CHEMMEDCHEM

DOI: 10.1002/cmdc.200700249

Identification of Putative Binding Sites of P-glycoprotein Based on its Homology Model

Christoph Globisch,[a] Ilza K. Pajeva,[a, b] and Michael Wiese*[a]

A homology model of P-glycoprotein based on the crystal structure of the multidrug transporter Sav1866 is developed, incorporated into a membrane environment, and optimized. The resulting model is analyzed in relation to the functional state and potential binding sites. The comparison of modeled distances to distances reported in experimental studies between particular residues suggests that the model corresponds most closely to the first ATP hydrolysis step of the protein transport cycle. Comparison to the protein 3D structure confirms this suggestion. Using SiteID and Site Finder programs three membrane related binding regions are identified: a region at the interface between the membrane and cytosol and two regions located in the transmembrane domains. The regions contain binding pockets of different size, orientation, and amino acids. A binding pocket located inside the membrane cavity is also identified. The pockets are analyzed in relation to amino acids shown experimentally to influence the protein function. The results suggest that the protein has multiple binding sites and may bind and/or release substrates in multiple pathways.

Introduction

In recent years significant progress has been made towards understanding the role and way of functioning of the multidrug resistance (MDR) transport proteins of the ABC superfamily.[1] Among the transporters studied P-glycoprotein (P-gp) has been paid special attention because of its key role in a number of cellular processes. Being naturally expressed in excreting organs and epithelial membranes P-gp is involved in the xenobiotic protection of the cells. Its substrates cover a broad spectrum of compounds of different structure and pharmacological function including antibiotics, HIV protease inhibitors, steroids, chemotherapeutic, and immunosuppressive drugs.[2] In tumor cells the protein recognizes a large variety of different antineoplastic agents (anthracyclines, *Vinca* alkaloids, taxanes) as substrates for an ATP-dependent efflux, minimizing their intracellular concentrations and, consequently, decreasing the therapeutic effect.[3] Many compounds have been investigated for their ability to inhibit the P-gp function, thus leading to development of several generations of MDR-modulators.[4]

Identification of the binding sites of the P-gp substrates and inhibitors has always been within the focus of research interest. Elucidation of the binding regions and their amino acid residues is a key step in understanding the molecular basis of drug transport and function of P-gp as a MDR transporter. Although various experimental data have been accumulated in the last decade, the number and location of the drug binding sites remains unclear. Initially a large common binding site was assumed,[5] later a minimum of two binding sites was proposed to explain the complex behavior of P-gp when cooperative, competitive, and noncompetitive interactions between MDR modulators are observed.[6] From experimental studies with P-gp containing proteoliposomes Shapiro and Ling postulated two distinct binding sites, the rhodamine (R)- and Hoechst 33342 (H)- sites, which interacted in a positively cooperative manner[7] and a third, regulatory binding site for progesterone and prazosin.[8] From radioligand binding experiments at least four different binding sites were inferred that were able to allosterically communicate.[9] Summarizing data obtained by competition with photolabeling drugs Safa proposed seven binding sites that partly interacted with each other: a vinblastine binding site, which also bound verapamil and cyclosporine A; a taxol binding site; a binding site for dihydropyridine type calcium channel blockers; a binding site for bepridil, prenylamine, and megesterol acetate; for flupentixol; for prazosin-like structures; and Hoechst 33342.[10] Loo and Clarke reported many results from cross-linking experiments on the interaction and transport characterization of P-gp.[11–14] Differences in the cross-linking patterns observed upon binding of different drugs led them to suggest that P-gp could accommodate varying substrates through an induced-fit mechanism. The authors proposed that the transmembrane domains TM2/TM11 and TM5/TM8 between the two halves must enclose the drug-binding pocket at the cytoplasmic side of P-gp.[15,16] Using photoaffinity labeling and MALDI-TOF mass spectrometry[17] Ecker et al. reported that TM3, TM5, TM8, and TM11 were mostly labeled by the benzophenone-type photoaffinity ligands.[18]

Thus, the localization and structure of the ligand binding sites continues to be an open question. In the absence of 3D

[a] *C. Globisch, Prof. Dr. I. K. Pajeva, Prof. Dr. M. Wiese*
Pharmaceutical Institute, University of Bonn, 53121 Bonn (Germany)
Fax: (+49) 228-737929
E-mail: mwiese@uni-bonn.de

[b] *Prof. Dr. I. K. Pajeva*
Center of Biomedical Engineering
Bulgarian Academy of Science, 1113 Sofia (Bulgaria)

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

structural data of a protein–substrate or a protein–inhibitor complex, studies using P-gp molecular models could contribute to this topic. Recently the crystal structure of the bacterial ABC transporter from *Staphylococcus aureus* Sav1866 has been resolved at 3 Å[19,20] and the first biochemical evidence has been provided supporting the correspondence in the domain arrangement between Sav1866 and P-gp as MDR exporters. Using cysteine mutagenesis and cross-linking Zolnerciks et al.[21] showed that the long intracellular loops of TM2 contacted the first nucleotide binding domain (NBD) confirming in this way the physiological relevance of the Sav1866 structure. Thus, the 3D structure of Sav1866 can be considered as an appropriate template for modeling of the human P-gp.

In this paper we report a homology model of P-gp based on the X-ray structure of Sav1866. The model has been incorporated into a membrane environment (phospholipids, water, and ions) and optimized to closely simulate the real environment of the protein. The resulting structure has been analyzed in relation to its correspondence to the functional state of the protein transport cycle. The model has further been used for generation of protein binding pockets. The binding regions identified have been analyzed for the amino acids involved and their role for the drug interaction and transport. The modeling results point to the possibility of multiple binding sites and multiple pathways for drug transport.

Results

Homology model of P-gp

To explore the influence of the force field and the environment on the homology structure five models were generated (see Computational Methods): optimized with CHARMM27 in vacuum (c1); with an OPLS-AA based force field in vacuum (c2) and in a membrane environment (c3); with GROMACS implemented force field (ffgmx) in vacuum (c4), and in a membrane environment (c5). Table 1 reports the comparison between the structures of Sav1866 and the P-gp models generated. Table 1 part A shows the root mean squared deviation (RMSD) values of matching the C-alpha atoms for the X-ray and optimized structures of Sav1866 and the CHARMM27 optimized structures of P-gp (initial and final models, see Computational Methods). The RMSD value of 1.32 Å for matching all residues is rather satisfactory, as it is well below the average value of 1.5 Å according to the relationship between the expected RMS deviation and sequence identity.[22] There are also no profound differences between the NBD and the TMD parts of the structures (Supporting Information Table S1). If the TMD-NBD loops (P-gp sequences 372–383 and 1014–1026) are excluded from comparison, RMSD-values lower than one are recorded. Table 1 part B reports the RMSD values of matching between all heavy atoms for the five P-gp models. The model obtained with OPLS-AA in vacuum deviates less from the CHARMM27 optimized one (0.34 Å) compared to the ffgmx optimized model (0.49 Å). The deviations between the models from optimization in vacuum and in the membrane are higher for ffgmx than for OPLS-AA, respectively 0.29 Å and 0.20 Å, suggesting that the OPLS-AA optimized model is closer to the CHARMM27 optimized one. Additionally the distances between particular amino acids were measured in all five models and very similar values were recorded indicating that the force field used did not significantly influence the geometry of the model structure (Supporting Information Table S1). Considering the small differences in the models generated, the OPLS-AA membrane optimized model was used for further analysis. We took into account that the OPLS-AA force field performs better in a lipid environment (the ffgmx force field tends to overestimate the interactions between the lipids and the protein[23]) and that the model optimized in a membrane environment may more closely simulate the real structure of the protein.

Table 1. RMSD values of matching C-alpha atoms (part A) and C-alpha, C-beta, C, N, and O atoms (part B).

A	1[a]	2[b]	3[c]	4[d]
1. Sav1866 X-ray	0.00			
2. Sav1866 X-ray min	0.71	0.00		
3. Initial P-gp model	0.74	1.02	0.00	
4. Final P-gp model	1.32	1.29	1.11	0.00

B	c1[e]	c2[f]	c3[g]	c4[h]	c5[i]
c1	0.00				
c2	0.34	0.00			
c3	0.21	0.20	0.00		
c4	0.49	0.44	0.45	0.00	
c5	0.35	0.29	0.32	0.29	0.00

[a] The X-ray structure of Sav1866; [b] After optimization with CHARMM27; [c] Initial P-gp model (after crude minimization with CHARMM27); [d] Final P-gp model (see Computational Methods). P-gp models optimized with: [e] CHARMM27 in vacuum; [f] OPLS-AA in vacuum; [g] OPLS-AA in a membrane environment; [h] ffgmx in vacuum; [i] ffgmx in a membrane environment.

Figure 1 A presents the homology model of P-gp based on the Sav1866 structure and Figure 1 B shows the model incorporated in the membrane with the electrostatic potential mapped on the Gaussian contact surface (see Computational Methods). Clearly seen are the neutral parts of the helices buried into the membrane. Inspection of the helices points to differences in their lengths located in the membrane compared to those reported in the Swissprot database[24] (Table 2). In most cases the lengths of the membrane located helices exceed the Swissprot lengths of the TMs; the extracellular domains between TM3-TM4 (EC2) and TM9-TM10 (EC5) are fully buried in the membrane.

Estimation of the putative functional state

According to Dawson and Locher[19] Sav1866 was crystallized in the presence of ADP, however, the conformation, at the resolution used, was undistinguishable from that of the isolated NBDs containing trapped ATP. The authors concluded that al-

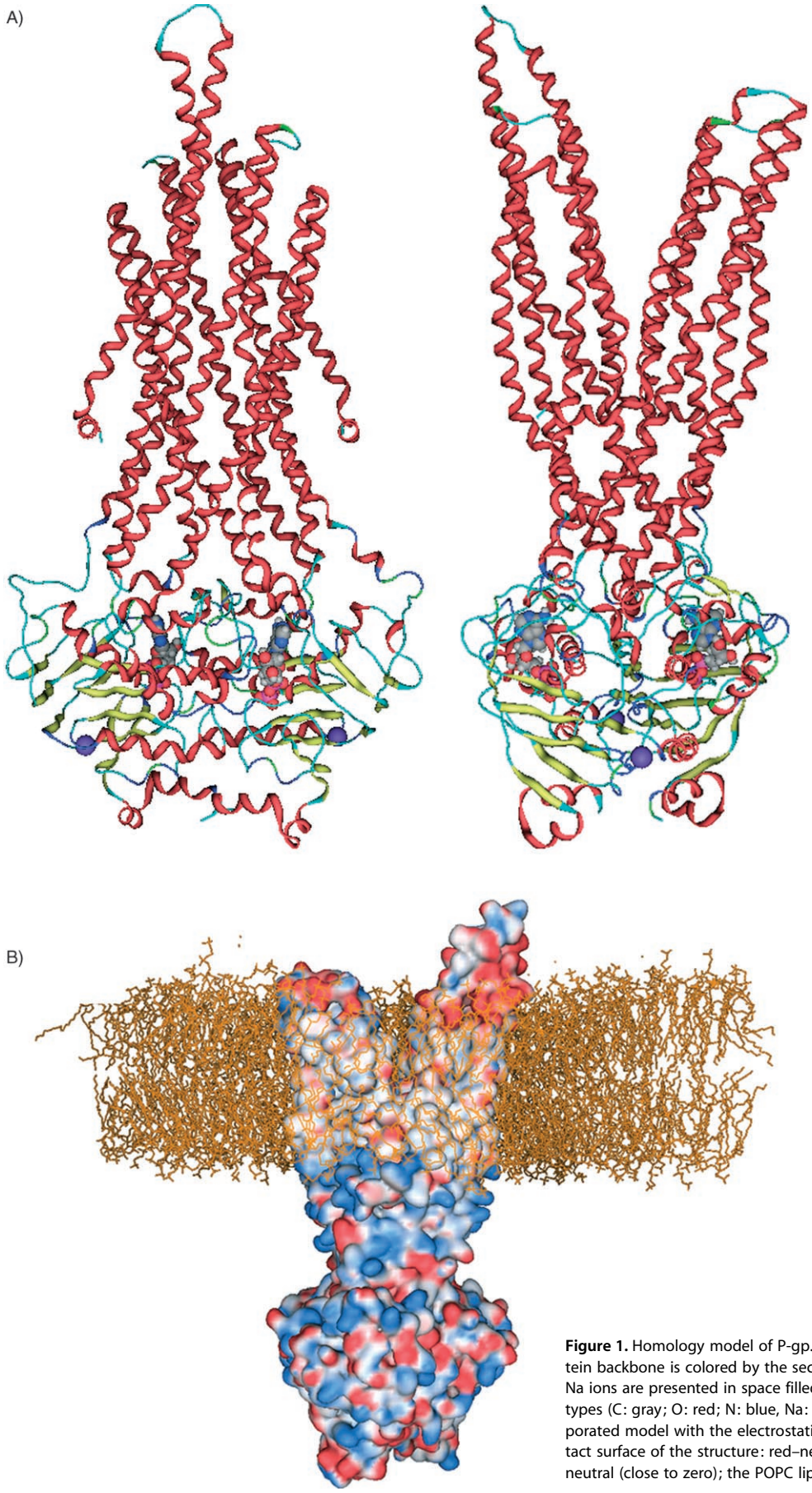

Figure 1. Homology model of P-gp. A) Face views (front and side); the protein backbone is colored by the secondary structure; the ADP molecules and Na ions are presented in space filled form and colored according to the atom types (C: gray; O: red; N: blue, Na: big blue spheres); B) A membrane incorporated model with the electrostatic potential mapped on the Gaussian contact surface of the structure: red–negative values; blue–positive values; gray–neutral (close to zero); the POPC lipids are orange.

Table 2. Alignment of the TM helices.[a]

TM domain	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM8	TM9	TM10	TM11	TM12
Swissprot	52–72	120–140	189–209	216–236	297–317	326–346	711–731	757–777	833–853	854–874	937–957	974–994
Model	49–76	104–138	186–…	…–233	292–323	328–354	707–734	745–777	829–…	…–879	934–965	971–1000

[a] According to the Swissprot database and approximate locations of the TM residues estimated from the membrane incorporated homology model of P-gp (Figure 1 B).

though ADP rather than ATP was bound, the NBDs in Sav1866 exhibited the conformation of the ATP bound state. In a later study a crystal structure of Sav1866 in a complex with AMP-PNP (nonhydrolysable ATP analogue) at 3.4 Å was resolved and the comparison with the previously determined structure of Sav1866 with bound ADP revealed no significant conformational changes.[20] The authors concluded that the X-ray structure represented the ATP-bound state of the transporter in agreement with the outward-facing conformation of the transmembrane domains.

Table 3. Comparison of the distances between amino acids recorded in different functional states of P-gp.[a]

Cross-linked amino acid pairs		Cross-linker/length [Å]	Modeled distance [Å]	Functional state
L332(TM6)	Q856(TM10)	M6M, M8M, M11M 18.0–23.6	33.5	no ATP + cross-linker
L332(TM6)	L976(TM12)	M5M, M6M, M8M 16.8–20.0	20.2	ATP/VO_4 + cross-linker 1st ATP hydrolysis
L332(TM6)	L975(TM12)	disulfide bond	19.4	2nd ATP hydrolysis

[a] In the experiments the amino acids were mutated to cysteine and cross-linked with thiol-reactive compounds (see Figure 3 and Table 2 in ref. [26]); in the model the distances were measured between the C-alpha atoms of the amino acids in the OPLS-AA optimized model of P-gp in a membrane environment; the lengths of the cross-linkers are corrected by adding 5.5 Å (Figure 2).

Considering the above facts we tried to more precisely analyze the putative functional state of P-gp corresponding to our model. Two ways were explored: 1) the inter-distances between the C-alpha atoms of different residues were measured and compared to distances that have been reported for different functional states of P-gp in experimental studies; 2) the P-gp model was qualitatively compared to the 3D structure of P-gp resolved at 8 Å.[25]

First the distances between the C-alpha atoms of the residues L332(TM6) and Q856(TM10), L975(TM12), and L976(TM12) shown to be cross-linked during different steps of the transport cycle of the protein[26] were measured. The distances were than compared to the lengths of the cross-linkers – thiol-reactive methanesulfonate derivatives, MnM. The lengths of the cross-linkers were defined based on their structures[12] and their conformations were built and optimized as described in our previous study.[27] Figure 2 shows the cross-linkers used and their lengths. An additional length of approximately 5.50 Å (2×2.75 Å) was added to each linker length as the distance between the C-alpha atoms of the residues were compared. The results are summarized in Table 3. As seen from the table the distance between L332 and L976 is close to the distance corresponding to the functional state of the first ATP molecule hydrolysis. In the model L332 and L976 as well as L975 faced each other and could be cross-linked; in contrast Q856 faced a direction opposite to L332.

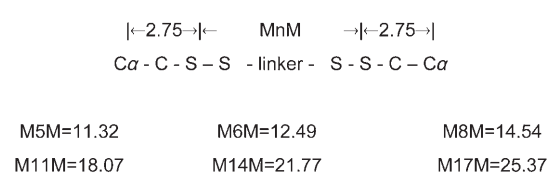

Figure 2. Lengths (in Å) of the cross-linkers; the distances between the cysteine S atoms correspond to the extended optimized conformations of the MnM structures[27] (structures of the MnM cross-linkers are taken from ref. [12]).

Table 4 reports the experimental and modeled distances for several more cross-linked residues registered in Cys-P-gp mutants. Using no ATP and a nonhydrolysable analogue of ATP (AMP-PNP) Loo and Clarke[26] found no difference in the cross-linking patterns for the reported amino acids. However, as noticed by themselves,[12] at the concentrations used, the cross-linkers were able to stimulate the ATP activity of Cys-less P-gp by about 50% thus, suggesting that the cross-linking may also introduce some conformational changes in the protein structure. In Table 4 the lengths of the shortest linkers able to cross-link the particular residues are reported. In general, as seen from the table, there is no correlation between the experimental distances and those measured in the model. It is notable that the biggest difference is recorded for the pairs S222(TM4)-I868(TM10) and S222(TM4)-G872(TM10): in the cross-linking experiments they are about 17 Å apart from each other and in the model this distance is above 40 Å. This observation shows that TM4 and TM10 may undergo a big repositioning during the protein transport cycle.

Furthermore the distances between TM2 and TM11, on one side, and TM5 and TM8, on the other side, were compared. Amino acids from these domains are shown to be directly linked by oxidative cross-linking with copper phenanthroline of human P-gp mutants and this linking is inhibited during the vanadate trapping of the nucleotide or the presence of some drug substrates.[15,16] Vanadate trapping is considered a possibility to characterize the first hydrolysis step because P-gp is stably inhibited upon trapping of ADP at a single catalytic site.[26] Table 5 presents the correspondence between the reported and modeled distances. The experimental distances

Table 4. Comparison between the experimental[26] and modeled distances.[a]

Amino acid pairs	Cross-linker/length [Å]	Modeled distance [Å]
S222 (TM4)–I868 (TM10)	M5M / 16.8	40.7
S222 (TM4)–G872 (TM10)	M5M / 16.8	40.7
I306 (TM5)–I868 (TM10)	M8M / 20.0	34.6
I306 (TM5)–G872 (TM10)	M8M / 20.0	34.8
I306 (TM5)–T945 (TM11)	M8M / 20.0	32.2
I306 (TM5)–V982 (TM12)	M8M / 20.0	23.1
I306 (TM5)–V984 (TM12)	M8M / 20.0	25.3
L339 (TM6)–I868 (TM10)	M8M / 20.0	30.7
L339 (TM6)–G872 (TM10)	M8M / 20.0	32.4
L339 (TM6)–F942 (TM11)	M17M / 30.1	27.3
L339 (TM6)–T945 (TM11)	M14M / 27.3	28.5
L339 (TM6)–V982 (TM12)	M11M / 23.6	16.6
L339 (TM6)–V985 (TM12)	M14M / 27.3	19.7

[a] Only the shortest from the reported linkers are shown; the distances are measured between the C-alpha atoms of the amino acids; the lengths of the linkers are corrected by adding 5.5 Å (Figure 2).

Table 5. Comparison of the modeled and experimental distances between the pairs TM2-TM11, TM5-TM8, and TM1-TM11.[a]

Cross-linked amino acids	Cross-link distance [Å]	Modeled distance [Å]
V133 (TM2)–R939 (TM11)	5–7	6.0
C137 (TM2)–A935 (TM11)	5–7	6.3
N296 (TM5)–G774 (TM8)	5–7	6.3
I299 (TM5)–G774 (TM8)	5–7	9.4
I299 (TM5)–F770 (TM8)	5-7	8.3
G300 (TM5)–F770 (TM8)	5–7	6.0
M68 (TM1)–Y950 (TM11)	5–7	10.0
M68 (TM1)–Y953 (TM11)	5–7	8.4
M68 (TM1)–A954 (TM11)	5–7	8.4
M69 (TM1)–A954 (TM11)	5–7	11.8
M69 (TM1)–F957 (TM11)	5–7	10.9

[a] Cross-linked by copper phenanthroline.[15,16,28] In the model the distances between the C-alpha atoms of the amino acids are measured.

correlate well with those measured in the model, indicating that the conformation represents the functional state close to the first step of ATP hydrolysis.

Additionally, the possibility for cross-linking of the cysteine-mutated residues M68 to Y950, Y953, and A954, as well as of M69 to A954 and F957 was explored in the model. As reported by Loo and Clarke these residues could be cross-linked with copper phenanthroline in the presence of ATP but not in the presence of ADP or AMP-PNP.[28] Table 5 reports the modeled distances. Analysis of the residues' location in the model reveals that Y950, Y953, A954, and F957 all face the same direction being well exposed to M68; in contrast, the cross-linking to M69 although possible (considering the flexibility of the side chain) could be more difficult in agreement with the larger differences in the distances of this amino acid compared to the experimental ones (Table 5). This implies that the model may not exactly represent the suggested ATP-bound functional state upon which the cross-linking of these residues has been performed.

Supporting Information Table S2 reports the correspondence of the model distances to the experimentally reported ones for all five P-gp homology models. Additional to the Loo and Clarke experimental data, the distances between TM2-TM11 and TM5-TM8 defined in the cross-linking experiments of Stenham et al.[29] are also compared. A length of about 13 Å has been recorded between the positions of the S atoms in the structure of the cross-linker *N*,*N*-phenylenedimaleimide used (built and optimized similarly to the above mentioned MnM linkers). The modeled distances varied from 8.2 Å up to 16.3 Å, thus showing a relatively good correlation with the experimental distance. Together with the distances reported in Table 5 this comparison confirms that TM2 remains in close proximity to TM11 and TM5 is also in close proximity to TM8 in the presumed functional state.

The homology model was further qualitatively compared to the protein model created with cryoelectron microscopy data of the P-gp 3D structure resolved at 8 Å.[25] We analyzed the possible correspondence of the helices in both models and assigned them accordingly. Figure 3 represents both models: in the homology model TMs are shown in colors similar to those used by Rosenberg et al.[25] The suggested TM assignment is based on the fact that the microscopic model corresponds to the nucleotide-bound state (AMP-PNP) of P-gp, in which the protein has already undergone significant conformational changes (see Discussion). Generally, there is a good corre-

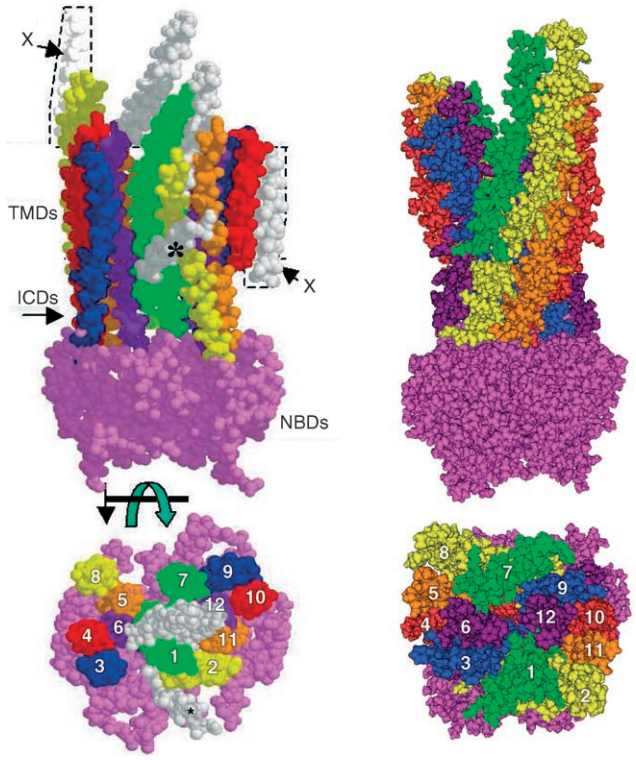

Figure 3. Qualitative comparison of the space-filling displays of the P-gp side and top views with numbered TMs. A) The model in the nucleotide-bound form based on cryoelectron microscopy data (extracted and adapted from ref. [25]. B) The homology model based on Sav1866 structure; purple: TM6 and TM12; orange: TM5 and TM11; red: TM4 and TM10; blue: TM3 and TM9; yellow: TM2 and TM8; green: TM1 and TM7; NBDs: magenta.

spondence between both models, suggesting that the homology model has a reasonable structure. However, despite the general similarity, differences are seen in the conformations of the helices as well as in the NBDs. Based on these comparisons we conclude that the homology model of P-gp most probably represents a conformation closer to the first step of ATP hydrolysis rather than a conformation corresponding to the ATP-bound functional state.

Analysis of the protein surface properties

We further investigated the surface properties of the protein model. Both, the outside and inside (cavity) surfaces were analyzed by generating their lipophilic potentials. Figure 4 illustrates these potentials mapped on the Connolly surfaces of the protein as well as on the Connolly channel surface of the cavity. The hydrophobic character is well seen on the membrane facing surface of the protein (Figure 4A and B, brown regions). The Connolly channel surface illustrates a large intramolecular cavity that is neutral to hydrophobic in the TM part (Figure 4C).

Identification of the binding regions and pockets

For identification of the protein binding sites two different approaches have been combined. First, the mode "Find pockets" with the "Grid" method in SiteID[30] was applied to get a general presentation about the distribution and localization of the pockets in the structure. Next, the module "Site Finder"[31] was employed to get a more precise characterization of the potential binding sites. Below the term "binding region" will be used as representing an aggregate of binding pockets; "binding pocket" will mean pockets identified by SiteID and Site Finder; "binding site" will mean a site where a substrate or an inhibitor may bind to the protein (in this sense "binding site" may be composed of one or more binding pockets).

Figure 5 illustrates the binding pockets found. SiteID identified 40 non-overlapping pockets belonging to different structural units of the protein (Figure 5A). Three main TM related binding regions can be outlined: binding region 1 located at the interface between the membrane and cytosol, and binding regions 2 and 3 located in the transmembrane part of the protein. The Site Finder (Figure 5B) generally resembled the SiteID pockets and additionally identified a big cavity pocket (Figure 5C). A number of binding pockets were also found in the NBD part of P-gp (shown only in Figure 5A). Considering that there might be a number of apparent binding pockets that could be a result of packing problems during the homology modeling and minimization procedures we further analyzed the reasonableness of the binding regions identified for presence of amino acids shown in mutation, cross-linking, and photolabeling experiments to affect the transport function or substrate specificity of P-gp.

Analysis of the binding regions and cavity pocket

Binding region 1. Table 6 reports the amino acids found in the pockets of binding region 1 and Figure 6 illustrates the region. The residues are predominantly from the intracellular domains IC1 (TM2/TM3), IC2 (TM4/TM5), IC3 (TM8/TM9), and IC4 (TM10/TM11). Single amino acids belonging to the lower parts of TM2, TM5, TM8, and TM11 as well as to the linkers of TM6 and TM12 to the NBDs are also involved. The residues that have been investigated for their effect on the transport function of P-gp and its substrate specificity in experimental studies are given in bold in Table 6. Supporting Information Table S3 reports a detailed list of the experimentally studied amino acids with the corresponding references. In Figure 6 the experimentally studied amino acids are labeled, additionally several other residues are also labeled that are located in close proximity to the experimentally studied ones.

Kwan and Gros investigated the influence of random mutations in IC1 and the flanking domains of the mouse P-gp and found a segment of four consecutive amino acids in positions 169–172 (human T173, L175, T176, are shown in Table 6, Table S3) of the central portion of IC1, in which mutations caused a severe overall loss of the protein function.[32] The authors suggested that these mutations could influence drug binding, or, alternatively, they could reduce the basal or inducible ATPase activity of P-gp. A number of mutations in the region have also been found to affect the substrate specificity of different cytotoxic agents. Loo and Clarke performed several glycine to valine mutations in the cytoplasmic loops of human P-gp, including G141, and demonstrated that these mutations increase relative resistance to colchicine and doxorubicin, but do not alter the resistance to vinblastine.[33] Mutations in an eight amino-acid segment of murine IC2 238–245 (239–246 in human P-gp) have been shown to cause reduction in the resistance to vincristine and paclitaxel and to have no influence on daunorubicin and colchicines.[34,35] Hana et al.[36] mutated a number of residues in TM11 of P-gp by alanine and found that the mutation in positions G935 and I936 (G939, I940 in human P-gp) displayed an overall decrease in resistance to actinomycin, doxorubicin, vinblastine, and colchicine. These amino acids have not been found in the pocket, however, close to them F938 has been identified. Furthermore four amino acids (L281, A284, K285, and G288) constitute a part of the so-called EAA motif in IC2 (278–294) that is considered to mediate the interactions between the TMs and NBDs.[37] Ecker et al.[17] identified several protein parts that contributed mostly to the binding domains of propafenones: 272–286 (IC2), 755–784 (TM8), and 789–798 (IC3), of them residues L281, A284, K285, E782, and R789 are involved in the pocket. In this region, mutation of F938 to alanine has been shown to cause a moderate but reproducible 2–3-fold decrease in resistance to actinomycin D, doxorubicin, colchicine, and vinblastine.[36] In the same region MTS-rhodamine has been able to inhibit by 51% the S993C mutant of P-gp.[13]

Binding region 2. Table 7 reports the amino acids identified in the binding pockets of binding region 2. The residues belong to TM3, TM4, TM5, TM6, and TM8, as well as to the extracellular

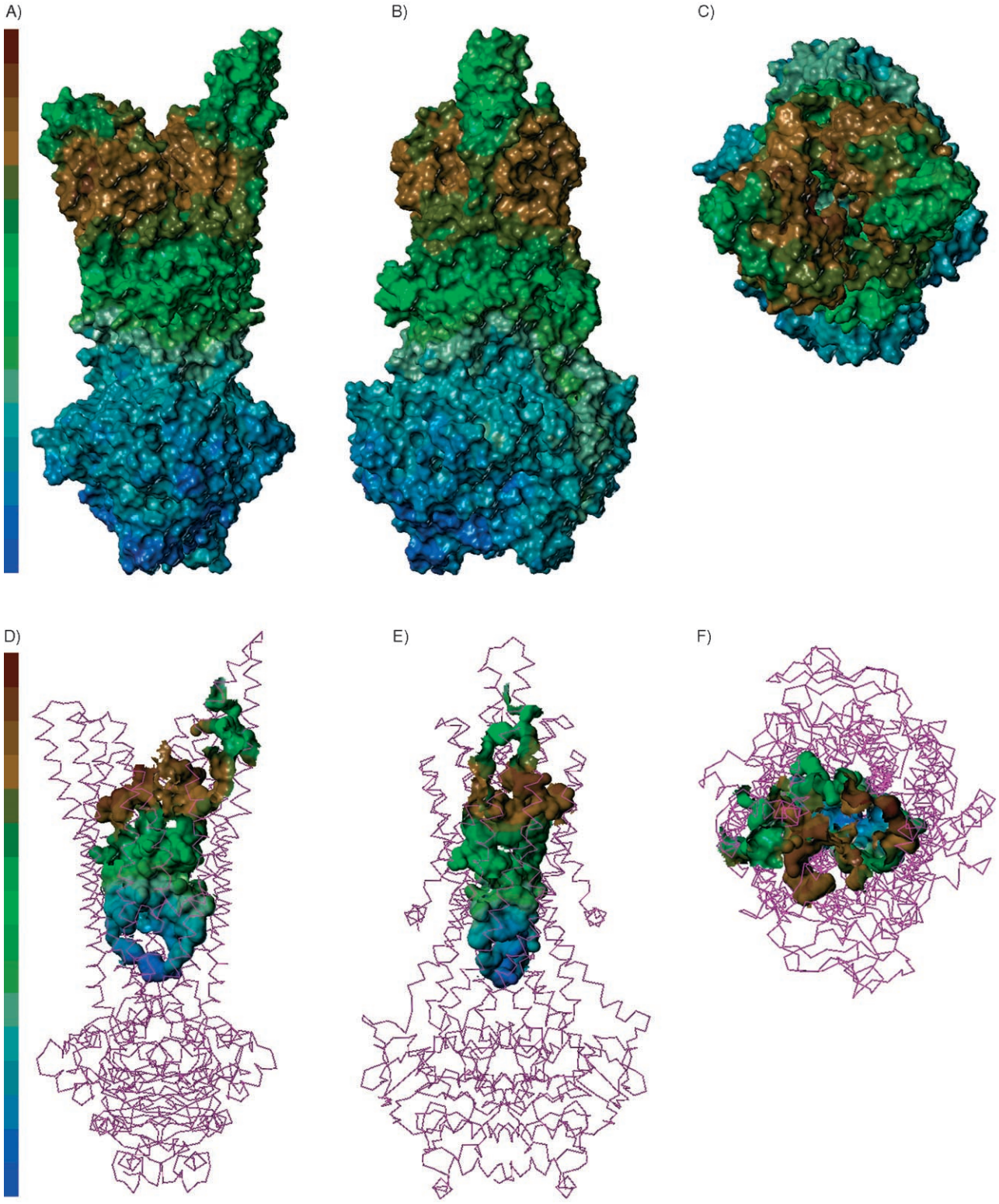

Figure 4. Lipophilic potentials of the P-gp model mapped on Connolly surfaces (A, B, and C) and Connolly channel (D, E, and F). Left: side view; Middle: front view; Right: top view (from outside, corresponds to side view); brown: highest hydrophobic area; blue: highest hydrophilic area; magenta: the Cα-chain of the protein.

loops EC2 (TM3/TM4), EC3 (TM5/TM6), and EC4 (TM7/TM8). Those mostly contributing are TM5 and TM6, each participating with 13 residues. Figure 7 illustrates binding region 2. Similarly to binding region 1, many of these residues have been experimentally shown to affect the transport function of P-gp and its substrate specificity (shown in bold in Table 7, see also Table S3 and Figure 7). A number of amino acids (196–207, 304–315, 330–342, 758–767, Table 7) belong to the sequences

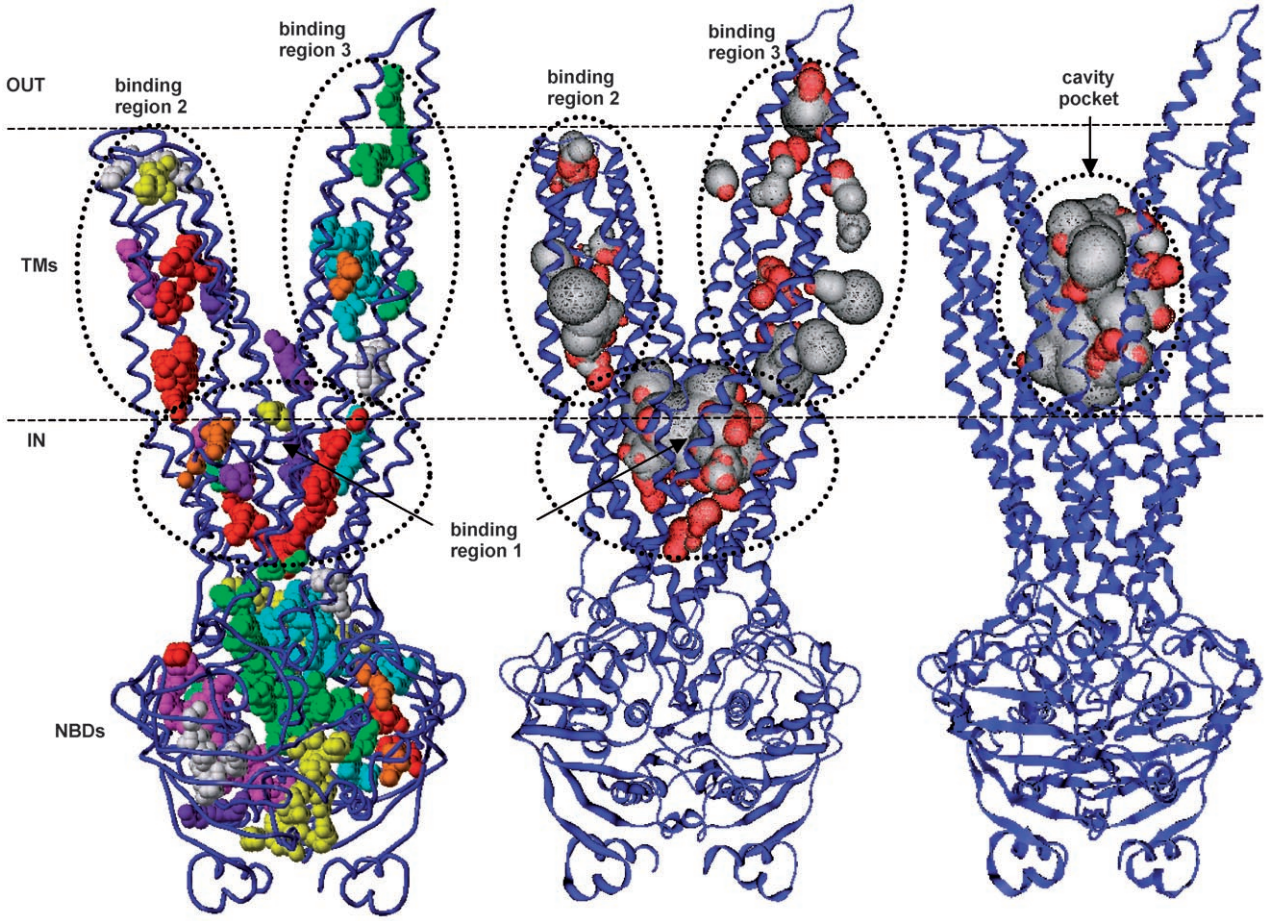

Figure 5. Binding regions of P-gp. A) Binding pockets identified by SiteID;[30] the pockets are filled with spheres set around the oxygen atoms of the water molecule; the same color is used for filling the spheres belonging to the same pocket. B and C) Binding pockets identified by Site Finder;[31] the pockets are filled with alpha spheres; gray spheres indicate hydrophobic atoms and red spheres, hydrophilic atoms; the protein backbone is shown as a blue tube; the membrane-related binding regions and the cavity pocket are outlined with dotted lines; the horizontal dotted lines show the approximate borders of the membrane.

194–208, 304–316, 327–343, and 755-784 related to the binding domains of propafenones, T199, and S309 have been suggested to act as a HB-donor for these compounds.[17,18] W212, K213, and L214 correspond to mouse W208, K209, and L214 (Table S3): these residues are found to be sensitive to alterations and critical for P-gp functioning.[32] In this region cross-linking of cysteine residues located in positions S222, I306, I340, and A342 have been protected by MTS-rhodamine, MTS-verapamil, and dibromobimane.[13] Furthermore, several mutations in TM6 have been found to modulate the activity and substrate specificity of human P-gp: mutation of V338 causes enhanced resistance to colchicine and reduced relative resistance to vinblastine; mutant G341 to valine confers decreased resistance to colchicine or doxorubicin, while resistance to vinblastine or actinomycin D

Table 6. Amino acids (AAs) involved in binding region 1.[a]

TM2: 4 AAs				IC1: 16 AAs							
W136	**C137**	**L138**	A140	**G141**	I144	H145	K146	R148	G169	N172	**T173**
								IC2: 16 AAs			
L175	**T176**	D177	**S180**	**K181**	I182	**N183**	**E184**	T240	E243	L244	Y247
G251	A252	A254	E255	**L281**	**A284**	**K285**	**G288**	**K291**	A292	A295	**N296**
TM5	Linker TM6-NBD: 6 AAs						TM8	IC3: 9 AAs			
I299	P350	E353	A354	N357	A358	A361	**G774**	**G778**	**G781**	**E782**	T785
					IC4: 17 AAs						
K786	**R789**	T815	A819	N820	A883	D886	K887	L890	E891	A893	G894
										TM11	
A897	T898	Y920	L924	P927	Y928	S931	L932	K934	**A935**	**F938**	
Linker TM12-NBD: 8 AAs											
S993	P996	D997	A999	K1000	A1001	I1003	S1004				

[a] AAs investigated by mutations, cross-linking, and photolabeling experiments are shown in bold (see also Table S3); AAs shown in Figure 6 are underlined; TM helices are assigned according to Swissprot (see Table 2).

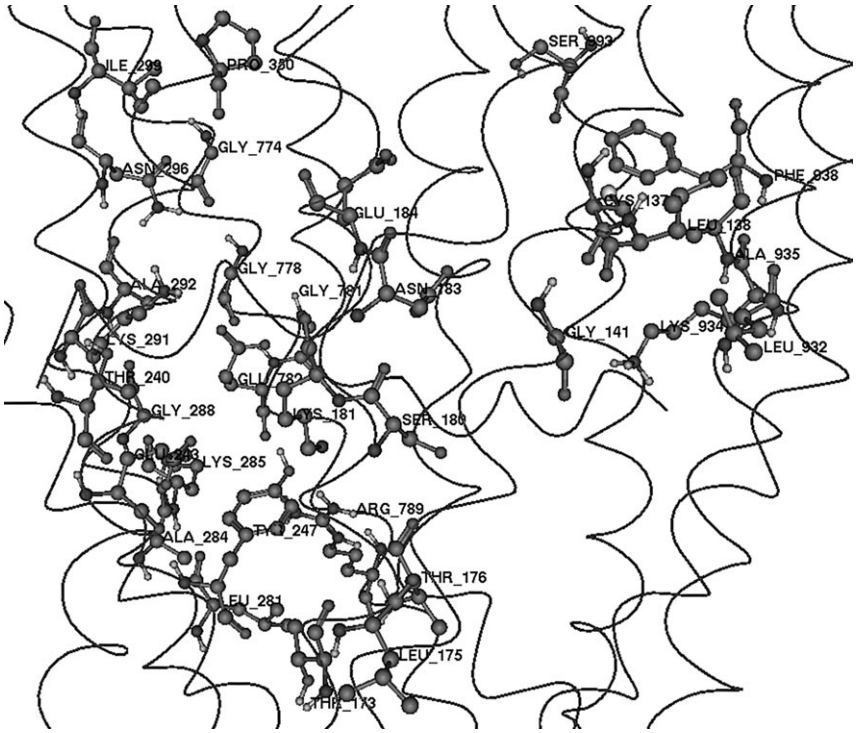

Figure 6. Binding region 1 with labeled amino acids shown in experimental studies to affect the transport function of P-gp (see Table 6); the orientation of the helices corresponds to Figure 5B; the residues are presented as balls and sticks (the nonpolar hydrogens are hidden).

Table 7. Amino acids (AAs) involved in binding region 2.[a]

TM3: 9AAs									EC2: 6AAs		
G193	**S196**	**M197**	**T199**	**F200**	**F201**	**G203**	**F204**	**G207**	R210	G211	**W212**
			TM4: 10 AAs								
K213	**L214**	**T215**	**I218**	L219	I221	**S222**	L225	G226	S228	A229	**W232**
	TM5: 13AAs										
L236	S298	**I299**	A301	**A302**	**L304**	**L305**	**I306**	**Y307**	**A308**	**S309**	**A311**
		EC3: 13AAs					TM6: 13 AAs				
L312	**W315**	**T318**	**T319**	V321	**L322**	E325	Y326	**Q330**	**V331**	**V334**	**S337**
								EC4: 6 AAs			
V338	**I340**	**G341**	**A342**	**S344**	V345	A348	S349	I735	F739	T740	T747
		TM8: 5AAs									
K748	N751	**L758**	**F759**	**L762**	**G763**	**F767**					

[a] AAs investigated by mutations, cross-linking, and photolabeling experiments are shown in bold (see also Table S3); AAs shown in Figure 7 are underlined; TM helices are assigned according to Swissprot (see Table 2).

is retained; A342 mutation results in reduction of P-gp ability to confer resistance to the above four drugs; finally mutation of S344 has been shown to reduce drug stimulated ATPase activity of P-gp.[38] Mutation of I299 to methionine, T319 to serine, and L322 to isoleucine decreases resistance to colchicine and taxol.[39] Mutations of murine W231, A301, S308 in TM4 and TM5 (human W232, A302, S309, Table 7) influence the resistance profile of the P-gp mutants in relation to steroids and are distinct from those related to *Vinca* alkaloids (human F239, K242, and A246 binding region 1) located within the cytoplasmic portion of the protein.[34]

Figure 8 shows binding region 2 as generated by Site Finder. The front views from the membrane (Figure 8A), the pore (Figure 8B), and the top view from outside the cell (Figure 8C) clearly show that the region is composed of pockets that face the membrane (identified by the gray spheres representing hydrophobic atoms) as well as the pore (identified by the red spheres representing hydrophilic atoms).

Binding region 3. Table 8 summarizes the residues identified in the binding pockets of binding region 3; the residues found in different experimental studies to affect the function of P-gp are presented in bold (see also Table S3). Figure 9 illustrates binding region 3. The amino acids from TM1, TM2, TM10, TM11, and TM12 as well as from the extracellular loops EC1 (TM1/TM2) and EC6 (TM11/TM12) are involved. As seen from Figure 5A and B a pocket partially located outside the membrane (green color) has been found, correspondingly a number of amino acids that belong to the extracellular loop EC1 are involved (Table 8). For this part of the loop (78–97) deletion mutants show dramatically altered drug-stimulated ATPase activity for different P-gp substrates. Calcein AM has only a little effect; rhodamine 123 and vinblastine are not able to stimulate the mutant at all. For verapamil the deletion is not so critical showing the same level of maximum stimulation but strong decrease in affinity by a factor of 100.[40] A number of amino acids involved in this region are also labeled by P-gp substrates. MTS-rhodamine labels residues Y118, V125, G872, S943 showing at least 50% inhibition; however, these residues are not protected from labeling by pre-incubation with rhodamine. This could indicate that they are accessible by substrates such as MTS-rhodamine but are not close enough to the rhodamine binding site. The amino acid A841 is not involved in the pocket but is adjacent to the identified I840; A841C is labeled by MTS-rhodamine but, in contrast to the above mentioned residues, it could be protected by pre-incubation with rhodamine. The residues L975C and V981C are also labeled by MTS-rhodamine and protected through pre-incubation by rhodamine, suggesting that A841, L975, and V981 should be close to

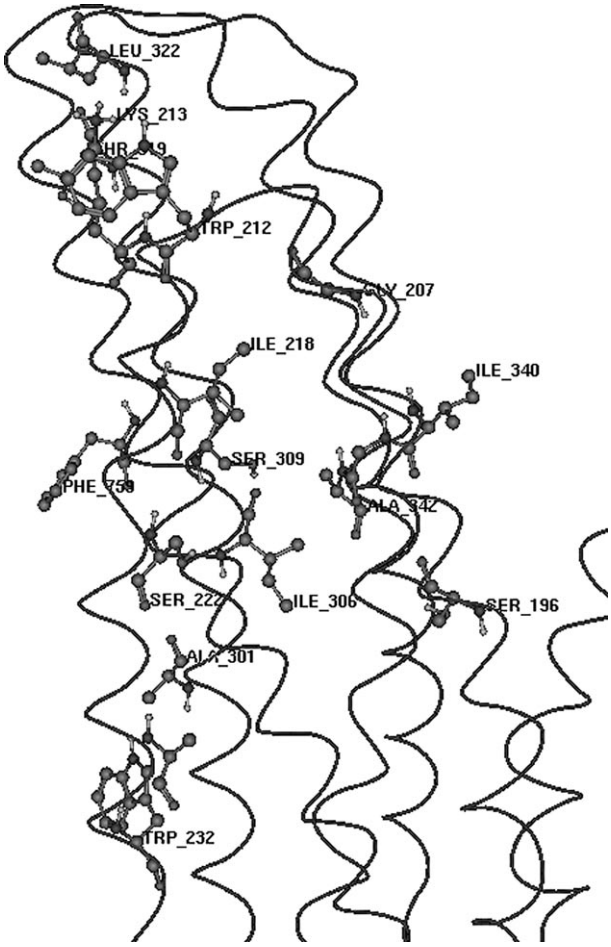

Figure 7. Binding region 2 with labeled amino acids shown in experimental studies to affect the transport function of P-gp (see Table 7); the orientation of the helices corresponds to Figure 5B; the residues are presented as balls and sticks (the nonpolar hydrogens are hidden).

the binding domain of rhodamines. MTS-verapamil labels cysteine-mutated residues I868, F942, Q945, and G984, and dibromobimane labels the residues I868, G872, Q945, L975, and A985. In case of MTS-verapamil the amino acids could be protected from cross-linking by pre-incubation with verapamil and in the case of dibromobimane by pretreatment with verapamil, vinblastine, or colchicine.[11,13,41] The double mutation of human I840 and N842 (hamster I837 and N839) together with G341 and A342 from TM6 show a synergistic effect on the resistance towards actinomycin D suggesting that TM6 cooperates with TM9 to mediate drug resistance.[42] In alanine scanning-mutagenesis of the TM11 helix in mice (pockets involved are 939–940, 942, 945, 947, 950, 952, and 953) changes in the resistance levels of the mutants towards vinblastine, doxorubicin, colchicine, and actinomycin D is are recorded: Y949A and F953A (human Y953 and F957) cause a 5–10-fold reduction in resistance to actinomycin D, colchicines, and daunorubicin and in the case of F953A to vinblastine.[36] For seven conserved residues in TM12 L975, V981, F983, M986, V988, V991, and A999 (M986 and A999 not identified in the pocket) the alanine-scanning mutagenesis identifies the importance of TM12 for the substrate specificity of P-gp. The V981A/F983A double mutation completely inhibits the transport of rhodamine 123, daunorubicin, and the L975A/V981A mutant by 50%.[43] In a study on murine P-gp five mutations have been identified (N350I, I862F, L865F, L868W, and A933T) that reduce the capacity of the compound tRA-96023 to inhibit P-gp-dependent drug resistance. The mutations also affect paclitaxel resistance.[35] Finally, a number of propafenone labeled residues are also present: 856, 858–862, 864, 939, 940, 942–953, 955, 958–963, 965, 968, 975, 982, 983–985, 987–989, 991, and 992.[17]

Figure 10 illustrates the top view of binding region 3 filled with the alpha spheres as generated by Site Finder. Clearly

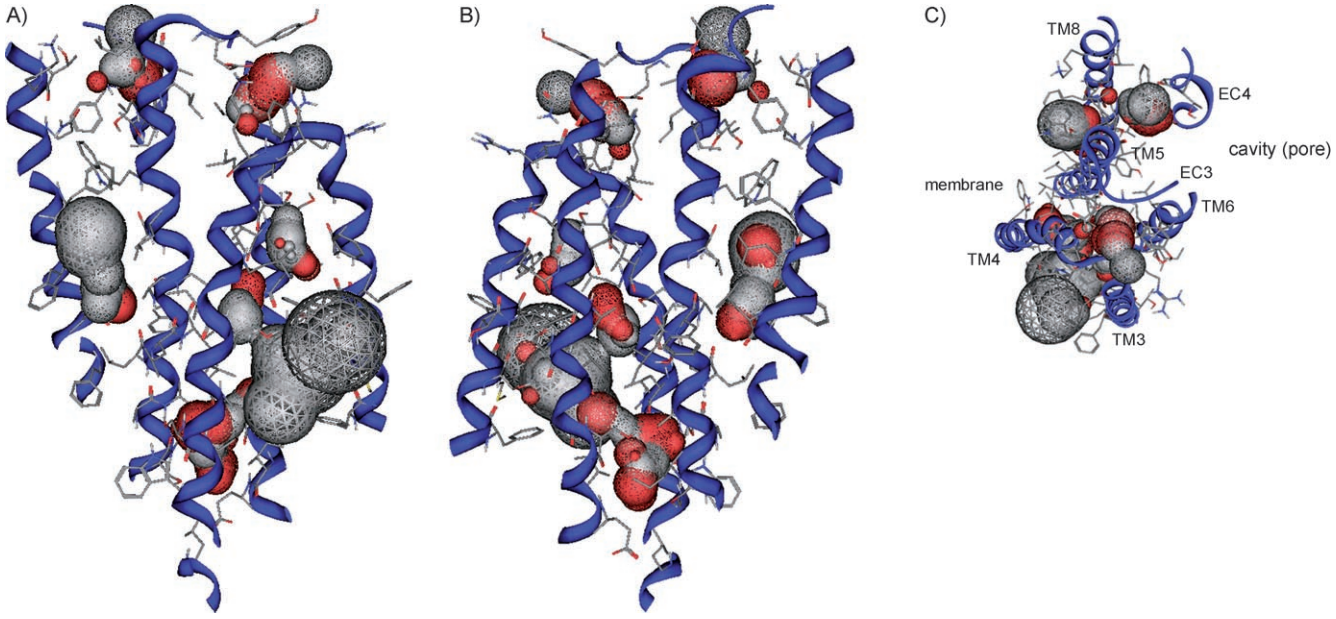

Figure 8. Binding region 2 filled with alpha spheres: A) Side view from the membrane; B) Side view from the pore; C) Top view (from outside); gray spheres indicate hydrophobic atoms and red spheres hydrophilic atoms; the amino acids are shown in sticks and colored according to the atom types (C: gray; O: red; N: blue; S: yellow; the nonpolar hydrogens are hidden); the TM backbones are shown as a blue tube.

Table 8. Amino acids (AAs) involved in binding region 3.[a]

TM1: 4 AAs				EC1: 18 AAs							
M69	L70	V71	F72	G73	E74	M75	T76	D77	**I78**	**F79**	**A82**
G83	**E86**	**D87**	**R95**	I98	G102	F103	N106	L107	E109	D110	M111
						TM2					TM9
T112	R113	Y114	Y116	**Y117**	**Y118**	I121	G122	L126	V127	Y130	R832
							TM10				
I836	N839	**I840**	L843	G846	I847	S850	**Q856**	**T858**	**L859**	**L860**	**L861**
						TM11					
L862	**I864**	V865	**I868**	A869	**G872**	**G939**	**I940**	**S943**	**F944**	**T945**	**A947**
							EC6				
M948	**M949**	**Y950**	**F951**	**S952**	**Y953**	**G955**	**R958**	**F959**	**A961**	**Y962**	**L963**
		TM12									
A965	**L968**	**L975**	**V981**	**F983**	**G984**	**A985**	**A987**	**V988**	**G989**	**V991**	**S992**

[a] AAs investigated by mutations, cross-linking, and photolabeling experiments are shown in bold (see also Table S3); AAs shown in Figure 9 are underlined; TM helices are assigned according to Swissprot (see Table 2).

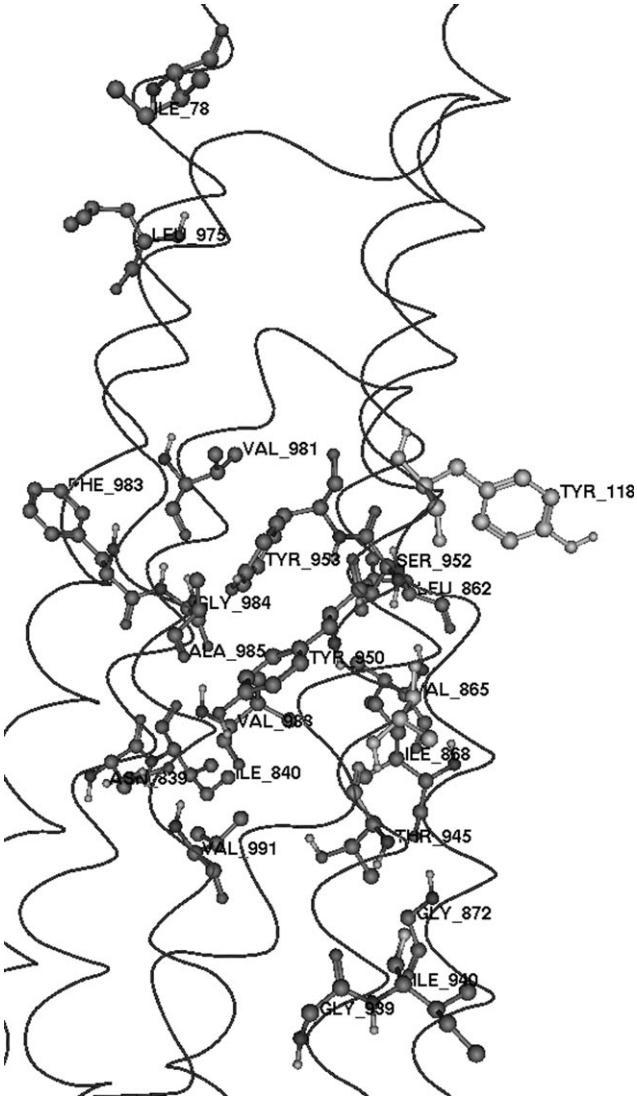

Figure 9. Binding region 3 with labeled amino acids shown in experimental studies to affect the transport function of P-gp (see Table 8); the orientation of the helices corresponds to Figure 5B; the residues are presented as balls and sticks (the nonpolar hydrogens are hidden).

seen are pockets that are simultaneously exposed to the membrane (gray spheres) and the pore (red spheres) similarly to the pockets observed in binding region 2.

Cavity pocket. Table 9 lists the residues involved in the cavity pocket. Some of them overlap with amino acids found in the binding regions described above (underlined in the table), but new residues are also identified. In the table the experimentally studied amino acids are shown in bold (see also Table S3). Among them are residues from TM1 (H61, G64, L65, M69) mutation of which has been shown to alter the substrate specificity of P-gp.[44,45] Residues Q128, E180, G183, and D184

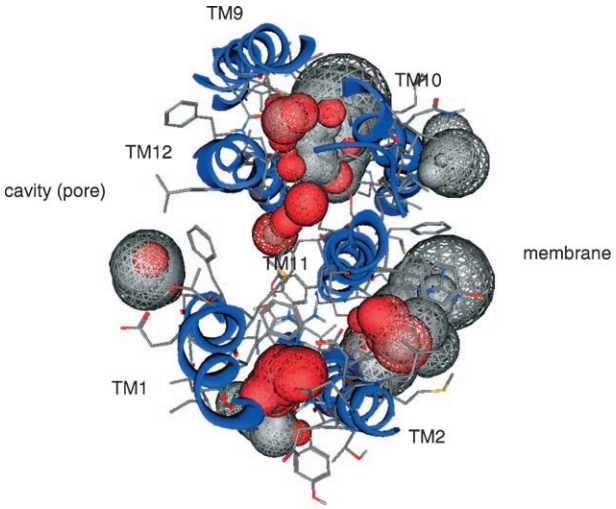

Figure 10. A top view (from outside) of binding region 11 3 filled with alpha spheres: gray spheres: hydrophobic atoms; red spheres: hydrophilic atoms; the amino acids are shown in sticks and colored according to the atom types (C: gray; O: red; N: blue; S: yellow; the nonpolar hydrogens are hidden); the TM backbones are shown as a blue tube.

(human Q132, E184, G187, and D188) belong to mutants that show reduction in activity compared to the wild-type P-gp; among them Q183 is characterized as a key residue for the altered protein function.[32] Residues labeled by propafenones are also involved (194–206, 306–310, 336–343, 762–777, and 978–994).[17] Mutation of hamster L338 (TM6) and N988 (TM12) (human L339 and Q990) causes reduction in the efficacy of the cyclosporin analogue PSC833, whereas murine L338 has been shown to cause a profound change in steroid binding.[35] This result is consistent with the proposed interaction of steroids with TM4, TM5, and TM6 domains of the protein.[46] The amino acid V982 has been labeled by TMEA and protected from labeling by cyclosporin A, colchicine, verapamil, and vinblastine.[41]

Table 9. Amino acids (AAs) involved in the cavity pocket.[a]

TM1											
G54	T55	A58	I59	**H61**	G62	A63	**G64**	**L65**	P66	L67	**M69**
	TM2							IC1			TM3
F72	Y117	I121	**V125**	A129	**Q132**	**V133**	W136	**E184**	**G187**	**D188**	G191
									TM5		
M192	**F194**	**Q195**	**S196**	**A198**	**T199**	T202	**G203**	**V206**	**N296**	**I299**	F303
			TM6								Link
I306	**Y307**	Y310	**F336**	**S337**	**L339**	**I340**	**A342**	**F343**	**S344**	G346	Q347
	TM7							TM8			
P350	A718	N721	G722	Q725	P726	**F728**	**A729**	**L762**	**F770**	**Q773**	**G774**
	TM9										TM11
F777	S831	A834	V835	I836	Q838	N839	**N842**	T845	G846	I849	**F938**
						TM12					
F942	**Q946**	**M949**	**Y950**	**Y953**	**F957**	**F978**	S979	**V981**	**V982**	**S983**	**A985**
M986	**A987**	**G989**	**Q990**	**S992**	**S993**	**F994**	D997				

[a] AAs investigated by mutations, cross-linking, and photolabeling experiments are shown in bold (see also Table S3); underlined are AAs that overlap with binding regions 1, 2, and 3; TM helices are assigned according to Swissprot (see Table 2).

Included are also residues F728 and A729 belonging to TM7 that is suggested by Loo and Clarke to be a part of the protein binding pocket;[47] A729, G774, and S993 are labeled by MTS-rhodamine.[13] In the pocket F343 (TM6) is involved; this amino acid is reported to be specific for MTS-rhodamine in stimulating ATP activity.[14] The F343 cysteine mutant has also been labeled by the hydrophilic fluorescein maleimide under basal conditions and the labeling has been lost after nucleotide binding or hydrolysis.[48] This suggests that the TM6 segment, which the residue belongs to, alters its conformation during the transport cycle. It should be noted that F343 is identified in neither of the binding regions; its protrusion to the pore presumes that the cavity pocket may not represent a binding site, but a site from where the compounds bound to the protein from the membrane are released. This suggestion is in agreement with involvement of residues from binding regions 2 and 3 in the pocket (see the underlined amino acids in Table 9): these residues compose the pore-facing parts of the binding regions.

Figure 11 A represents the top view of the pocket—it is formed by all TMs except TM4 and TM10. As already noted these TMs are apart from each other and have the largest distances in the model (see Table 4). Shown are the residues proven by Loo and Clarke to relate to verapamil (L339, A342, F942, in light gray) and rhodamine (I340, V981, V982, in dark gray) binding sites.[11,13] Figure 11 B and C illustrate these amino acids in a space filled form. In the figure, in addition to the domains involved in the cavity pocket (Figure 11 A), TM4 and TM10 are also shown, taken from binding regions 2 and 3, respectively. They also contain residues related to verapamil binding—S222 in TM4 and I868 in TM10. In the figures residues T945 and G984 for verapamil, and A841 and L975 for rhodamine are also displayed to get a general impression of the key amino acids related to the interactions of these substrates with P-gp. As seen from Figure 11 B and C the location of the labeled residues reveals two possible binding sites for rhodamine and verapamil related to either of the binding regions 2 and 3.

Discussion

After the withdrawal of the X-ray MsbA structures that have previously been used as templates for homology modeling of MDR transporters, the recently obtained Sav1866 structures[19,20] gave new possibilities for modeling of P-gp. These structures show a good correspondence to the 8 Å resolution electron microscopy data of P-gp.[25] The interaction observed in Sav1866 between IC4 and the neighboring NBD has also been experimentally shown in P-gp; additionally both proteins share some common substrates.[19] This indicates that Sav1866 is an appropriate template for homology modeling of P-gp. Recently O'Mara and Tieleman[49] published a homology model of P-gp for the closed, open, and semi-open protein conformations. Our homology model (Figure 1), based on the same Sav1866 template, represents the outward conformation of P-gp and closely resembles the open conformation model of O'Mara and Tieleman (compare the distances in Tables 3, 4, 5, and S1 and the distances reported in ref. [49]). According to several quality measures our homology model represents a valid and feasible protein structure: the PROCHECK Ramachandran plot shows a high percentage (87.6%) of residues in the most favored angle regions (for Sav1866 it is 86.6%) and no residues in disallowed regions (Figure S2). The initial P-gp model shows 0.72 Å RMSD to the Sav1866 template for all residues including loops. The difference in the RMSD values between the initial and final P-gp model (0.58 Å) is comparable to the same difference for the initial and minimized Sav1866 structures (0.71 Å). The RMSD value of 1.32 Å for matching all residues in the optimized structures is also very satisfactory, according to the relationship between sequence identity and the expected RMS deviation,[22] thus confirming the quality of the model. All five P-gp models generated showed very close positions of the non-hydrogen atoms (Table 1) suggesting only a minor influence of the applied force field. For identification of putative binding sites the OPLS-AA optimized structure was selected as it was obtained from a membrane incorporated model.

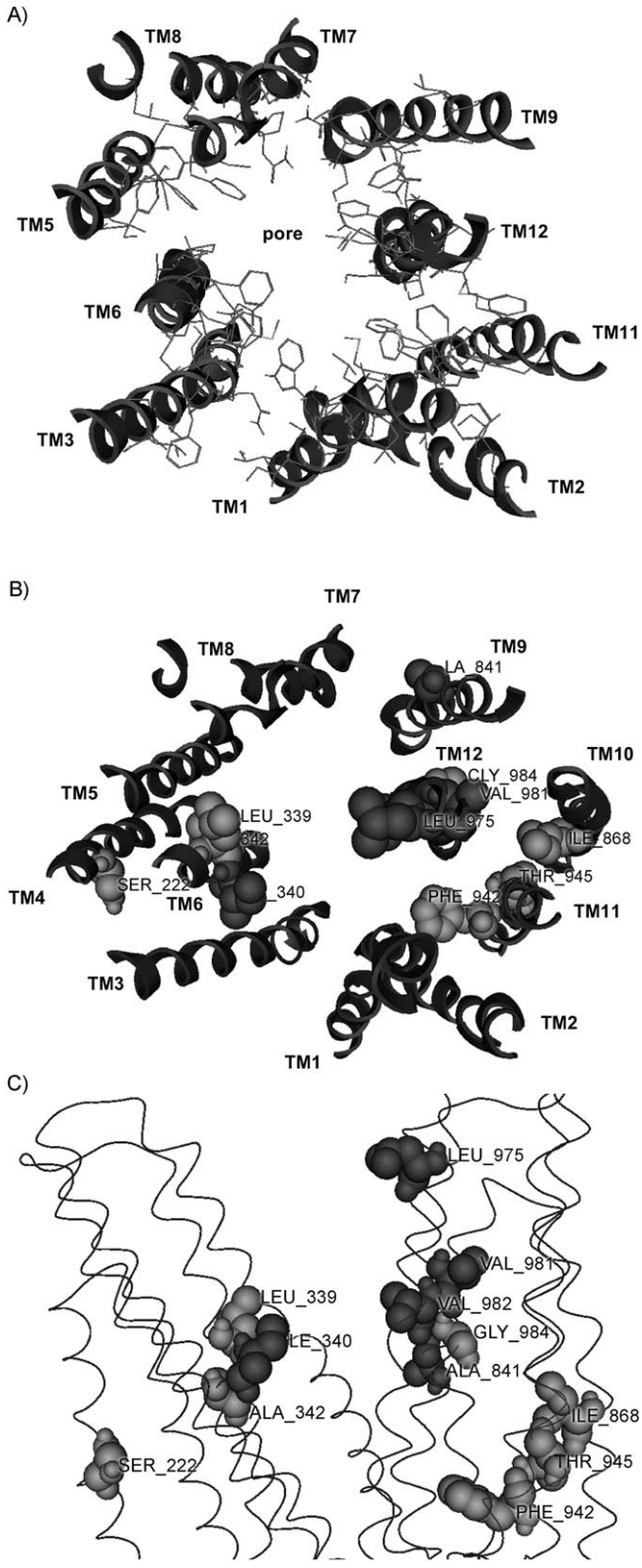

Figure 11. Top views: A) Cavity pocket; B) Cavity pocket complemented with TM4 and TM10; C) Side view of the TM domains with residues related to rhodamine binding (gray): I340, A841, L975, V981, V982; residues related to verapamil binding (light gray): S222, L339, A342, F942, G984 (data taken from cross-linking studies)[11,13] the amino acids are shown in space filled forms; the backbone of the protein is shown as a dark gray line.

An important question concerning the model is: what is the functional state of the protein? Dawson and Locher proposed that in the outward-facing conformation of TMs the bound substrate can escape into the outer leaflet of the lipid bilayer or into the aqueous medium surrounding the cell depending on its hydrophobicity.[19] Analyzing the cavity observed in Sav1866 they established the predominant presence of polar and charged amino acids and suggested that rather than being a high affinity binding site, it may reflect the extrusion pathway with little or no affinity for hydrophobic drugs.

The detailed comparison of the modeled distances to those evaluated in the numerous cross-linking experimental studies of Loo and Clarke (Tables 3, 4, and 5) and to the highest resolution microscopic structure of P-gp (Figure 3) points to the first step of ATP hydrolysis as the most probable functional state for the model. The low-affinity binding site is thought to be exposed to extracellular medium in the post-hydrolytic state (vanadate trapped).[50] In this state the protein is suggested to have already undergone the most essential conformational changes and might be in a transition from the "open to inside" to the "close to inside" conformation of the TMs, correspondingly the drug release must have already occurred. Thus, a low-affinity binding status or a status upon transition from a low to a high-affinity (basal) binding conformation can be assigned to the protein binding sites in the model. Unfortunately, at the moment there is no clear understanding about what does "high" and "low" affinity binding site mean at an atomic level. Probably, the difference between both relates to the conformational rearrangements in the site involving the same core of amino acids, although, some changes of the participating residues could also be expected. Thus, analysis of the protein binding regions at either step of the protein catalytic cycle could contribute to identification of the core amino acid residues involved in binding.

Three main membrane-related binding regions in P-gp were outlined (Figure 5A and B). Binding region 1 is located at the interface between the membrane and the cytosol and two other binding regions are located in the transmembrane parts of the protein. The regions contain multiple binding pockets and involve amino acids from all TMs but also from the IC and EC structural units of the protein. In the regions the pockets are close to each other (Figure 5); thus it is possible that drugs, depending on their structural properties, may bind to either more hydrophobic or more hydrophilic pockets or even to more than one pocket simultaneously.

Additionally, a big binding pocket has been found in the protein cavity (Figure 5C). Inspection shows that it involves residues from binding regions 1, 2, and 3 and may represent an "escaping" site, where the compounds that bind to any of these regions are released from the protein. The analysis of the cavity illustrates that it maintains hydrophobic to neutral surface properties and could possess potential affinity for hydrophobic drugs (Figure 4). Binding regions 2 and 3 contain pockets exposed simultaneously to the pore and to the membrane (Figure 8, 10). Thus, it can be presumed that more hydrophobic drugs may also be released by the protein to the outer leaflet of the membrane.

Analysis of the binding pockets revealed that they contain a large number of amino acids found to have an important impact on functioning of the protein in various mutational, cross-linking, and labeling experiments. The residues are shown to influence the transport activity for a number of substrates and inhibitors of P-gp, thus implying that the pockets identified might be reasonable. Unfortunately, from the experimental data no consensus conclusion about possible involvement of particular amino acids in the binding of particular drugs can be outlined. In the absence of 3D structural data of a ligand bound to P-gp, it is difficult to precisely define the location of the binding sites. The experimental data from mutation, cross-linking, and photolabeling studies may reflect effects that are not directly related to the drug binding and transport. The picture becomes even more complicated considering the induced-fit possibility in the drug–protein interactions, so that one cannot differentiate between the residues directly involved in binding and those located in close proximity that might be influenced by this binding. Thus, the presence of residues in the pockets which have not directly been proven to influence the transport function of the protein in experimental studies, does not necessarily mean that they are not involved in binding.

Residues that affect the transport of P-gp substrates and inhibitors (anthracyclines, *Vinca* alkaloids, colchicine, propafenones, rhodamine 123, verapamil, and their derivatives) have been found in different binding regions. A number of experimentally studied amino acids related to the transport of *Vinca* alkaloids, taxol, and PSC388 were identified in binding region 1 (Table 6, Figure 6). It is likely that big compounds, such as *Vinca* alkaloids, cyclosporine A, and their analogues, may interact with P-gp within the cytoplasmic compartment. In this region, as suggested by Gruol et al.[34] initial interactions of smaller drugs with the cytoplasmic loops of the protein, may also take place and these interactions can further influence the interaction of the protein with the ATP molecules. Using a FRET approach Logo and Sharom[51] mapped the location of the putative binding site of rhodamines (R-site) to 18–25 Å away from the "line" connecting C428 and C1071 (hamster P-gp) in the NBDs. We investigated the approximate location of this site in the P-gp model. The line between the corresponding human C431 and C1074 was built and the approximate location of the R-site was estimated around 25–32 Å below the membrane, thus suggesting that the drug binds to the protein in its cytosolic part. At the same time, as proven by the studies of Loo and Clarke[13] residues related to rhodamine binding are also found in the transmembrane region presuming also interactions in the transmembrane part of the protein. The rhodamine and verapamil labeled amino acids (Figure 11 B and C) point to possibility for these substrates to interact with the protein at more than one site. According to the figures, for example, rhodamines can interact with P-gp in the vicinity of V981 and V982 performing cross-linking with either L975 or I340. According to the location of the labeled amino acids verapamil may have two binding sites in binding regions 2 and 3. However, it cannot be excluded that these substrates, similar to the other ones, may interact with the protein in binding region 1 and use the pore for release from the cell. It is feasible that small drugs may have binding sites in more than one binding region of the protein. These results are in agreement with the assumption that P-gp has multiple drug-binding sites which may behave differently.[50]

In conclusion, our results suggest that P-glycoprotein has multiple binding sites and may bind and/or release substrates in multiple pathways. The binding regions and pockets identified may help further experimental studies, molecular modeling, and dynamics simulations that aim at a more precise location and identification of the protein binding sites.

Computational Methods

Homology modeling

Alignment: To find a suitable initial alignment of the conserved residues both sequences, Sav1866 (code Q99T13) used as a template and P-gp (code P08183) taken from the Swissprot database,[24] were scanned against with ScanProsite[52] and parts were identified that matched the ABC_TMF1 (ABC transporter integral membrane type-1 fused domain profile) and the ABC_TRANSPORTER2 profiles. The matching parts of the sequences were used as a template for the alignment of the whole sequences of P-gp on Sav1866 (PDB ID 2HYD[19]) by the "Align" tool in MOE.[31] The extracellular loops between TM1/TM2 (IC1) and TM7/TM8 (IC4) were assigned to have insertions or deletions in the non-helical regions. For IC4 different alignments of two amino acids' insertion were explored as leading to a modified loop composition. As an additional template for the Y-loop of the NBD, the TAP1 structure (PDB ID code 1JJ7[53]) was selected (in this segment it matches the sequence of P-gp better then the Sav1866 structure) and aligned in MOE. The overall sequence identity between P-gp and Sav1866 is 29.2% and for the TAP1 part it is 51.4%. The final alignment of the Sav1866 based model of P-gp is reported in Supporting Information Figure S1.

Force fields used: For the development of P-gp models the following force fields were employed: 1) CHARMM27;[54] 2) a mixture of lipid-optimized non-bonded parameters[55] and parameters based on the GROMOS87 force field implemented in GROMACS as ffgmx;[56–59] 3) a mixture of OPLS based lipid-parameters[55] and the OPLS-AA force field[60] using OPLS combination rules for the mixed Lennard–Jones parameters (coded as OPLS-AA for brevity).

Model development: Each transporter half was modeled within the environment. The first half of P-gp was built using the first chain of the Sav1866 structure as a template, whereas the remaining parts as ADP, Na^+-ions, water, and the second chain of Sav1866 were set as environment. The second half of P-gp was then built based on the second chain of Sav1866 using the model of the first half of P-gp and ADP, ions, and water from the Sav1866 structure as environment.

Two hundred models were generated using the best intermediate option and crude minimization with CHARMM 27 force field of each model to remove bad contacts. The intermediate models were further examined for the loop composi-

tion and the best combinations were selected according to the scoring function and the lowest number of outliers in the protein report function of MOE. For the Y-loop the TAP1 monomers were superimposed on the Sav1866 structure and a homology model for both NBDs of P-gp was created. The parts of the Y-loop were then incorporated into the Sav1866 based halves. The resulting model (initial model) was further minimized, first with weight constraints on the backbone atoms and later on without them. The quality of the model was proven by the protein report of MOE indicating no outliers and additionally checked with PROCHECK.[61] Supporting Information Figure S2 shows the PROCHECK Ramachandran plot indicating absence of outliers and 88% residues falling within the most favored regions.

Lipids: As a starting point we used a pre-equilibrated bilayer consisting of 128 1-palmitoyl-2-oleoyl-glycerophosphocholine (POPC) lipid molecules (popc128a.pdb and the corresponding topology file popc.itp) available from ref. [62]. Because the protein extends the dimensions of the membrane file, the lipids were extended to a bilayer of 512 lipids.

Membrane incorporation: Prior to the insertion into the membrane, the protein was minimized with the steepest descent algorithm followed by the conjugated gradient algorithm up to a gradient of 100 $kJ\,mol^{-1}\,nm^{-1}$. The protein was then placed into the membrane according to its surface properties and the overlapping lipids were removed. In cases of only partial overlapping the lipids were explored for possible chain rotamers and kept when steric clashes could be removed. The lipids were then minimized with strong position restrains (100 000 $kJ\,mol^{-1}\,nm^{-2}$) on the protein. Water was added and removed manually from the hydrophobic areas. Subsequently water molecules were randomly replaced by sodium and chloride ions to neutralize the system and to reach the physiologically relevant ion concentration of 154 $mmol\,L^{-1}$. Minimization with position restrains (1000 $kJ\,mol^{-1}\,nm^{-2}$) on protein followed by a minimization without restraints was performed.

The minimizations of the membrane incorporated models were carried out using the GROMACS 3.3.1 software package.[63–65] The Lennard-Jones interactions were calculated using a cutoff of 1.0 nm and a long range correction. The particle mesh Ewald method was used to calculate the electrostatic interactions[66,67] with a 0.9 nm cutoff for real space interactions and a 0.12 nm grid for the reciprocal space interactions using 6th order spline interpolation. The minimizations of the membrane incorporated models were carried out using periodic boundary conditions. The SPC (Simple Point Charge) water model was used in all minimizations.[68]

Identification of the binding sites

Two approaches, utilizing the programs SiteID[30] and Site Finder[31] have been employed. They belong to the category of the geometrical methods, but use different algorithms. The SiteID "Find pocket" mode with the grid method is appropriate when searching for likely binding sites for small molecules and when there are tight cavities and/or solvent inaccessible voids in the protein. The disadvantage of the grid method is that it depends on the orientation of the molecule in space. However, it produces non-overlapping and well-separated binding pockets on the protein surface. Site Finder is based on the methodology of convex hulls thus producing pockets invariant to rotation of the atomic coordinates. It treats the set of 3D points by triangulating them and associates each resulting simplex with a sphere, coded as "alpha sphere". The radii of the spheres are proportional to the planes of the convex hull of the point set. Each sphere is classified as either "hydrophobic" or "hydrophilic" depending on whether the sphere is a good hydrogen bonding point in the receptor. Hydrophilic spheres not near to hydrophobic ones are eliminated as they generally correspond to water sites. The generated pockets consist of one or more alpha spheres and at least one of them is hydrophobic.

SiteID settings: grid resolution 1.0 Å; film depth around protein 3.0 Å; minimum permissible distance between grid points and protein atom 2.5 Å; inclusion radius around grid point for protein atoms 8.0 Å; minimum number of protein heavy atoms within inclusion radius 80; inclusion radius for grid-grid distance 2.0 Å; minimum number of grid points within above radius 6. To find out the solvent accessible AAs the minimum possible value of 3 Å was used. Involvement of water bridges in the interactions with the protein was not considered.

Site Finder settings: the radius of a hypothetical hydrophilic hydrogen bonding atom 1.4 Å; the radii of a hypothetical hydrophobic atoms 1.8 Å ; isolate donor/acceptor distance 3 Å ; connection distance between clusters of alpha spheres 2.5 Å; minimum site size 3; minimum site radius 2 Å.

RMS values: The MOE-ProSuperpose module was used for calculating the root mean squared deviation (RMSD) values.

Surface potentials: The surface potentials were generated by MOLCAD[30] and "Surface and Maps" module.[31] The fast Connolly, Connolly channel (a portion of the Connolly surface that identifies an intramolecular channel or a cavity), and Gaussian contact surfaces were generated using the default settings. The surfaces were color coded by lipophilic and electrostatic properties (CHARMM charges).

Acknowledgements

I.P. and M.W. gratefully acknowledge the Alexander von Humboldt foundation, Germany, for the financial support of this study (grant BUL/1021057). I.P thanks also the NSF of Bulgaria (L-1416).

Keywords: binding site · homology models · MDR transporter · molecular modeling · p-glycoprotein

[1] B. Sarkadi, L. Homolya, G. Szakács, A. Váradi, *Physiol. Rev.* **2006**, *86*, 1179–1236.
[2] A. H. Schinkel, J. W. Jonker, *Adv. Drug Delivery Rev.* **2003**, *55*, 3–29.
[3] U. Hoffmann, H. K. Kroemer, *Drug Metab. Rev.* **2004**, *36*, 669–701.
[4] C. H. Lee, *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 43–52.
[5] M. J. Borgnia, G. D. Eytan, Y. G. Assaraf, *J. Biol. Chem.* **1996**, *271*, 3163–3171.
[6] S. Ayesh, Y. M. Shao, W. D. Stein, *Biochim. Biophys. Acta Mol. Basis Dis.* **1996**, *1316*, 8–18.
[7] A. B. Shapiro, V. Ling, *Eur. J. Biochem.* **1997**, *250*, 122–129.
[8] A. B. Shapiro, K. Fox, P. Lam, V. Ling, *Eur. J. Biochem.* **1999**, *259*, 841–850.

[9] C. Martin, G. Berridge, C. F. Higgins, P. Mistry, P. Charlton, R. Callaghan, *Mol. Pharmacol.* **2000**, *58*, 624–632.
[10] A. R. Safa, *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 1–17.
[11] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2001**, *276*, 14972–14979.
[12] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2001**, *276*, 36877–36880.
[13] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2002**, *277*, 44332–44338.
[14] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2003**, *278*, 50136–50141.
[15] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2004**, *279*, 7692–7697.
[16] T. W. Loo, M. C. Bartlett, D. M. Clarke, *J. Biol. Chem.* **2004**, *279*, 18232–18238.
[17] G. F. Ecker, E. Csaszar, S. Kopp, B. Plagens, W. Holzer, W. Ernst, P. Chiba, *Mol. Pharmacol.* **2002**, *61*, 637–648.
[18] K. Pleban, S. Kopp, E. Csaszar, M. Peer, T. Hrebicek, A. Rizzi, G. F. Ecker, P. Chiba, *Mol. Pharmacol.* **2005**, *67*, 365–374.
[19] R. J. P. Dawson, K. P. Locher, *Nature* **2006**, *443*, 180–185.
[20] R. J. P. Dawson, K. P. Locher, *FEBS Lett.* **2007**, *581*, 935–938.
[21] J. K. Zolnerciks, C. Wooding, K. J. Linton, *FASEB J.* **2007**, *21*, 3937–3948.
[22] C. Chothia, A. M. Lesk, *EMBO J.* **1986**, *5*, 823–826.
[23] D. P. Tieleman, J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu, L. Monticelli, *J. Phys. Condens. Matter* **2006**, *18*, S1221-S1234.
[24] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, A. Bairoch, *Nucleic Acids Res.* **2003**, *31*, 3784–3788.
[25] M. F. Rosenberg, R. Callaghan, S. Modok, C. F. Higgins, R. C. Ford, *J. Biol. Chem.* **2005**, *280*, 2857–2862.
[26] T. W. Loo, D. M. Clarke, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3511–3516.
[27] I. K. Pajeva, C. Globisch, M. Wiese, *J. Med. Chem.* **2004**, *47*, 2523–2533.
[28] T. W. Loo, M. C. Bartlett, D. M. Clarke, *Biochemistry* **2005**, *44*, 10250–10258.
[29] D. R. Stenham, J. D. Campbell, M. S. P. Sansom, C. F. Higgins, I. D. Kerr, K. J. Linton, *FASEB J.* **2003**, *17*, 2287–2289.
[30] SYBYL 7.3, Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63114–2917.
[31] MOE 2006.08 (Molecular Operating Environment), Chemical Computing Group, 1010 Sherbrooke Street West, Suite 910; Montreal, Que., Canada H3 A 2R7.
[32] T. Kwan, P. Gros, *Biochemistry* **1998**, *37*, 3337–3350.
[33] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **1994**, *269*, 7243–7248.
[34] D. J. Gruol, M. N. King, M. E. Kuehne, *Mol. Pharmacol.* **2002**, *62*, 1238–1248.
[35] D. J. Gruol, J. Bernd, A. E. Phippard, I. Ojima, R. J. Bernacki, *Mol. Pharmacol.* **2001**, *60*, 104–113.
[36] M. Hanna, M. Brault, T. Kwan, C. Kast, P. Gros, *Biochemistry* **1996**, *35*, 3625–3635.
[37] N. Shani, A. Sapag, D. Valle, *J. Biol. Chem.* **1996**, *271*, 8725–8730.
[38] T. W. Loo, D. M. Clarke, *Biochemistry* **1994**, *33*, 14049–14057.
[39] T. Shoshani, S. Zhang, S. Dey, I. Pastan, M. M. Gottesman, *Mol. Pharmacol.* **1998**, *54*, 623–630.
[40] E. Welker, K. Szabó, Z. Holló, M. Müller, B. Sarkadi, A. Váradi, *Biochem. Biophys. Res. Commun.* **1995**, *216*, 602–609.
[41] T. W. Loo, D. M. Clarke, *J. Biol. Chem.* **2001**, *276*, 31800–31805.
[42] J. Song, P. W. Melera, *Mol. Pharmacol.* **2001**, *60*, 254–261.
[43] P. Hafkemeyer, S. Dey, S. V. Ambudkar, C. A. Hrycyna, I. Pastan, M. M. Gottesman, *Biochemistry* **1998**, *37*, 16400–16409.
[44] Y. Taguchi, K. Kino, M. Morishima, T. Komano, S. E. Kane, K. Ueda, *Biochemistry* **1997**, *36*, 8883–8889.
[45] Y. Taguchi, M. Morishima, T. Komano, K. Ueda, *FEBS Lett.* **1997**, *413*, 142–146.
[46] Q. D. Vo, D. J. Gruol, *J. Biol. Chem.* **1999**, *274*, 20318–20327.
[47] T. W. Loo, M. C. Bartlett, D. M. Clarke, *Biochem. J.* **2006**, *399*, 351–359.
[48] A. Rothnie, J. Storm, J. Campbell, K. J. Linton, I. D. Kerr, R. Callaghan, *J. Biol. Chem.* **2004**, *279*, 34913–34921.
[49] M. L. O'Mara, D. P. Tieleman, *FEBS Lett.* **2007**, *581*, 4217–4222.
[50] K. J. Linton, C. F. Higgins, *Pfluegers Arch.* **2007**, *453*, 555–567.
[51] M. R. Lugo, F. J. Sharom, *Biochemistry* **2005**, *44*, 14020–14029.
[52] E. de Castro, C. J. A. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, N. Hulo, *Nucleic Acids Res.* **2006**, *34*, W362–W365.
[53] R. Gaudet, D. C. Wiley, *EMBO J.* **2001**, *20*, 4964–4972.
[54] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400–1415.
[55] O. Berger, O. Edholm, F. Jähnig, *Biophys. J.* **1997**, *72*, 2002–2013.
[56] W. F. van Gunsteren, H. J. C. Berendsen, Gromos-87 manual, Biomos BV, Nijenborgh 4, 9747 AG Groningen, The Netherlands, **1987**.
[57] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, H. J. C. Berendsen, Gromacs User Manual version 3.3, www.gromacs.org, **2005**.
[58] D. P. Tieleman, M. S. Sansom, H. J. Berendsen, *Biophys. J.* **1999**, *76*, 40–49.
[59] D. P. Tieleman, H. J. Berendsen, M. S. Sansom, *Biophys. J.* **1999**, *76*, 1757–1769.
[60] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
[61] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
[62] http://moose.bio.ucalgary.ca/index.php?page=Structures_and_Topologies.
[63] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **1995**, *91*, 43–56.
[64] E. Lindahl, B. Hess, D. van der Spoel, *J. Mol. Model.* **2001**, *7*, 306–317.
[65] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.
[66] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[67] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.
[68] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, *Intermolecular Forces*, D. Reidel Publishing Company, Dordrecht, **1981**, pp. 331–342.

Received: September 11, 2007
Revised: October 18, 2007
Published online on January 3, 2008

