4:   List of substructure fingerprints that contributed to the P-gp Inhibitor models.

SI-Figure 1:   Distribution plots physicochemical properties.

SI-Figure 2:   Decision tree model.

SI-Figure 3:   Applicability domain experiment using ED approach for the training and test sets, as well as for the FDA approved drugs.

2

SI-Table1: List of physicochemical properties used for PCA

Properties	Description		
logP(o/w)	Log of the partition coefficient (octanol/water)	a_count	Number of atoms
SMR	Molar refractivity (including implicit hydrogens)	a_heavy	Number of heavy atoms
TPSA	Total polar surface area	a_hyd	Number of hydrophobic atoms
weight	Molecular weight	a_nC	Number of carbon atoms
apol	Sum of the atomic polarizabilities	a_nH	Number of hydrogen atoms
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms	a_nO	Number of oxygen atoms
density	Density	b_ar	Number of aromatic bonds
logs	Log of the solubility in water	b_count	Number of bonds
MR	Molar refractivity	b_double	Number of double bonds
pmi	Principal moment of inertia	b_heavy	Number of bonds between heavy atoms
pmiX	x component of pmi	b_single	Number of single bonds
pmiY	y component of pmi	diameter	Diameter
pmiZ	z component of pmi	lip_acc	Number of O and N atoms
vdw_area	Area of van der Waals surface	lip_don	Number of OH and NH atoms
vdw_vol	Van der Waals volume	radius	Radius
vol	Volume		

SI-Table 2: Results of training set prediction and 10 fold cross-validation

Validation	Descriptors	Models	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	G-Mean	MCC
Training Set	MOE	RF	840	355	5	1	1.00	0.99	1.00	0.99	0.99
		SVM	787	204	156	54	0.94	0.57	0.83	0.73	0.56
		KNN	841	360	0	0	1.00	1.00	1.00	1.00	1.00
		B-QSAR	751	270	90	90	0.89	0.75	0.85	0.82	0.64
	MACCS	RF	799	241	119	42	0.95	0.67	0.87	0.80	0.67
		SVM	785	142	218	56	0.93	0.39	0.77	0.61	0.40
		KNN	809	231	129	32	0.96	0.64	0.87	0.79	0.67
		B-QSAR	616	255	105	225	0.73	0.71	0.73	0.72	0.41
	SS-FP	RF	812	164	196	29	0.97	0.46	0.81	0.66	0.53
		SVM	807	141	219	34	0.96	0.39	0.79	0.61	0.46
		KNN	816	160	200	25	0.97	0.44	0.81	0.66	0.53
		B-QSAR	725	191	169	116	0.86	0.53	0.76	0.68	0.41
	Combined	RF	841	356	4	0	1.00	0.99	1.00	0.99	0.99
		SVM	788	223	137	53	0.94	0.62	0.84	0.76	0.61
		KNN	841	360	0	0	1.00	1.00	1.00	1.00	1.00
		B-QSAR	732	261	109	99	0.88	0.71	0.83	0.79	0.59
Trainig	MOE	RF	780	245	115	61	0.93	0.68	0.85	0.79	0.64
		SVM	788	199	161	53	0.94	0.55	0.82	0.72	0.55
		KNN	746	258	102	95	0.89	0.72	0.84	0.80	0.61
		B-QSAR	747	266	94	94	0.89	0.74	0.84	0.81	0.63

4

154

Fingerprint	Method								
MACCS	RF	751	197	163	90	0.89	0.55	0.79	0.47
	SVM	782	137	223	59	0.93	0.38	0.77	0.38
	KNN	770	178	182	71	0.92	0.49	0.79	0.46
	B-QSAR	616	255	105	225	0.73	0.71	0.73	0.41
SS-FP	RF	795	142	218	46	0.95	0.39	0.78	0.43
	SVM	805	140	220	36	0.96	0.39	0.79	0.45
	KNN	800	130	230	41	0.95	0.36	0.77	0.41
	B-QSAR	725	191	169	116	0.86	0.53	0.76	0.41
Combined	RF	780	253	107	61	0.93	0.70	0.86	0.66
	SVM	786	216	144	55	0.93	0.60	0.83	0.59
	KNN	744	251	109	97	0.88	0.70	0.83	0.75
	B-QSAR	727	256	104	114	0.86	0.71	0.82	0.57

5

SI-Table 3: List of MACCS fingerprints that contributed to the MACCS FP based models and their frequency of occurrences in inhibitor and non-inhibitor

MACCS key		Description	Inhibitors [%]	Non-Inhibitors [%]
8	OAA@1	a 4-membered heterocycle	0.19	2.82
17	CTC	two carbon atoms connected via a triple bond	0.37	1.50
50	C=C(C)	propene	93.03	74.44
54	QHAAQH	two hetero atoms linked via two atoms	5.48	21.24
69	QQH	two successive hetero atoms	1.12	9.77
75	A!N$A	piperidine nitrogen	70.17	44.55
76	C=C(A)A	disubstituted ethene	98.51	90.60
84	NH2	primary amine	3.62	17.48
86	CH2QCH2	hetero atom connected to two aliphatic carbon atoms	72.03	40.04
102	QO	hetero atom bound to oxygen	2.70	11.28
112	AA(A)(A)A	branched substructure of 5 atoms of any type	99.63	94.36
125	aromatic ring > 1	more than one aromatic ring	84.85	49.44
129	ACH2AACH2A	chain of 6 atoms with the second and the fourth being aliphatic carbon atoms	75.74	48.50
139	OH	hydroxyl group	27.97	50.75
145	6M ring > 1	more than one 6-membered ring	92.94	70.11
155	A!CH2!A	an aliphatic carbon atom connected with two atoms of any type	94.70	79.70
162	AROMATIC	at least one aromatic ring	94.70	81.58

6

SI-Table 4: List of substructure fingerprints that contributed to the substructure FP based models and their frequency of occurrences in inhibitor and non-inhibitor

Substructure Fingerprint	Description	Inhibitor [%]	Non-Inhibitor [%]
SubFP2	secondary carbon	78.53	55.26
SubFP6	alkine	0.37	1.50
SubFP18	alkylarylether	48.42	24.44
SubFP23	amine	60.59	32.33
SubFP41	1,2-diol	0.56	5.83
SubFP84	carboxylic acid	0.93	13.91
SubFP90	carbothioic S ester	0.09	0.38
SubFP128	peptide C term	0.09	3.01
SubFP151	guanidine	0.09	1.88
SubFP169	phenol	6.13	15.23
SubFP170	1,2-diphenol	0.19	3.01
SubFP172	arylfluoride	7.53	2.26
SubFP214	sulfonic derivative	1.21	6.77
SubFP274	aromatic	94.70	81.58
SubFP287	conjugated double bond	96.84	87.41
SubFP302	rotatable bond	98.88	91.54

7
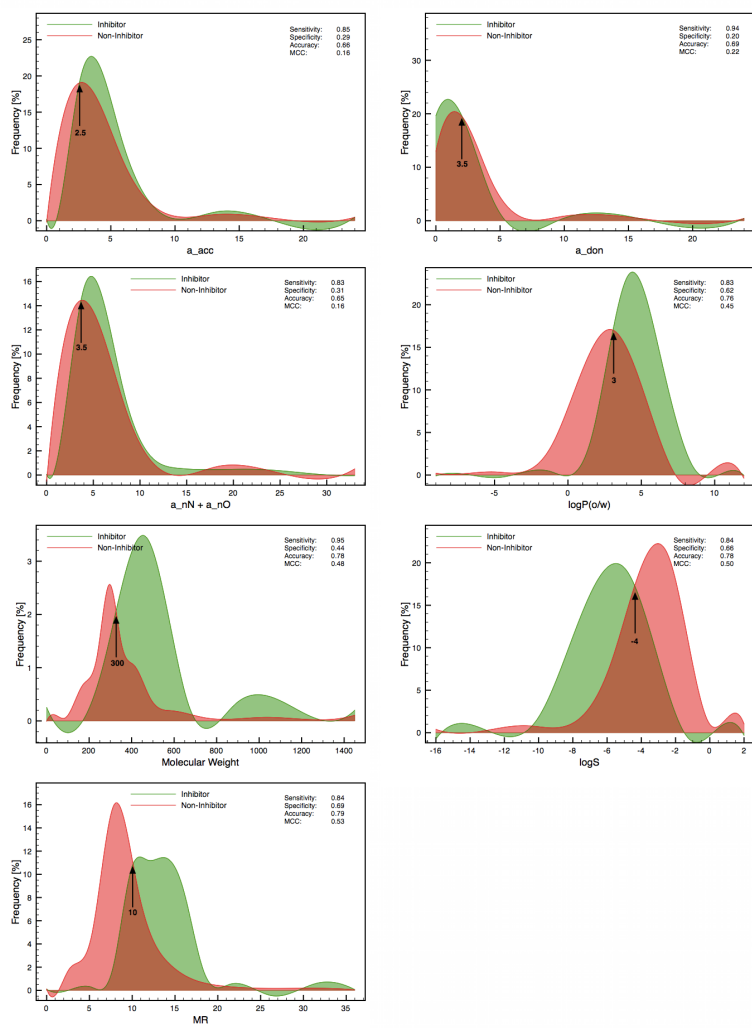
SI-Figure 2: Distribution plots from physicochemical properties
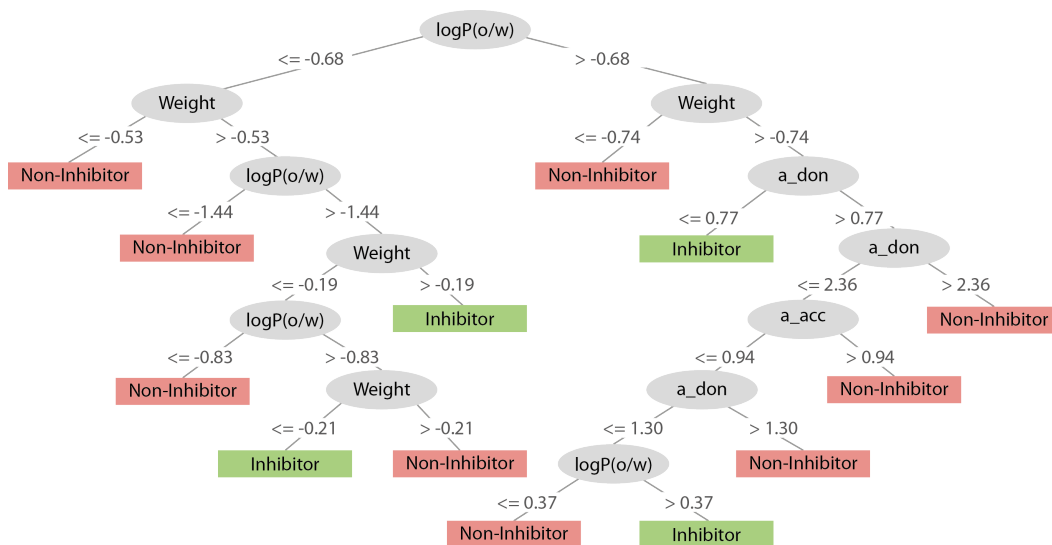
SI-Figure 3: Decision tree generated by using "Rule-of-Five" descriptors.
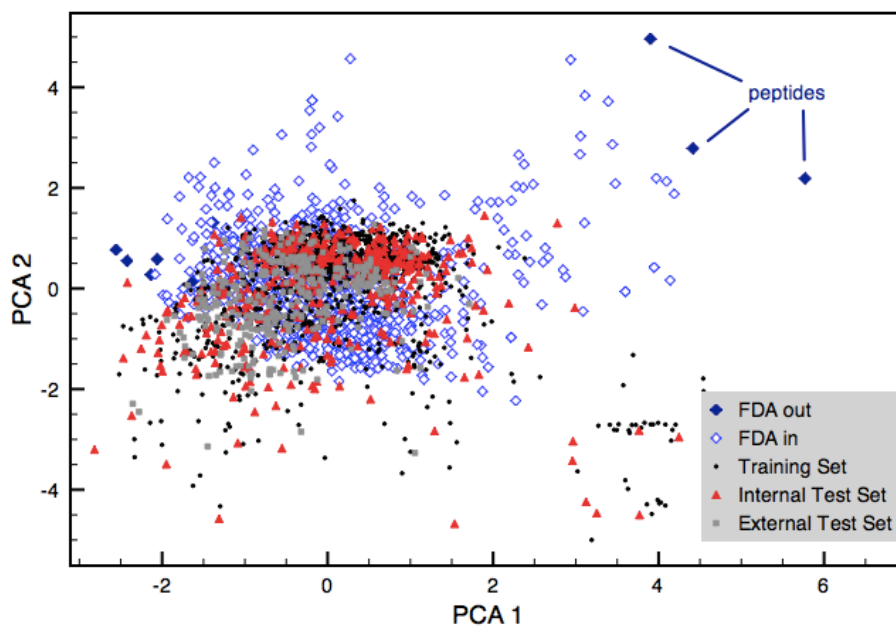
SI-Figure 4: Applicability domain experiment using "ED approach" Compounds shown as follows: Training compounds: Gray dots, FDA compounds: red square, Test compounds: Black cross.

9

Chapter 4

Concluding Discussion

The ABC transporter P-glycoprotein (P-gp) represents a major obstacle in drug discovery. The polyspecific efflux pump prevents numerous therapeutics from entering their target cells and thus triggers the occurrence of MDR. Additionally, compounds that modulate P-gp activity can lead to unwanted drug-drug interactions when being co-administered with other therapeutics.[1–3] These transporter-based drug-drug interactions prompted the FDA to publish a guidance that commits new drug candidates to be screened for P-gp activity.[4] Because of these implications, the early identification of P-gp modulators is an important goal in drug research. Successful *in silico* studies about P-gp and other ABC transporters have already been performed and are summarized in the reviews in Chapter 2.[5–8] However, most of the *in silico* studies done so far focused on ligand-based approaches, comprising 2D and 3D QSAR, ligand-based pharmacophore modeling and others. An explanation for that is the lack of high resolution structural information for membrane proteins. Section 2.2.1 already pointed out the difficulties regarding the application of structure-based approaches on transmembrane transporters. Although they account for about 50 % of all drug targets, the number of unique high-resolution structures of membrane proteins available in the Brookhaven's Protein Databank (PDB, www.pdb.org)[9] lies around 350.[10] Nevertheless, understanding the underlying mechanism of drugs binding to P-glycoprotein, is a major prerequisite for overcoming the problems

associated with P-gp. In that sense, structure-based design represents the method of choice.

Still, the application of such has to be performed with caution. As already pointed out in the review in section 2.1.3, performing docking with low resolution structures can lead to well-scored artifacts that do not represent the correct binding mode.[11] Especially when docking into homology models, an "erroneous" scoring function can lead to wrong results. Thus, if the scoring procedure cannot be validated, it should be limited to a minimum.

That is why in the publications presented in Section 3.1 and 3.2, information about the ligands was taken as criterion for the selection of the correct binding mode. Both studies applied the theory that compounds bearing a similar structure, will bind in a similar fashion. These common binding modes were identified by a thorough clustering process (common scaffold clustering). In case of the paper by Jabeen et al.[12] (Section 3.1), this led to two different clusters for the diastereoisomeric series of benzopyrano[3,4-b][1,4]oxazines. Finally, the binding modes identified by the method are in agreement with lipophilicity analyses performed with both series, thus giving a hint about the stereoselective activity of these P-gp inhibitors.

Similarly, the study performed by Klepsch et al.[13] (Section 3.2) could identify a highly probable binding mode of the P-gp inhibitor propafenone and its derivates. Applying the common scaffold clustering workflow and additionally integrating information from SAR-studies about propafenone derivatives turned out to be a successful strategy to deal with this fuzzy protein. The binding mode identified could not only explain SAR- and experimental data, its stability in a dynamic environment could also be proved by performing MD simulation studies. Moreover, the prospective quality of this binding mode has been confirmed by its ability to identify new P-gp inhibitors. Screening a large vendor database with a structure-based pharmacophore model on basis of the protein-ligand complex generated by docking, yielded two highly active compounds.

When facing the facts, it is clear that structure-based methods can only be as good as the information they are provided. As the availability of high-resolution structures is thus essential, working with membrane proteins is a

161

very challenging scenario. Highly resolved ($\leqslant 2\,\text{Å}$) 3D structures of membrane proteins are extremely rare and account only for 20 % of the already small number of available structures.[14] The lack of resolution therefore generates a blurry layer of uncertainness on top of the investigated problem. Additionally, applying scoring functions for pose selection again adds another layer and thus increases the risk of wrong results. That is why dealing with low resolution must be compensated by additional hard facts gathered about the protein and the ligands of interest.

While this can be successfully applied when investigating the binding mode of a class of compounds, as explained above, it is not really applicable when working with almost 2000 structurally diverse compounds, as demonstrated in the classification study in Section 3.4. The study showed, that the machine learning techniques SVM and Random Forrest were suitable methods for classifying P-gp inhibitors, achieving prediction accuracies of 73-75 % with an external test set.

The structure-based method docking, on the other hand, was not capable of performing equally reliable predictions. However, also in this case did the integration of additional ligand information improve accuracy. The fact that P-gp inhibitors have to diffuse into the lipid bilayer to reach the protein's active site was implemented by combining the docking score with the ligands' logP value.

The improved results advise further investigation of the combination of scoring functions with ligand descriptors. More elaborated associations could probably lead to good classifications with the advantage of providing additional binding information.

In the end, it is necessary to face and respect the limitations of a method. And although structure-based methods might not yet be able to globally classify P-gp inhibitors and non-inhibitors, the application of such is essential for understanding binding and thus transport or blocking respectively.

Bibliography

[1] Hochman JH, Yamazaki M, Ohe T, Lin JH (2002) Evaluation of drug interactions with p-glycoprotein in drug discovery: in vitro assessment of the potential for drug-drug interactions with p-glycoprotein. Curr Drug Metab 3: 257-73.

[2] Marchetti S, Mazzanti R, Beijnen JH, Schellens JHM (2007) Concise review: Clinical relevance of drug drug and herb drug interactions mediated by the abc transporter abcb1 (mdr1, p-glycoprotein). Oncologist 12: 927-41.

[3] Keogh JP (2012) Membrane transporters in drug development. Adv Pharmacol 63: 1-42.

[4] Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (2012). Drug interaction studies - study design, data analysis, implications for dosing, and labeling recommendations. http://www.fda.gov/downloads/Drugs/GuidanceCompliance RegulatoryInformation/Guidances/UCM292362.pdf.

[5] Klepsch F, Stockner T, Erker T, Müller M, Chiba P, et al. (2010) Using structural and mechanistic information to design novel inhibitors/substrates of p-glycoprotein. Curr Top Med Chem 10: 1769-74.

[6] Klepsch F, Jabeen I, Chiba P, Ecker GF (2010) Pharmacoinformatic approaches to design natural product type ligands of abc-transporters. Curr Pharm Des 16: 1742-52.

[7] Jurik A, Klepsch F, Zdrazil B (2012) Medicinal Chemistry and Drug Design, InTech, chapter Molecular Modeling and Simulation of Membrane Transport Proteins. pp. 1–34.

[8] Klepsch F, Ecker GF (2010) Impact of the recent mouse p-glycoprotein structure for structure-based ligand design. Molecular Informatics 29: 276–286.

[9] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The protein data bank. Nucleic Acids Res 28: 235-42.

[10] White S (2012) Membrane proteins of known 3d structure. http://blancobiomoluciedu/mpstruc/listAll/list .

[11] Ferrara P, Jacoby E (2007) Evaluation of the utility of homology models in high throughput docking. J Mol Model 13: 897-905.

[12] Jabeen I, Wetwitayaklung P, Klepsch F, Parveen Z, Chiba P, et al. (2011) Probing the stereoselectivity of p-glycoprotein-synthesis, biological activity and ligand docking studies of a set of enantiopure benzopyrano[3,4-b][1,4]oxazines. Chem Commun (Camb) 47: 2586-8.

[13] Klepsch F, Chiba P, Ecker GF (2011) Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of p-glycoprotein. Plos Comput Biol 7: e1002036.

[14] Raman P, Cherezov V, Caffrey M (2006) The membrane protein data bank. Cellular and Molecular Life Sciences 63: 36–51.

Appendix A

MD simulation setup

The MD simulations were conducted using the freely available software package GROMACS.[1] Detailed settings of the simulations are shown in Table A.1.

Table A.1: Simulation settings.

Program	Gromacs 4.5.3
Force Field	Gromos 53a6
Temperature coupling	$310\,\mathrm{K}$ (nose-hoover)
Pressure coupling	$1\,\mathrm{bar}$ (Parrinello-Rahman)
Periodic boundary conditions	on
Equilibration (NVT/NPT)	$100\,\mathrm{ps}$ / $1\,\mathrm{ns}$
Production run	$10\,\mathrm{ns}$
Water model	SPC/E
Membrane	POPC

Topologies and coordinates The binding mode complexes used for this study were taken from the docking study presented in Section 3.2.[2] The topologies for the protein have been automatically generated by GROMACS, using the *pdb2gmx* program. For the ligands, the topologies were prepared manually (topology files are included in the appendix). The coordinates and topologies for the POPC membrane were taken from Pieter Tieleman's webpage[3] and were preprocessed by Dr. Thomas Stockner from the Medical

165

University of Vienna. During the preprocessing the size of the membrane had been adjusted and an equilibration had been conducted. In Figure A.1 the schematic workflow for the MD setup is depicted. After the topology and coordinate files for the protein-ligand complex and the membrane were prepared, they were minimized separately to correct locally unfavorable conformations.

Membrane insertion After the minimization the protein-ligand complex was subsequently embedded into the membrane by using *g_ membed*,[4] a built-in program in GROMACS. Briefly, *g_ membed* shrinks the protein, positions it in the membrane and lets it grow again to its initial size. Overlapping lipid molecules will be removed and the system is minimized for 2 ns.

Water box Thus, a rectangular box has been created (*editconf* program), which was subsequently filled with water molecules using the water model SPC/E[5] (*genbox* program). The SPC/E (extended simple point charge) water model is a so-called 3-site model, that comprises of three centers of charge. Positive charges are conferred to the hydrogen atoms of the molecule, while the oxygen is assigned a negative charge. The H-O-H angle is slightly larger than the experimental angle (109.42 compared to 104.45), which should correct the dipole moment in this simplified water model. In order to neutralize the full system, 13 Cl^- counter ions have been introduced randomly via the program *genion*, implemented in GROMACS. As the presence of ions in the binding site was not desired in this case, the location of the counter ions was inspected visually. Minimization of the system with subsequent position-restrained MD should result in the relaxation of the added solvent atoms. For this position-restrained MD and all other simulation runs (equilibration and production run), the molecules were not coupled seperately. They rather have been divided into following coupling groups, which have been generated using the program *make_ ndx*:

- protein + ligand + membrane

- solvent (water) + ions

166
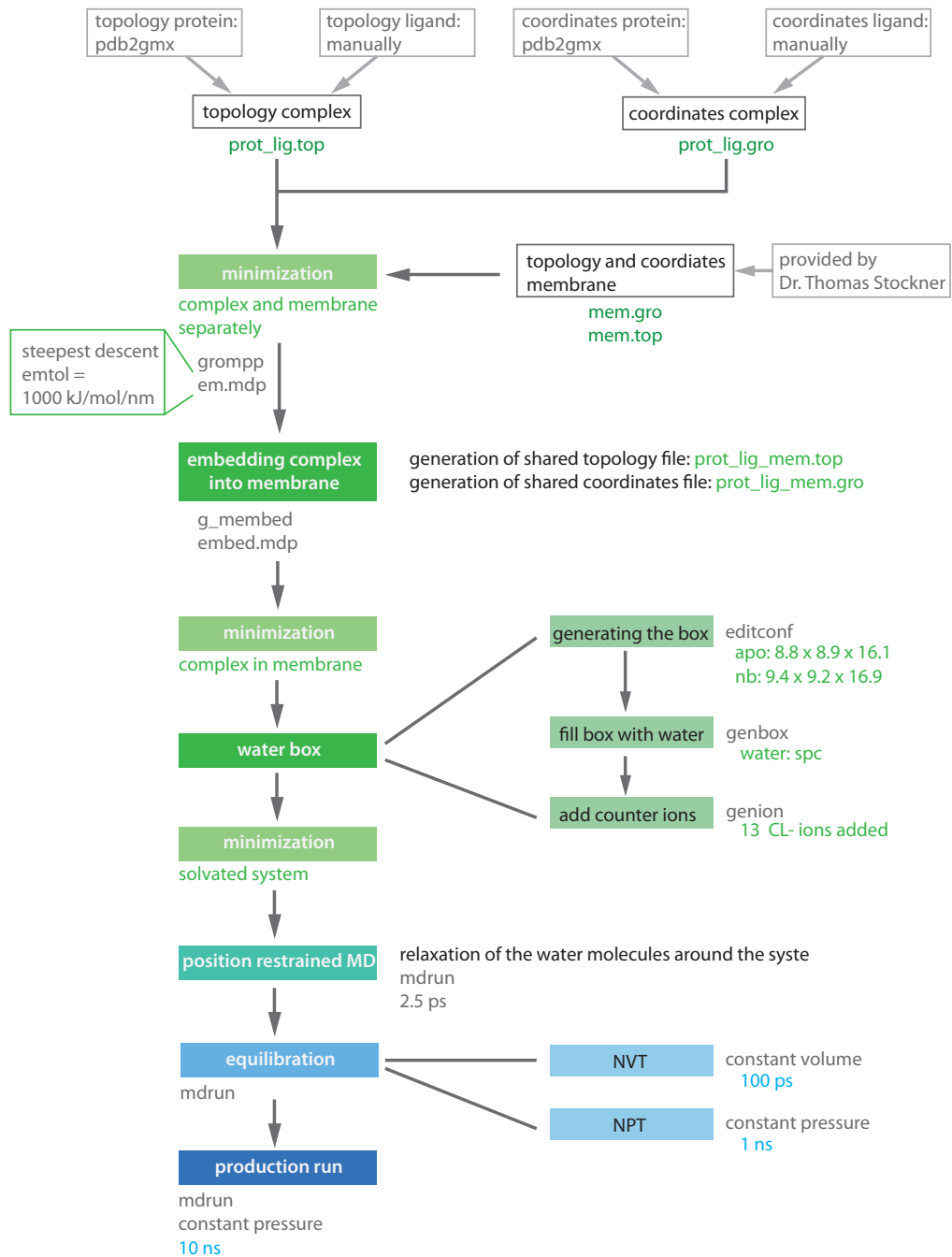
Figure A.1: MD setup for all simulations.

Equilibration In the first step, the system was simulated at NVT, referring to constant number of particles, volume and temperature. During this run the temperature coupling was activated using a Berendsen thermostat[6] and the system was able to relax to these new conditions. The temperature was set to 310 K, which was above the reported phase transition temperature for POPC membranes.[7] In the second equilibration step the system was simulated at NPT (constant number of particles, constant pressure and constant temperature). At this point, the pressure coupling to 1 bar was switched on, by using a Berendsen barostat.[6] Because of the presence of the membrane, semi-isotropic pressure coupling was chosen for the simulation. Using that setting the x and y direction (bilayer plane) were coupled isotropically, but the z direction (normal to the bilayer) was coupled separately. This provides a more stable surface area per lipid during the simulation.

Production run After the equilibration phase, the production run was started. For this simulation step, the position restraints of the protein and the ligand could be removed. However, as the nucleotide-binding domains seemed to be quite unstable during previous test runs, they were restrained, applying a force of $1000 \, \text{kJ} \, \text{mol}^{-1} \, \text{nm}^{-1}$. In total 10 ns per production run have been simulated.

Bibliography

[1] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation 4: 435–447.

[2] Klepsch F, Chiba P, Ecker GF (2011) Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of p-glycoprotein. Plos Comput Biol 7: e1002036.

[3] Tieleman P (2012). Structures and topologies. URL `http://moose.bio.ucalgary.ca/`.

[4] Wolf MG, Hoefling M, Aponte-Santamaria C, Grubmueller H, Groenhof G (2010) g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. J Comput Chem 31: 2169–2174.

[5] Berendsen H, Grigera J, Sraatsma T (1987) The missing term in effective pair potentials. J Phys Chem-Us 91: 6269–6271.

[6] Berendsen H, Postma J, Van Gunsteren W, Dinola A, Haak J (1984) Molecular dynamics with coupling to an external bath. Journal of Chemical Physics 81: 3684–3690.

[7] Keough K (2003) How thin can glass be? new ideas, new approaches. Biophys J 85: 2785-6.

Ligand topologies

GPV062

```
[ moleculetype ]
;  Name            nrexcl
   gp062n           3

  [ atoms ]

;nr   type   resnr   residu   atom   cgnr   charge      mass
1       C      1      62N     C01     1       -0.1    12.011
2      HC      1      62N     H01     1        0.1     1.008
3       C      1      62N     C02     2       -0.1    12.011
4      HC      1      62N     H02     2        0.1     1.008
5       C      1      62N     C03     3       -0.1    12.011
6      HC      1      62N     H03     3        0.1     1.008
7       C      1      62N     C04     4       -0.1    12.011
8      HC      1      62N     H04     4        0.1     1.008
9       C      1      62N     C05     5       -0.1    12.011
10     HC      1      62N     H05     5        0.1     1.008
11      C      1      62N     C06     6          0    12.011
12    CH2      1      62N     C07     7          0    14.027
13    CH2      1      62N     C08     8          0    14.027
14      C      1      62N     C09     9       0.38    12.011
15      O      1      62N     O09     9      -0.38    15.999
16      C      1      62N     C10    10          0    12.011
17      C      1      62N     C11    11       -0.1    12.011
18     HC      1      62N     H11    11        0.1     1.008
19      C      1      62N     C12    12       -0.1    12.011
20     HC      1      62N     H12    12        0.1     1.008
21      C      1      62N     C13    13       -0.1    12.011
22     HC      1      62N     H13    13        0.1     1.008
23      C      1      62N     C14    14       -0.1    12.011
24     HC      1      62N     H14    14        0.1     1.008
25      C      1      62N     C15    15       0.16    12.011
26     OA      1      62N     O16    15      -0.36    15.999
27    CH2      1      62N     C17    15        0.2    14.027
28    CH1      1      62N     C18    16       0.15    13.019
29     OA      1      62N     O18    16     -0.548    15.999
30      H      1      62N     H18    16      0.398     1.008
```

170

```
31     CH2     1     62N    C19    17      0.15    14.027
32      NT     1     62N    N20    17     -0.45    14.007
33    CH2r     1     62N    C21    17      0.15    14.027
34    CH2r     1     62N    C25    17      0.15    14.027
35    CH2r     1     62N    C24    18         0    14.027
36     CH0     1     62N    C23    19      0.15    12.011
37      OA     1     62N    O23    19    -0.548    15.999
38       H     1     62N    H23    19     0.398     1.008
39    CH2r     1     62N    C22    20         0    14.027
40       C     1     62N    C26    21         0    12.011
41       C     1     62N    C27    22      -0.1    12.011
42      HC     1     62N    H27    22       0.1     1.008
43       C     1     62N    C28    23      -0.1    12.011
44      HC     1     62N    H28    23       0.1     1.008
45       C     1     62N    C29    24      -0.1    12.011
46      HC     1     62N    H29    24       0.1     1.008
47       C     1     62N    C30    25      -0.1    12.011
48      HC     1     62N    H30    25       0.1     1.008
49       C     1     62N    C31    26      -0.1    12.011
50      HC     1     62N    H31    26       0.1     1.008

   [ bonds ]

;nr  ai   aj   funct
1     2    2   gb_3
1     3    2   gb_16
1    11    2   gb_16
3     4    2   gb_3
3     5    2   gb_16
5     6    2   gb_3
5     7    2   gb_16
7     8    2   gb_3
7     9    2   gb_16
9    10    2   gb_3
9    11    2   gb_16
11   12    2   gb_27
12   13    2   gb_27
13   14    2   gb_27
14   15    2   gb_5
14   16    2   gb_23
16   17    2   gb_16
16   25    2   gb_16
17   18    2   gb_3
17   19    2   gb_16
```

```
19    20    2    gb_3
19    21    2    gb_16
21    22    2    gb_3
21    23    2    gb_16
23    24    2    gb_3
23    25    2    gb_16
25    26    2    gb_13
26    27    2    gb_18
27    28    2    gb_27
28    29    2    gb_18
28    31    2    gb_27
29    30    2    gb_1
31    32    2    gb_21
32    33    2    gb_21
32    34    2    gb_21
33    39    2    gb_26
34    35    2    gb_26
35    36    2    gb_26
36    37    2    gb_18
36    39    2    gb_26
36    40    2    gb_27
37    38    2    gb_1
40    41    2    gb_16
40    49    2    gb_16
41    42    2    gb_3
41    43    2    gb_16
43    44    2    gb_3
43    45    2    gb_16
45    46    2    gb_3
45    47    2    gb_16
47    48    2    gb_3
47    49    2    gb_16
49    50    2    gb_3

   [ pairs ]

;nr   ai   aj   funct
1     13   1
9     13   1
12    15   1
13    17   1
13    25   1
15    17   1
15    25   1
```

172

```
16    27   1
23    27   1
25    28   1
26    29   1
26    31   1
27    30   1
27    32   1
28    33   1
28    34   1
29    32   1
30    31   1
31    35   1
31    39   1
32    36   1
33    35   1
33    37   1
33    40   1
34    37   1
34    39   1
34    40   1
35    38   1
35    41   1
35    49   1
37    41   1
37    49   1
38    39   1
38    40   1
39    41   1
39    49   1
```

[angles]

```
;nr  ai   aj  ak   funct
2    1    3   2    ga_25
2    1    11  2    ga_25
3    1    11  2    ga_27
1    3    4   2    ga_25
1    3    5   2    ga_27
4    3    5   2    ga_25
3    5    6   2    ga_25
3    5    7   2    ga_27
6    5    7   2    ga_25
5    7    8   2    ga_25
5    7    9   2    ga_27
```

8	7	9	2	ga_25
7	9	10	2	ga_25
7	9	11	2	ga_27
10	9	11	2	ga_25
1	11	9	2	ga_27
1	11	12	2	ga_27
9	11	12	2	ga_27
11	12	13	2	ga_15
12	13	14	2	ga_15
13	14	15	2	ga_27
13	14	16	2	ga_27
15	14	16	2	ga_27
14	16	17	2	ga_27
14	16	25	2	ga_27
17	16	25	2	ga_27
16	17	18	2	ga_25
16	17	19	2	ga_27
18	17	19	2	ga_25
17	19	20	2	ga_25
17	19	21	2	ga_27
20	19	21	2	ga_25
19	21	22	2	ga_25
19	21	23	2	ga_27
22	21	23	2	ga_25
21	23	24	2	ga_25
21	23	25	2	ga_27
24	23	25	2	ga_25
16	25	23	2	ga_27
16	25	26	2	ga_27
23	25	26	2	ga_27
25	26	27	2	ga_13
26	27	28	2	ga_15
27	28	29	2	ga_13
27	28	31	2	ga_15
29	28	31	2	ga_13
28	29	30	2	ga_12
28	31	32	2	ga_15
31	32	33	2	ga_15
31	32	34	2	ga_15
33	32	34	2	ga_15
32	33	39	2	ga_15
32	34	35	2	ga_15
34	35	36	2	ga_15

```
35    36    37    2    ga_13
35    36    39    2    ga_13
35    36    40    2    ga_13
37    36    39    2    ga_13
37    36    40    2    ga_13
39    36    40    2    ga_13
36    37    38    2    ga_12
33    39    36    2    ga_15
36    40    41    2    ga_27
36    40    49    2    ga_27
41    40    49    2    ga_27
40    41    42    2    ga_25
40    41    43    2    ga_27
42    41    43    2    ga_25
41    43    44    2    ga_25
41    43    45    2    ga_27
44    43    45    2    ga_25
43    45    46    2    ga_25
43    45    47    2    ga_27
46    45    47    2    ga_25
45    47    48    2    ga_25
45    47    49    2    ga_27
48    47    49    2    ga_25
40    49    47    2    ga_27
40    49    50    2    ga_25
47    49    50    2    ga_25

   [ dihedrals ]

;nr  ai   aj   ak   al   funct   phi0  cp   mult
1    11   12   13   1    gd_39
11   12   13   14   1    gd_34
12   13   14   16   1    gd_39
13   14   16   17   1    gd_10
16   25   26   27   1    gd_11
25   26   27   28   1    gd_34
26   27   28   31   1    gd_34
27   28   29   30   1    gd_23
27   28   31   32   1    gd_34
28   31   32   33   1    gd_29
34   32   33   39   1    gd_34
33   32   34   35   1    gd_34
32   33   39   36   1    gd_34
32   34   35   36   1    gd_34
```

```
34   35   36   40   1   gd_34
35   36   37   38   1   gd_23
35   36   39   33   1   gd_34
35   36   40   41   1   gd_39

   [ dihedrals ]

;nr  ai   aj   ak   al   funct
1    2    3    11   2   gi_1
1    3    5    7    2   gi_1
3    1    4    5    2   gi_1
3    1    11   9    2   gi_1
3    5    7    9    2   gi_1
5    3    6    7    2   gi_1
5    7    9    11   2   gi_1
7    5    8    9    2   gi_1
7    9    11   1    2   gi_1
9    7    10   11   2   gi_1
11   1    3    5    2   gi_1
11   1    9    12   2   gi_1
14   13   15   16   2   gi_1
16   14   17   25   2   gi_1
16   17   19   21   2   gi_1
17   16   18   19   2   gi_1
17   16   25   23   2   gi_1
17   19   21   23   2   gi_1
19   17   20   21   2   gi_1
19   21   23   25   2   gi_1
21   19   22   23   2   gi_1
21   23   25   16   2   gi_1
23   21   24   25   2   gi_1
25   16   17   19   2   gi_1
25   16   23   26   2   gi_1
28   27   31   29   2   gi_2
40   36   41   49   2   gi_1
40   41   43   45   2   gi_1
41   40   42   43   2   gi_1
41   40   49   47   2   gi_1
41   43   45   47   2   gi_1
43   41   44   45   2   gi_1
43   45   47   49   2   gi_1
45   43   46   47   2   gi_1
45   47   49   40   2   gi_1
47   45   48   49   2   gi_1
```

```
49   40   41   43   2   gi_1
49   40   47   50   2   gi_1
```

GPV019

```
[ moleculetype ]
;  Name            nrexcl
   gp019n           3

  [ atoms ]

;nr   type   resnr   residu   atom   cgnr   charge     mass
1      C      1       19N      C01    1       -0.1    12.011
2      HC     1       19N      H01    1        0.1     1.008
3      C      1       19N      C02    2       -0.1    12.011
4      HC     1       19N      H02    2        0.1     1.008
5      C      1       19N      C03    3       -0.1    12.011
6      HC     1       19N      H03    3        0.1     1.008
7      C      1       19N      C04    4       -0.1    12.011
8      HC     1       19N      H04    4        0.1     1.008
9      C      1       19N      C05    5       -0.1    12.011
10     HC     1       19N      H05    5        0.1     1.008
11     C      1       19N      C06    6          0    12.011
12     CH2    1       19N      C07    7          0    14.027
13     CH2    1       19N      C08    8          0    14.027
14     C      1       19N      C09    9       0.38    12.011
15     O      1       19N      O09    9      -0.38    15.999
16     C      1       19N      C10    10         0    12.011
17     C      1       19N      C11    11      -0.1    12.011
18     HC     1       19N      H11    11       0.1     1.008
19     C      1       19N      C12    12      -0.1    12.011
20     HC     1       19N      H12    12       0.1     1.008
21     C      1       19N      C13    13      -0.1    12.011
22     HC     1       19N      H13    13       0.1     1.008
23     C      1       19N      C14    14      -0.1    12.011
24     HC     1       19N      H14    14       0.1     1.008
25     C      1       19N      C15    15      0.16    12.011
26     OA     1       19N      O16    15     -0.36    15.999
27     CH2    1       19N      C17    15       0.2    14.027
28     CH1    1       19N      C18    16      0.15    13.019
29     OA     1       19N      O18    16     -0.548   15.999
```

177

```
30       H      1     19N   H18   16      0.398    1.008
31     CH2      1     19N   C19   17       0.15    14.027
32      NT      1     19N   N20   17      -0.45    14.007
33    CH2r      1     19N   C21   17       0.15    14.027
34    CH2r      1     19N   C25   17       0.15    14.027
35    CH2r      1     19N   C24   18       0.15    14.027
36      NT      1     19N   N23   18      -0.35    12.011
37    CH2r      1     19N   C22   18       0.15    14.027
38       C      1     19N   C26   18       0.05    12.011
39       C      1     19N   C27   19       -0.1    12.011
40      HC      1     19N   H27   19        0.1    1.008
41       C      1     19N   C28   20       -0.1    12.011
42      HC      1     19N   H28   20        0.1    1.008
43       C      1     19N   C29   21       -0.1    12.011
44      HC      1     19N   H29   21        0.1    1.008
45       C      1     19N   C30   22       -0.1    12.011
46      HC      1     19N   H30   22        0.1    1.008
47       C      1     19N   C31   23       -0.1    12.011
48      HC      1     19N   H31   23        0.1    1.008

   [ bonds ]

;nr  ai  aj   funct
1    2   2    gb_3
1    3   2    gb_16
1    11  2    gb_16
3    4   2    gb_3
3    5   2    gb_16
5    6   2    gb_3
5    7   2    gb_16
7    8   2    gb_3
7    9   2    gb_16
9    10  2    gb_3
9    11  2    gb_16
11   12  2    gb_27
12   13  2    gb_27
13   14  2    gb_27
14   15  2    gb_5
14   16  2    gb_23
16   17  2    gb_16
16   25  2    gb_16
17   18  2    gb_3
17   19  2    gb_16
19   20  2    gb_3
```

```
19    21   2   gb_16
21    22   2   gb_3
21    23   2   gb_16
23    24   2   gb_3
23    25   2   gb_16
25    26   2   gb_13
26    27   2   gb_18
27    28   2   gb_27
28    29   2   gb_18
28    31   2   gb_27
29    30   2   gb_1
31    32   2   gb_21
32    33   2   gb_21
32    34   2   gb_21
33    37   2   gb_26
34    35   2   gb_26
35    36   2   gb_23
36    37   2   gb_23
36    38   2   gb_12
38    39   2   gb_16
38    47   2   gb_16
39    40   2   gb_3
39    41   2   gb_16
41    42   2   gb_3
41    43   2   gb_16
43    44   2   gb_3
43    45   2   gb_16
45    46   2   gb_3
45    47   2   gb_16
47    48   2   gb_3

  [ pairs ]

;nr   ai   aj   funct
1     13   1
9     13   1
12    15   1
13    17   1
13    25   1
15    17   1
15    25   1
16    27   1
23    27   1
25    28   1
```

```
26    29   1
26    31   1
27    30   1
27    32   1
28    33   1
28    34   1
29    32   1
30    31   1
31    35   1
31    37   1
32    36   1
33    35   1
33    38   1
34    37   1
34    38   1
35    39   1
35    47   1
37    39   1
37    47   1
37    39   1
```

[angles]

```
;nr  ai   aj   ak   funct
2    1    3    2    ga_25
2    1    11   2    ga_25
3    1    11   2    ga_27
1    3    4    2    ga_25
1    3    5    2    ga_27
4    3    5    2    ga_25
3    5    6    2    ga_25
3    5    7    2    ga_27
6    5    7    2    ga_25
5    7    8    2    ga_25
5    7    9    2    ga_27
8    7    9    2    ga_25
7    9    10   2    ga_25
7    9    11   2    ga_27
10   9    11   2    ga_25
1    11   9    2    ga_27
1    11   12   2    ga_27
9    11   12   2    ga_27
11   12   13   2    ga_15
12   13   14   2    ga_15
```

13	14	15	2	ga_27
13	14	16	2	ga_27
15	14	16	2	ga_27
14	16	17	2	ga_27
14	16	25	2	ga_27
17	16	25	2	ga_27
16	17	18	2	ga_25
16	17	19	2	ga_27
18	17	19	2	ga_25
17	19	20	2	ga_25
17	19	21	2	ga_27
20	19	21	2	ga_25
19	21	22	2	ga_25
19	21	23	2	ga_27
22	21	23	2	ga_25
21	23	24	2	ga_25
21	23	25	2	ga_27
24	23	25	2	ga_25
16	25	23	2	ga_27
16	25	26	2	ga_27
23	25	26	2	ga_27
25	26	27	2	ga_13
26	27	28	2	ga_15
27	28	29	2	ga_13
27	28	31	2	ga_15
29	28	31	2	ga_13
28	29	30	2	ga_12
28	31	32	2	ga_15
31	32	33	2	ga_15
31	32	34	2	ga_15
33	32	34	2	ga_15
32	33	39	2	ga_15
32	34	35	2	ga_15
34	35	36	2	ga_15
35	36	37	2	ga_13
35	36	39	2	ga_13
35	36	40	2	ga_13
37	36	39	2	ga_13
37	36	40	2	ga_13
39	36	40	2	ga_13
36	37	38	2	ga_12
33	39	36	2	ga_15
36	40	41	2	ga_27

```
36    40   49   2    ga_27
41    40   49   2    ga_27
40    41   42   2    ga_25
40    41   43   2    ga_27
42    41   43   2    ga_25
41    43   44   2    ga_25
41    43   45   2    ga_27
44    43   45   2    ga_25
43    45   46   2    ga_25
43    45   47   2    ga_27
46    45   47   2    ga_25
45    47   48   2    ga_25
45    47   49   2    ga_27
48    47   49   2    ga_25
40    49   47   2    ga_27
40    49   50   2    ga_25
47    49   50   2    ga_25
```

[dihedrals]

```
;nr  ai   aj   ak   al   funct   phi0  cp   mult
1    11   12   13   1    gd_39
11   12   13   14   1    gd_34
12   13   14   16   1    gd_39
13   14   16   17   1    gd_10
16   25   26   27   1    gd_11
25   26   27   28   1    gd_34
26   27   28   31   1    gd_34
27   28   29   30   1    gd_23
27   28   31   32   1    gd_34
28   31   32   33   1    gd_29
34   32   33   37   1    gd_34
33   32   34   35   1    gd_34
32   33   37   36   1    gd_34
32   34   35   36   1    gd_34
34   35   36   38   1    gd_29
35   36   37   33   1    gd_29
35   36   38   39   1    gd_41
```

[dihedrals]

```
;nr  ai   aj   ak   al   funct
1    2    3    11   2    gi_1
1    3    5    7    2    gi_1
3    1    4    5    2    gi_1
```

3	1	11	9	2	gi_1
3	5	7	9	2	gi_1
5	3	6	7	2	gi_1
5	7	9	11	2	gi_1
7	5	8	9	2	gi_1
7	9	11	1	2	gi_1
9	7	10	11	2	gi_1
11	1	3	5	2	gi_1
11	1	9	12	2	gi_1
14	13	15	16	2	gi_1
16	14	17	25	2	gi_1
16	17	19	21	2	gi_1
17	16	18	19	2	gi_1
17	16	25	23	2	gi_1
17	19	21	23	2	gi_1
19	17	20	21	2	gi_1
19	21	23	25	2	gi_1
21	19	22	23	2	gi_1
21	23	25	16	2	gi_1
23	21	24	25	2	gi_1
25	16	17	19	2	gi_1
25	16	23	26	2	gi_1
28	27	31	29	2	gi_2
38	36	39	47	2	gi_1
38	39	41	43	2	gi_1
39	38	40	41	2	gi_1
39	38	47	45	2	gi_1
39	41	43	45	2	gi_1
41	39	42	43	2	gi_1
41	43	45	47	2	gi_1
43	41	44	45	2	gi_1
43	45	47	38	2	gi_1
45	43	46	47	2	gi_1
47	38	39	41	2	gi_1
47	38	45	48	2	gi_1

Abstract English

The ABC transporter P-glycoprotein (P-gp) is responsible for the translocation of a broad variety of different substances across the lipid bilayer. As a result, P-gp on the one hand exhibits important barrier and detoxifying functions, on the other hand it affects the ADMET properties of drugs, triggers unwanted drug-drug interactions and its overexpression is responsible for the development of multidrug-resistance (MDR), one major reason for the failure of antibiotic or anticancer therapies.

Thus, reliable *in silico* methods for the early identification of P-gp modulators or virtual screening approaches for discovering new inhibitors is of high interest in drug discovery.

This thesis outlines in three independent studies how structure-based methods can be used for tackling the problems triggered by P-gp. Although structure-based design has to be performed with precaution when dealing with membrane proteins, it is highly necessary for understanding intermolecular interactions between the ligand and its target protein. Two studies are presented, that describe the use of docking for the identification of binding modes of two classes of P-gp inhibitors. The workflows applied show how the implemention of external information is able to reduce the debatable use of scoring functions to a minimum. The results obtained provided useful information for screening for new P-gp inhibitors.

Furthermore, a third study that compares the classification performance of ligand-based and structure-based *in silico* methods is also included in this thesis. Although the former models tend to be more accurate, the implementation of some ligand information could improve the structure-based models, emphasizing the importance of combined scoring functions (merging scoring and descriptor values) for future studies.

Abstract Deutsch

Der membranständige ABC-Transporter P-glycoprotein (P-gp) ist für den Export einer großen Anzahl von Substanzklassen aus der Zelle verantwortlich. Einerseits wird ihm dadurch eine wichtige Schutz- und Entgiftungsfunktion zugeschrieben, andererseits wird P-gp aus diesem Grund sowohl mit veränderten ADMET Eigenschaften, als auch mit unerwünschten Arzneistoffwechselwirkungen und der Resistenzentwicklung von Tumoren und Krankheitserregern in Zusammenhang gebracht.

Aus diesem Grund ist die frühzeitige Erkennung potentieller P-gp Substrate und Inhibitoren mittels verlässlicher *in silico* Methoden von großer Bedeutung in der Wirkstoffentwicklung. In diesem Sinne, werden in dieser Dissertation drei Studien präsentiert, die die erfolgreiche Anwendung strukturbasierter *in silico* Modelle zur Vorhersage von P-gp-Wirkstoff-Interaktionen darstellen.

In zwei Studien wurde Protein-Ligand Docking dazu verwendet, um die Bindungsmodi für zwei Klassen von P-gp-Inhibitoren zu identifizieren. Der Augenmerk lag hierbei auf der Implementierung von externen Informationen, die die Verwendung umstrittener Fitnessfunktionen limitiert. Die daraus gewonnenen Resultate konnten erfolgreich für ein virtuelles Screening für neue P-gp Inhibitoren weiterverwendet werden.

Außerdem wurden in einer dritten Studie potentielle Methoden zur Klassifizierung von P-gp Inhibitoren untersucht. Dabei wurden liganden- und struktur-basierte Techniken gegenübergestellt. Obwohl sich herausstellte, dass Erstere in puncto Treffsicherheit deutlich überlegen waren, konnten Letztere durch Einbeziehung von Ligandeninformation die Vorhersagekraft verbessern. Fitnessfunktionen, die Scoring- und Ligandendeskriptorwerte kombinieren, zeigen daher großes Potential für Methoden des strukturbasierten Drug Designs.

185

List of Abbreviations

ABC   ATP binding cassette

ABCB1   ATP binding casettte transporter, subfamily B1

FDA   Food and Drug Administration

MD   molecular dynamics

MDR   multidrug resistance

NB   nucleotide binng

P-gp   P-glycoprotein

POPC   1-palmitoyl-2-oleoyl-sn-glycero-3-phosophocholine

RMSD   root rean square deviation

SAR   Structure-Activity Relationship

SHED   Shannon entropy descriptors

SPC/E   extended simple point charge (water model)

TM   transmembrane

WHO   World Health Organization

Freya Klepsch CURRICULUM VITAE

Date of birth:
1st November 1986 (Vienna)

Nationality: Austrian

Apostelgasse 17/50-51
1030 Vienna
M +43-699-18154935 (private)
P +43-1-4277-55112
freya.klepsch@univie.ac.at

Education

Nov 2008 - Jul 2012	**PhD in Molecular Biology**
Organization	Pharmacoinformatics Research Group University of Vienna Althanstraße 14, 1090 Vienna, AUSTRIA
Principal subject	Molecular Biology
Major	Pharmacoinformatics
Supervisor	Prof. Gerhard F. Ecker
Title	Molecular basis of drug-ABC transporter interaction

Sep 2004 - Sep 2008	**Diploma study in Biotechnology**
Qualification awarded	DI (FH) (equivalent to M.Sc.)
Organisation	FH Campus Wien Vienna Biocenter II, 1030 Vienna, AUSTRIA
Principal subject	Biotechnology
Major	Chemistry of Active Substances
Main subjects included	Chemistry (Organic, Synthetic, Analytic), Biochemistry, Rational Drug Design, Molecular Biology, Genetics
Thesis title	*Screening of highly active essential oils and other plant extracts investigating their antibacterial action against different pathogens of the intestinal tract.*

Professional experience

Nov 2008 - present	**Research assistant, PraeDoc**
Organisation	Pharmacoinformatics Research Group University of Vienna Althanstraße 14, 1090 Vienna, AUSTRIA

2009 - 2011	**Personal assistant to the EFMC president (Prof. Ecker)**
Organisation	European Federation for Medicinal Chemistry www.efmc.info

Jul 2007 - Sep 2008	**Research assistant**
Organisation	Biomin GmbH TZT, Technopark 1, 3430 Tulln, AUSTRIA

International experience

Freya Klepsch

Apostelgasse 17/50-51
1030 Vienna
M +43-699-18154935 (private)
P +43-1-4277-55112
freya.klepsch@univie.ac.at

Apr 2011 - Jul 2011	**Internship**
Organization	Division of Molecular and Computational Toxicology Vrije Universiteit Amsterdam De Boelelaan 1083, 1081 HV, The NETHERLANDS
Project	MD simulations of P-glycoprotein embedded in a membrane
Supervisors	Dr. DP Geerke, Prof. NPE Vermeulen

2008 - 2012	**Regular attendance at scientific meetings and conferences**
	see sections "Talks" and "Posters" below

Teaching experience

2011 - 2012	Computational Life-Sciences University of Vienna
2011 - 2012	Analytics and development of drugs - *In silico* module University of Vienna
2009 - 2010	Rational Drug Design FH Campus Wien, University of Applied Sciences

Skills

Languages

German	native
English	fluent, both written and oral
French	basic, both written and oral

Computer skills

Operating systems	MacOS, Linux, Windows
Programming	Shell scripting, SVL (MOE), python
Office programs	MS Office, OpenOffice, iWork
Scientific programs	MOE, Schrödinger Suite, LigandScout, GOLD, AutoDock, GROMOS, GROMACS, Modeller, VMD, weka

Awards

2011
Scholarship for Scientific Excellence
243rd ACS National Meeting, San Diego

Interests

Organizations

Europin (European Pharmaceutical Initiative), PhD program

OePhG (Austrian Pharmaceutical Society)

American Chemical Society

Freya Klepsch

Apostelgasse 17/50-51
1030 Vienna
M +43-699-18154935 (private)
P +43-1-4277-55112
freya.klepsch@univie.ac.at

Activities

Since 2010	member of the Europin editorial board
2009 - 2010	volunteer at the OeHTB Austrian relief organisation for deafblind people

Sports	running, beach volleyball, snowboarding

Publications

2012

Klepsch F, Vasanthanathan P, Ecker GF (2012) *Ligand- and Structure-based Methods for Prediction of P-gp inhibitors,* **in preparation**

2011

Jurik A, Klepsch F, Zdrazil B (2011) *Molecular Modelling and Simulation of Membrane Transport Proteins*. In: Medicinal Chemistry and Drug Design. in Press

Klepsch F, Chiba P, Ecker GF (2011) *Exhaustive sampling of docking poses reveals binding hypotheses for propafenone type inhibitors of P-glycoprotein.* PLoS Comp Biol 7(5): e1002036.

Malle E, Zhou H, Neuhold J, Spitzenberger B, Klepsch F, Pollak T, Berner O, Ecker GF, Stolt-Bergner PC (2011) *Random mutagenesis of the prokaryotic peptide transporter YDGR identifies potential periplasmatic gating residues.* J Biol Chem 286(26):23121-23131

Jabeen I, Wetwitayaklung P, Klepsch F, Parveen Z, Chiba P, Ecker GF (2011) *Probing he stereoselectivity of P-glycoprotein - synthesis, biological activity and preliminary docking studies of a set of enantiopure benzopyrano[3,4b][1,4]oxazines*. Chem Comm 47(9):2586-2588

2010

Klepsch F, Stockner T, Erker T, Mueller M, Chiba P, Ecker GF (2010) *Using structural and mechanistic information to design novel inhbitors/substrates of P-glycoprotein.* Top Med Chem 10(6):1769-1774

Klepsch F, Ecker GF (2010) *Impact of the recent mouse P-glycoprotein structure for structure-based ligand design.* Mol Inform 29(4):276-286

Klepsch F, Jabeen I, Ecker GF (2010) *Pharmacoinformatic approaches to design natural product type ligands of ABC-transporters.* Curr Pharm Design 16(15):1742-1752

Talks

2011

Klepsch F, Geerke DP, Ecker GF. *Analysis of binding modes of propafenones in P-glycoprotein by means of molecular dynamics simulations.* Joint Meeting of the Ausrian and German Pharmaceutical Societies, Innsbruck

Klepsch F, *Structure based studies on P-glycoprotein*. Summer School on Drug Design, Vienna

2009

Klepsch F, Chiba P, Ecker GF. *Ligand docking of propafenones into ABCB1 using FlexX and Glide* (upgraded poster). VII European Workshop on Drug Design, Siena

Freya Klepsch

Apostelgasse 17/50-51
1030 Vienna
M +43-699-18154935 (private)
P +43-1-4277-55112
freya.klepsch@univie.ac.at

Posters

2012

Klepsch F, Prokes K, Parveen Z, Chiba P, Ecker GF. *Structure based pharmacophore screening for new P-gp inhibitors.* 243rd ACS National Meeting Spring 2012, San Diego

2011

Klepsch F, Vosmeer CR, Stockner T, Chiba P, Geerke DP, Ecker GF. *Application of post-docking molecular dynamics simulations to evaluate binding modes of propafenone analogues in P-gp.* SFB Symposium 2011, Vienna

Klepsch F, Vosmeer CR, Geerke DP, Oostenbrink C, Ecker GF. *nsights into the binding mode of propafenone type P-glycoprotein inhibitors.* 241st ACS National Meeting Spring 2011, Anaheim - MEDI division

Klepsch F, Chiba P, Ecker GF. *Exhaustive Docking Protocol with SAR based Pose Selection.* 241st ACS National Meeting Spring 2011, Anaheim - CINF division

2010

Klepsch F, Ecker GF, *Exhaustive docking protocol with SAR based pose selection.* CCG MOE User group meeting, Cambridge

Klepsch F, Chiba P, Ecker GF. *Exhaustive sampling of docking poses reveals binding modes of propafenones in P-glycoprotein.* EFMC International Symposium on Medicinal Chemistry, Brussels

2009

Klepsch F, Chiba P, Ecker GF. *Docking studies to identify the binding mode of propafenones to P-glycoprotein.* Frontiers in Medicinal Chemistry, Barcelona

Klepsch F, Stockner T, Chiba P, Ecker GF. *Ligand docking of propafenone derivatives into homology models of P-glycoprotein.* ACS Fall 2009 National Meeting, Washington

Klepsch F, Stockner T, Chiba P, Ecker GF. *Ligand docking of propafenones into homology models of P-glycoprotein representing two different states of the catalytic cycle.* Joint Meeting on Medicinal Chemistry, Budapest

Klepsch F, Chiba P, Ecker GF. *Ligand docking of propafenones into ABCB1 using FlexX and Glide.* VII European Workshop on Drug Design, Siena

Klepsch F, Stockner T, Chiba P, Ecker GF. *Ligand docking of propafenones into a homology model of ABCB1.* Meeting of the Austrian Pharmaceutical Society, Vienna

Freya Klepsch

Apostelgasse 17/50-51
1030 Vienna
M +43-699-18154935 (private)
P +43-1-4277-55112
freya.klepsch@univie.ac.at