Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 837–851
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Towards understanding the mechanism of action of the multidrug resistance-linked half-ABC transporter ABCG2: A molecular modeling study

Yong-Fu Li [a], Orsolya Polgar [b], Masaki Okada [b], Lothar Esser [a], Susan E. Bates [b], Di Xia [a,*]

[a] *Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, United States*
[b] *Cancer Therapeutics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, United States*

Received 30 January 2006; received in revised form 22 August 2006; accepted 22 August 2006
Available online 30 August 2006

Abstract

The ATP-binding cassette protein ABCG2 is a member of a broad family of ABC transporters with potential clinical importance as a mediator of multidrug resistance. We carried out a homology and knowledge-based, and mutationally improved molecular modeling study to establish a much needed structural framework for the protein, which could serve as guidance for further genetic, biochemical, and structural analyses. Based on homology with known structures of both full-length and nucleotide-binding domains (NBD) of ABC transporters and structural knowledge of integral membrane proteins, an initial model of ABCG2 was established. Subsequent refinement to conform to the lipophilic index distributions in the transmembrane domain (TMD) and to the results of site-directed mutagenesis experiments led to an improved model. The complete ABCG2 model consists of two identical subunits facing each other in a closed conformation. The dimeric interface in the nucleotide-binding domain (NBD) involves a characteristic nucleotide sandwich and the interface in the TMD consists of the TM helices 1–3 of one subunit and the helices 5 and 6 of the other. The interface between the NBD and the TMD is bridged by the conserved structural motif between TM2 and TM3, the intracellular domain 1 (ICD1), and the terminal β-strand (S6) of the central β-sheet in the NBD. The apparent flexibility of the ICD1 may play a role in transmitting conformational changes from the NBD to the TMD or from the TMD to the NBD.
Published by Elsevier Inc.

Keywords: Molecular modeling; Structure; ABC transporter; ABCG2

1. Introduction

ATP-binding cassette (ABC) transporters represent the largest gene family of polytopic membrane-spanning proteins that use the energy of ATP hydrolysis to drive active transport of various substances across cellular membranes [1]. Members of this family are encoded within the genome of every living organism and are involved in importing and/or exporting a large variety of molecules [2]. Some clinically relevant transporters belong to this family, such as the multidrug resistance transporters ABCB1 (P-glycoprotein, P-gp) [3,4], the multidrug resistance related protein 1 (MRP1 or ABCC1) [5], the breast cancer resistance protein (BCRP or ABCG2) [6,7], and the chloride channel CFTR (cystic fibrosis transmembrane conductance regulator) [8–10]. Human ABC transporters are classified into seven subfamilies (ABCA to ABCG) on the basis of phylogenetic relationships defined in the analyses of NBD sequences from various human ABC genes [1,11].

ABCG2 (also called MXR/BCRP/ABCP) belongs to the ABCG subfamily of half-transporters and is found highly expressed in the placenta, small intestine, and breast cancer cell lines [6,7,12]. The human *ABCG2* gene is located on

Abbreviations: ABC, ATP-binding cassette; BCRP, breast cancer resistant protein; ABCP, ABC transporter expressed in the placenta; CFTR, cystic fibrosis transmembrane-conductance regulator; DPBS, Dulbecco's phosphate buffered saline; *Ec*MsbA, *E. coli* MsbA; EM, electron microscopy; FTC, fumitremorgin C; ICD, intra-cellular domain; MDR, multidrug resistance; MXR, mitoxantrone-resistant protein; NBD, nucleotide-binding domain; P-gp, P-glycoprotein; PhA, pheophorbide A; SNP, single nucleotide polymorphism; *St*MsbA, *S. typhimurium* MsbA; TMH, transmembrane helix; TMD, transmembrane domain; *Vc*MsbA, *V. cholera* MsbA

* Corresponding author at: Laboratory of Cell Biology, National Cancer Institute, NIH, 37 Convent Dr., Building 37, Room 2122C, Bethesda, MD 20892, United States. Tel.: +1 301 435 6315; fax: +1 301 480 2315.
E-mail address: dixia@helix.nih.gov (D. Xia).

1093-3263/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmgm.2006.08.005

chromosome 4q22 and encodes a polypeptide chain of 655 amino acid residues [1]. Overexpression of ABCG2 in tumor cell lines confers resistance to anticancer drugs such as mitoxantrone, methotrexate, topotecan derivatives, bisantrene, etoposide, the irinotecan metabolite 7-ethyl-10-hydroxycamptothecin (SN-38), and flavopiridol [13]. Members of the ABCG subfamily feature an N-terminal NBD and a C-terminal TMD, which is similar to the topology of Pdr5p [14], a yeast multidrug ABC transporter. This topology is the reverse of that observed for many other members of the ABC transporter family, such as for P-gp; the effect of this reverse orientation on function is not known. Experiments have demonstrated that the ABCG subfamily proteins engage in a homo- or hetero-dimeric association to form an active transporter [15,16].

A number of nucleotide changes have been identified in the *ABCG2* gene following drug selection or by searching for single nucleotide polymorphisms (SNPs) in the human population [17–23]; studies of these mutants aided our understanding of ABCG2's structure and function. However, the lack of a structural framework posed serious challenges in rationalizing results from mutagenesis studies and/or in formulating further experiments to characterize the protein. While the pursuit of a high-resolution experimental structure is underway, we decided in the interim to generate a knowledge-based structural model. To date, crystal structures of four full-length bacterial ABC transporters have been reported, namely, the lipid A transporter MsbA from *Escherichia coli* (*Ec*MsbA) [24], from *Vibrio cholera* (*Vc*MsbA) [25], and from *Salmonella typhimurium* (*St*MsbA) [26], and the vitamin B12 importer BtuCD from *Escherichia coli* [27]. In addition, crystal structures of a number of isolated NBDs have been reported, including HisP, MalK, TAP1, MJ0796, MJ1267, Rad50, and HlyB (see the review article [28] and references therein). Although most of these are structures of bacterial ABC proteins, they have facilitated mechanistic studies of ABC transporters and have been particularly useful for studies of eukaryotic transporters with clinical importance, for which experimental structural information is difficult to obtain. In this report, a structural model of ABCG2 was constructed by combining the available structural knowledge of ABC transporters with sequence homology and results from mutagenesis.

2. Methodology

2.1. Sequences and structure models

The amino acid sequence of human ABCG2 was obtained from GeneBank (accession no. NP_004818). The following models for the NBDs were taken from the Protein Data Bank (PDB): *E. coli* MalK (1q12), *T. litoralis* MalK (1g29), HisP from *S. typhimurium* (1b0u), human TAP1 (1jj7), and LmrA of *L. lactis* (1mv5). Also obtained from the PDB were the coordinates of the full-length bacterial transporter MsbAs from *V. cholera* (1pf4), from *E. coli* (1jsq), from *S. typhimurium* (1z2r), and BtuCD of *E. coli* (1l7v).

2.2. Modeling the NBD of ABCG2

For the NBD, significant sequence identities were found by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST) with the following structures in the PDB: *E. coli* MalK, *T. litoralis* MalK, HisP from *S. typhimurium*, human TAP1, MsbA from *V. cholera*, BtuD of *E. coli*, and LmrA of *L. lactis*. The crystal structure of MalK from *E. coli* (1q12) was used as the template for the homology modeling of the ABCG2 NBD. The initial monomeric NBD model was generated by the automated modeling program SWISS-model [29]; the dimeric model was then generated by superposition of the monomers onto the MalK dimer in the modeling program O [30] and subsequently refined in the refinement program CNS [31] to optimize van der Waals interactions and to conform to the standard stereochemistry values of bond angles, bond lengths, and dihedral angles.

2.3. Modeling the TMD of ABCG2

For modeling the TM domain, TM helix segments were predicted with various computer programs such as SOSUI [32], PHD [33], HMMTOP [34], and TMHMM [35], and a consensus of six TM helices (Table 2) emerged. The boundary of each predicted helix was adjusted to conform to a set of empirical rules (see Section 3) derived from known membrane protein structures [36,37]. The predicted TM sequence segments of ABCG2 from 361 to 655 were mapped to the sequence alignment based on MsbA structures (Fig. 3). An ABCG2 TM model was built using the *Vc*MsbA structure as a template in the program O. Adjustments were made to the model whenever residues in the model had stereochemistry violations.

2.4. Lipophilic index calculation

Lipophilic indices (lipid propensity) for predicted TM helices were calculated based on the empirical hydrophobic scales of individual amino acids [38]. Briefly, the membrane bilayer was conveniently divided into four zones of roughly equal thickness: two for lipid head groups, and two for lipid tail domains. Correspondingly, each TM helix was divided into four quadrants and two different hydrophobic scale sets were used for residues in the two outer quadrants and for those in the two inner quadrants, respectively. Residues in a TM helix are grouped into three categories: membrane-facing, internal cavity-facing and hydrophobic core. A lipophilic index for residues of a given category in a TM helix was obtained by summation over all residues of that category in the helix and the results are given in Table 3.

2.5. Site-directed mutagenesis and flow cytometry analysis

The mutagenesis procedure has been published previously [17]. Briefly, mutations in the *ABCG2* gene were generated by site-directed mutagenesis in a pcDNA3.1 vector. Human embryonic kidney (HEK) 293 cells were transfected with the plasmid DNA at 37 °C in 5% CO_2 for one hour in serum-free

medium containing a 1:2 ratio of DNA and TransFast transfection reagent (Promega). Cells were then incubated for an additional 48 h in media containing 10% fetal bovine serum before the addition of 2 mg/ml G418. For flow cytometry, cells were trypsinized and resuspended in complete media (phenol red-free IMEM with 10% fetal calf serum) containing 10 μM pheophorbide a (PhA, Frontier Scientific), with or without 10 μM of the ABCG2 blocker fumitremorgin C (FTC), and incubated for 30 min at 37 °C in 5% CO_2 [39]. Cells were then washed twice with cold Dulbecco's Phosphate Buffered Saline (DPBS) and placed on ice in the dark until analyzed. The analysis was done with a FACSort flow cytometer equipped with both a 488 nm argon laser and a 536 nm diode laser. In studies with the anti-ABCG2 antibody 5D3 [39], cells were incubated in 2% BSA/DPBS with either phycoerythrin-labeled negative control antibody (IgG2b) or phycoerythrin-labeled 5D3 antibody (both from eBioscience) according to the manufacturer's instructions, washed with DPBS, and subsequently analyzed.

2.6. Modeling of the full-length dimeric ABCG2

The TMD model was refined on the basis of the mutagenesis results; the resulting TMD model was combined with the NBD model manually in the program O with the two 2-fold axes superimposed. The ICD1 motif of the TMD was adjusted to interface with the β-strand S6 of the central β-sheet of NBD. The resulting model was subjected to several cycles of refinement with the program CNS to further remove unfavorable interactions. The final ABCG2 model was checked with the program PROCHECK [40]; the distribution of main chain dihedral angles obtained from Ramachandran plot shows 82.1% of residues in the most favorable region, 17.1% in the additionally allowed region, and none in the disallowed region.

3. Results

Homology modeling requires the model template that is sufficiently homologous in sequence to the query sequence. Sequence comparisons failed to identify a homolog of known structure with significant identity to the entire ABCG2 sequence. Based on sequence analysis, ABCG2 could roughly be divided into four parts: an N-terminal hydrophilic segment of 40 residues (1–40) followed by the core NBD of 260 residues (41-300); the C-terminal TM helix domain (361–655) follows a linker region of 60 residues (301–360) between the NBD and TM domain. To date, experimental structures such as those obtained crystallographically and capable of serving as templates for modeling could only be found for the NBD and for the TMD; no suitable models exist for the N-terminal hydrophilic segment and the linker between the NBD and TMD. Furthermore, noticeable sequence homology exists only for the NBD between ABCG2 and templates. It was therefore considered necessary to apply different approaches for modeling different parts of ABCG2. Fig. 1 depicts a flowchart of the strategy used for modeling ABCG2, which consists of three steps: high sequence similarity in the NBD allowed homology modeling for this domain, whereas the very low sequence similarity in the TMD warranted the use of a knowledge-based approach. In the third step, a complete ABCG2 model was generated by combining the separate TMD and NBD model. Although this approach resulted in an NBD model likely to be more reliable than the TMD model, the quality and reliability of the TMD structure were improved by incorporating experimental mutagenesis data as constraints for the modeling.

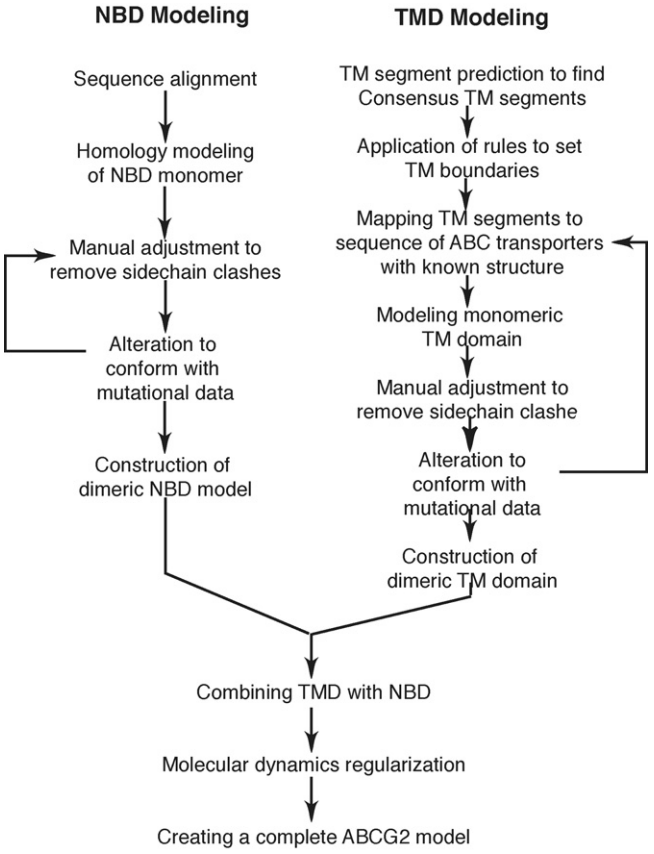

Fig. 1. Flowchart of the strategy for modeling the dimeric ABC transporter ABCG2.

3.1. Sequence alignment of the NBD domain of ABCG2 with members of the ABC transporter family

Sequence alignments of the full-length ABCG2 with members of the ABC transporter family showed an overall low identity or similarity of less than 20%, except for those from the same subfamily. Conservations are most pronounced in the NBD, especially for the Walker A and B motif, and in the signature VSGGE motif. These conserved sequence segments provided strong constraints for model building. Since a number of NBD structures were already available, ranging from prokaryotes such as *E. coli* to eukaryotes including humans, these structures were superimposed to generate a more precise, structure-based sequence alignment (the base alignment). In this alignment, additional conserved sequence motifs such as the A-, Q-, and the D-loop, and the H-motif were found precisely aligned. We further aligned the ABCG2

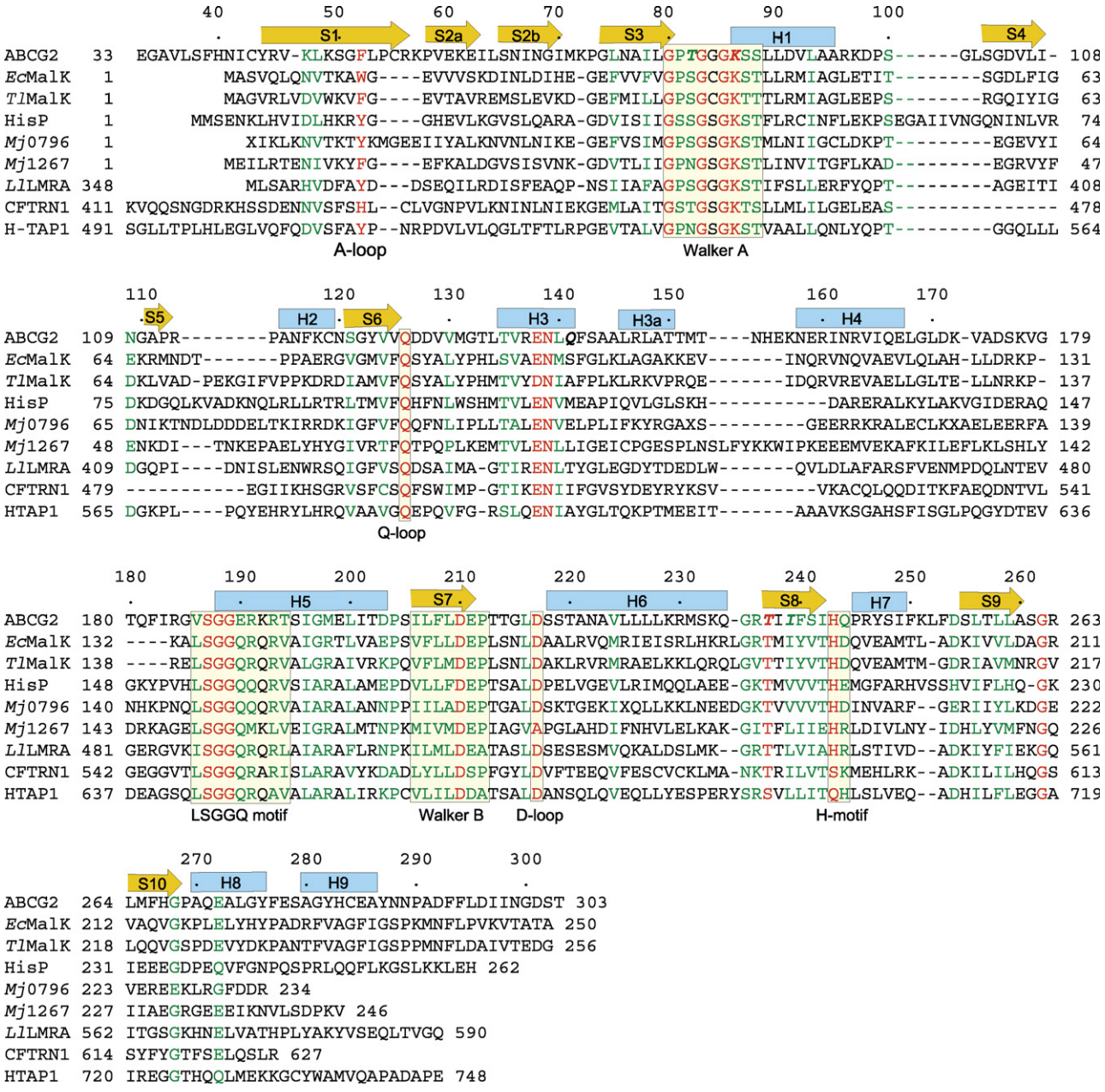

Fig. 2. Structure-based sequence alignment of NBDs. Eight sequences are aligned: the human ABCG2; *Ec*MalK, the ATP-binding protein of the *E. coli* maltose/maltodextrin transporter (PDB: 1Q12); *Tl*MalK, the ATPase subunit of the *T. litoralis* trehalose/maltose transporter (PDB: 1G29); HisP, the ATP-binding subunit of the histidine permease from *S. typhimurium* (PDB: 1B0U); *Mj*0796, the ATP-binding protein of the hypothetical ABC transporter from *M. jannaschii* (PDB: 1F30); *Mj*1267, another ATP-binding subunit of a hypothetical ABC transporter in *M. jannaschii* (PDB: 1G6H); *Ll*LMRA, the NBD of the MDR half-transporter from *L. lactis* (PDB: 1MV5); CFTRN1, the N-terminal NBD of the mouse cystic fibrosis translocation regulator (PDB:1ROW); TAP1, the NBD of the human TAP1 (PDB: 1JJ7). Functionally important sequence motifs, such as the Walker A, LSGGQ, Walker B, A-, Q-, D-, and H-loops, are boxed and labeled below the box. Modeled secondary structure elements for the ABCG2 NBD model are shown with yellow arrows for β-strands and blue rectangular bars for α-helices. Highly conserved residues among all sequences appear in red, conserved residues are in green. At positions in the NBD sequence where mutations are known, the residues are highlighted in bold–italic type. Gaps in the alignment are shown with "–".

sequence against the base alignment (Fig. 2) and obtained an accurate account of sequence identities of the ABCG2 NBD to those of other transporters (Table 1). Despite the low sequence identities between NBDs of ABCG2 and various ABC transporters, the sequence similarities are fairly constant with the human TAP1 and the *E. coli* MalK being the highest at 43% (Tables 2 and 3).

3.2. Homology modeling of the ABCG2 NBD

Alignment of the ABCG2 NBD with multiple sequences showed that the *E. coli* MalK, for which the crystal structure is known [41], is clearly among the closest to ABCG2, both in sequence identity and similarity. In contrast to the human TAP1 and *T. litoralis* MalK, which have a comparable homology, the

Table 1
Sequence identity and similarity in the NBD and TMD between ABCG2 and selected ABC transporters

ABC transporter	NBD domain (%)[a]		ABC transporter	TM domain (%)[b]	
	Identity	Similarity[c]		Identity	Similarity
ABCG2	100.0 (234)[d]	100.0 (234)	ABCG2	100.0 (295)	100.0 (295)
*Ec*MalK	26.5 (62)	42.7 (100)	*Vc*MsbA	9.5 (28)	19.5 (56)
*Tl*MalK	22.6 (53)	42.3 (99)	*Ec*MsbA	6.4 (19)	18.3 (54)
His-P	18.4 (43)	40.6 (95)	*Lc*LmrA	8.5 (25)	21.0 (62)
MJ1267	17.5 (41)	38.0 (89)	Pgp-n	6.1 (18)	19.7 (58)
LMR-A	20.9 (49)	40.2 (94)	Pgp-c	6.1 (18)	15.6 (46)
CFTR-N1	20.5 (48)	34.6 (81)	Dro-Wht	22.0 (65)	40.3 (119)
TAP1	17.1 (40)	42.7 (100)	Pgp-n-rev[e]	(16)	(30)

[a] Sequence identities and similarities are based on alignments in Fig. 2.
[b] Sequence identities and similarities are based on alignments in Fig. 4.
[c] Similarity is defined for amino acid residues in the alignment, which share common physical or chemical properties such as size, hydrophobicity or charge.
[d] Numbers in parentheses are number of residues.
[e] Order of the TM helices for ABCG2 is reversed and the helices are aligned with Pgp N-terminal TM domain helices.

E. coli MalK crystallizes in a dimeric form with bound ATP and its structure is believed to be more physiologically relevant. By homology modeling, we mapped the sequence of the ABCG2 NBD onto the structure of the *E. coli* MalK monomer. The model was then manually adjusted for side-chain clashes and for accommodation of deletions and insertions, and subsequently checked for consistency with known mutations. The resulting monomeric NBD (Fig. 3A) was fitted to the second monomer in the dimeric MalK to produce the dimeric NBD model of ABCG2 (Fig. 3B). The complete dimeric NBD of ABCG2 was subject to a few cycles of molecular dynamics to remove unfavorable interactions present in the initial model. The final NBD model from residue Y44 to N288 is very similar to that of *E. coli* MalK with a root-mean-square (rms) deviation of 1.109 Å for 229 Cα atoms. In this model, the ATP molecule also came from the MalK structure, whereas the position for the Mg^{2+} ion was derived from a consensus of various structures.

3.3. Agreement between the ABCG2 NBD model and available mutagenesis data

The monomeric model (Fig. 3A) of ABCG2 NBD has a shape typical of NBDs of ABC transporters and consists of two subdomains. The larger RecA-like domain (red) is made up of a central 5-stranded parallel β-sheet that is flanked by four α-helices. The smaller subdomain (green), specific to ABC transporters, is mostly helical. Mapping of the modeled secondary structure elements to the ABCG2 NBD sequence is given in Fig. 2. Like other monomeric NBDs, the ATP moiety is located at the surface with its phosphate groups near the conserved Walker A motif, fits comfortably into a crack between the two symmetry related NBD monomers, and interacts with the ABC signature motif from the symmetry mate (Fig. 3B). The NBD model satisfies all known requirements from prior structural work and biochemical studies, and thus should represent a reasonable and testable three-dimensional structure before the real crystal structure is obtained. We identified five mutations in the NBD, either from a literature survey or from mutations generated in our laboratory (Table 4); these were mapped onto the model and are shown in Fig. 3A. Mutations T82A, K86M, and K86I [42,43] are part of the Walker A motif; all lead to loss of transport activity. The T82A mutation results in the possible loss of interaction with residue R193 in the neighboring LSGGQ motif, potentially affecting dimer assembly. The K86M or the K86I mutation leads to loss of interaction with the tri-phosphate group of the bound ATP, and thus inactivates the enzyme. Q141K, a SNP [22,23], is located on the surface of the small sub-domain near the conserved H3 helix

Table 2
Predicted transmembrane segments in the ABCG2 TMD

Programs	Predicted TM segments (number of amino acid residues)							
	I	II	IIa	III	IV	V	Va	VI
Predicted TM segments made by various programs								
SOSUI [32][a]	393–415[b] (23)	426–448 (23)	454–476 (23)	427–500 (22)	506–528 (23)	540–562 (23)	–	–
HMMTOP [34]	394–416 (23)	429–450 (22)	–	478–499 (22)	505–527 (23)	533–555 (23)	561–581 (23)	631–651 (21)
TYHMM [35]	394–416 (23)	428–450 (23)	–	478–499 (23)	506–528 (23)	533–555 (23)	–	629–651 (24)
PHD [33]	393–414 (22)	420–444 (25)	450–478 (29)	480–490 (11)	502–530 (29)	534–549 (16)	562–586 (25)	629–651 (23)
Consensus[c]	394–416 (23)	429–450 (22)	–	478–499 (23)	505–530 (26)	533–555 (23)	–	629–651 (23)
Final TMH[d]	384–414 (31)	422–448 (27)	–	478–502 (25)	506–531 (26)	537–566 (30)	–	627–655 (29)

[a] Numbers in square brackets are reference numbers.
[b] Numbers represent start and end point of a predicted TM segment.
[c] The consensus for the number of TM segments in the ABCG2 was determined based on the number of times each potential TM segment is being predicted by various programs. Predicted potential TM segments IIa and Va were rejected because they were not predicted by two of the four methods used.
[d] The final length for each TM segment after giving considerations to various empirical rules.

Table 3
Lipophilic indices for three categories of residues on TM helices in the structures of *St*MsbA, *Vc*MsbA and the final ABCG2 model[a]

TMs	Residue category	*St*MsbA		*Vc* MsbA		ABCG2	
		Lipophilic index[b] (no. of residues)[c]	Total no. of residues[d]	Lipophilic index (no. of residues)	Total no. of residues	Lipophilic index (no. of residues)	Total no. of residues
TM1	Lipid	0.74 (13)	25	0.45 (12)	29	0.89 (12)	31
	Core	−2.45 (9)		−1.98 (12)		−3.19 (11)	
	Cavity	−0.46 (3)		−0.56 (5)		−0.57 (8)	
TM2	Lipid	0.48 (13)	26	0.67 (13)	28	0.52 (18)	27
	Core	−3.10 (13)		−3.59 (12)		−1.71 (7)	
	Cavity	–		−0.07 (3)		−0.75 (2)	
TM3	Lipid	0.68 (6)	22	0.57 (9)	23	1.11 (9)	22
	Core	−0.28 (12)		−1.20 (7)		−0.27 (6)	
	Cavity	−0.61 (6)		−2.10 (7)		0.33 (7)	
TM4	Lipid	0.59 (9)	24	0.46 (14)	23	0.49 (11)	23
	Core	−0.32 (10)		0.29 (5)		−0.57 (7)	
	Cavity	−0.61 (5)		0.69 (4)		−0.20 (5)	
TM5	Lipid	0.17 (1)	20	−0.99 (5)	24	0.5 (8)	27
	Core	−0.11 (14)		0.27 (6)		0.39 (7)	
	Cavity	0.35 (5)		0.35 (13)		−1.31 (12)	
TM6	Lipid	0.06 (9)	22	0.58 (10)	23	0.8 (9)	22
	Core	0.11 (7)		−0.65 (10)		0.44 (8)	
	Cavity	−0.29 (6)		0.26 (3)		−0.13 (5)	
Total	Lipid	2.75 (51)	139	1.74 (63)	150	4.31 (67)	152
	Core	−6.15 (52)		−6.86 (52)		−4.91 (46)	
	Cavity	−1.62 (25)		−1.43 (35)		−2.63 (39)	

[a] The hydrophobic scale for individual residue was obtained from Ref. [38].
[b] A more positive number indicates more hydrophobic, and a more negative number denotes more hydrophilic.
[c] Numbers in parentheses indicate number of residues in a TM helix, which belong to a particular category and are included in the calculation.
[d] Total number of residues in a TM helix.

(Table 4, Figs. 2 and 3A). This SNP frequently occurs in the Japanese population and was described as leading to lower surface expression and ATPase activity when compared to the wild-type protein. Patients carrying the Q141K SNP were shown to have elevated drug levels in plasma when treated with topotecan or diflomotecan [22,44–49]. From its location in the molecule, the Q141K polymorphism should not disrupt the folding of the subunit; rather it introduces an additional positive charge in an already very positively charged surface (residues R137, R147, K157, R160, and R163 are in close proximity). Two other NBD mutations, T237V and I239K (OP and SEB, unpublished results), are located near the conserved H-motif in a surface cleft; while the former was fully functional, the latter could not be detected on the cell surface. In the model, the side chain of T237 is fully exposed to the solvent, a prediction that is consistent with the normal activity of the T237V mutant. The side chain of I239, on the other hand, contributes to the hydrophobic core of the protein, and as might be expected from the model, the I239K mutation disrupts the proper folding of the structure and renders the molecule inactive.

3.4. Sequence alignment of the TMD of ABCG2 with members of the ABC transporter family

Next, a structure-based sequence alignment for the TMDs of *Ec*MsbA and *Vc*MsbA was generated, then the sequences of LmrA from *L. lactis*, the N- and C-terminal half of P-gp were aligned to the MsbA sequences, creating a set of reference alignments for the TMD (Fig. 4). Clearly, the reference set of sequences showed considerable similarity, but the alignment of the ABCG2 TMD sequence to the reference set turned out to be much more challenging, giving a very low sequence identity. We further aligned the sequence of the ABCG subfamily protein Drosophila white to the reference set, which again demonstrated a very low sequence similarity, suggesting an overall low homology in the TMD between ABCG subfamily proteins and proteins in the reference set. To evaluate whether the low homology is due in part to the reversed topology of ABCG subfamily proteins, in which the ATP-binding domain is located N-terminal to the TMD, various alignment schemes were devised to test for possible shuffling of TM helices in ABCG2 with respect to those in the reference set. All attempts failed to give a significant improvement in the alignment. As an example, there are only 16 identities and 30 similarities found in the TMDs (Table 1) of ABCG2 and the N-terminal half of P-gp when the order of the TM helices of ABCG2 is reversed.

An alternative (knowledge-based) approach for TMD modeling was taken. We used four different secondary structure prediction programs to identify putative TM segments in the ABCG2 sequence; six to eight TM segments were predicted

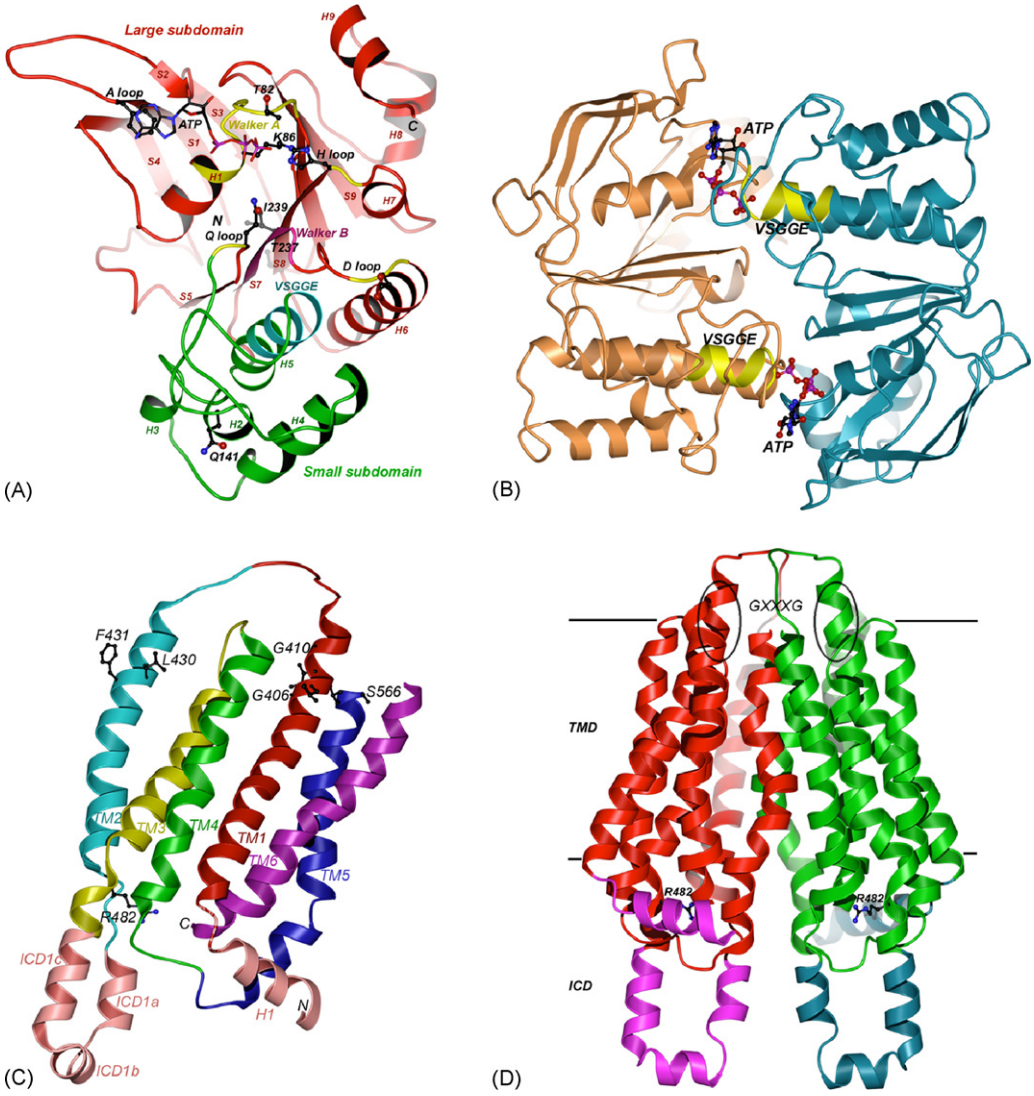

Fig. 3. Structural models of the NBD and the TMD of ABCG2. (A) The monomeric ABCG2 NBD is shown as a ribbon diagram produced with the programs Molscript [62], Bobscript [63], and Povray (http://www.povray.org) interfaced with Gl-render (http://convent.nci.nih.gov/glrhome). The large sub-domain is colored in red and the small sub-domain in green. Secondary structure elements are labeled corresponding to the notations in Fig. 2. Conserved sequence motifs are illustrated in different colors and as labeled, such motifs are the Walker A motif colored in yellow, the Walker B motif in magenta, and the VSGGE signature region in blue. Residues displayed in the ball-and-stick models are at positions where mutations were either identified or made. A bound ATP molecule is also shown, in which carbon atoms are shown in black, nitrogen in blue, oxygen in red. (B) The dimeric ABCG2 NBD model is given in ribbon format viewed along the two-fold symmetry axis. The two subunits are colored cyan and coral, respectively. The two ATP moieties in ball-and-stick model are shown located at the interface between the two monomers, where the nearby VSGGE motifs are also shown in yellow. (C) The monomeric TMD is given as a ribbon diagram showing six TM helices: the TM helices are color coded and as labeled. The ICDs appear in coral and are also as labeled. The residues that are displayed in the ball-and-stick models are those where mutations were found either affecting substrate specificity or changing the dimeric states of the transporter. (D) The dimeric ABCG2 TMD model is given in ribbon format viewed with the two-fold axis vertically placed in the plane of the paper. One monomer is colored red and the symmetry-related one green. The ICDs are colored cyan and magenta, respectively.

(Table 2). A consensus set of TM segments was determined based on the frequency of a particular segment with which it was predicted by different programs. The six predicted consensus TM sequence segments (Table 2) were then mapped to the reference alignment. However, we suspected that the mapping would not be accurate, due to the very low sequence identity, which necessitated adjustments according to rules based on prior knowledge of membrane protein structures; such rules included the ‘‘positive-inside’’ rule [36] and the ‘‘aromatic distribution’’ rule [37].

3.5. Model building for the TM domain

At the time of this writing, there were three bacterial ABC transporter crystal structures available to serve as potential templates for modeling the TMD: *Ec*MsbA [24], BtuCD [27], and *Vc*MsbA [25]. We eliminated BtuCD because of its 20 TM segments, even though its NBD exhibited a polypeptide fold and conformation similar to those observed in structures of isolated NBDs. We evaluated in detail the two MsbA structures available to us and concluded that *Vc*MsbA would serve as a

Table 4
Locations of mutations as predicted by the ABCG2 model and functional correlation

Mutation	Position in ABCG2	Phenotype	Reference
V12M	N-terminal	Membrane localization, SNP, and somewhat lower expression and lower resistance	[22]
S25P[a]	N-terminal	Low drug resistance for the cell line due to lower expression at cell surface	[42]
T82A[a]	NBD, Walker A	Low drug resistance for the cell line due to lower expression at cell surface	[42]
K86M	NBD, Walker A	No expression at cell surface, retained in the Golgi	[43]
K86I	NBD, Walker A	No expressed at cell surface	[43]
Q141K	NBD	SNP with lower protein expression and low drug resistance	[22,23]
T237V	NBD	Fully functional	[b]
I239K,R	NBD	Loss of expression may be due to structural disruption	[b]
R309G	Linker[c]	Low drug resistance	[42]
Δ315-6	Linker	Deletion mutant for A315 and T316. It is a splicing variant with a somewhat lower expression and lower resistance	[23]
A347T[a]	Linker	Low drug resistance	[42]
G406L, G410L, G406L/G410L	GXXXG motif TM1	Dimerization in related proteins. Expressed on surface. Exhibit reduced transport of mitoxantrone, pheophorbide and basal ATPase activity	[17]
G406A, G410A, G406A/G410A	TM1	Fully functional	[b]
E446A,D,G,H,K,R,S	TM2	Loss of drug resistance	[21]
R482G,T,M,S, N,D,K,Y	TM3	T or G change produces gain of function mutant. G mutant is critical to Rhodamine 123 but not to some antibiotics. T mutant binds more tightly to Rhodamin123 and inhibits rhodamine123 transport	[19,51,21]
G553L, G553E	TM5	In drosophila, mutation at this position yields monomer. In the ABCG2 model, this residue is involved in dimerization. The mutant showed low surface expression and appeared to be ER localized in mammalian cells. In sf9 cell, it is expressed on cell surface, but with no ATPase activity	[56]
L554P	TM5	Lowered drug resistance	[42]
N557D,E	TM5	Functional	[21]
S566A[a]	ECL (between TM5 and 6)	Lowered drug resistance for the cell line	[42]
N596Q	Between TM5 and 6	N-glycosylation site	[65]
Y605C[a]	Loop between TM5 and 6	Lowered drug resistance for the cell line	[42]
D620N	Loop between TM5 and 6	SNP polymorphism	[22]
H630E,L	TM6	Functional	[21]
A632V[a]	TM6	Lowered drug resistance for the cell line	[42]

[a] Mutants not well characterized.
[b] OP and SEB (unpublished results).
[c] Linker between NBD and TMD.

better template for ABCG2 modeling for four reasons: it showed a slightly better sequence similarity to the ABCG2 TMD (Table 1), had a higher resolution, included amino acid side chains in the model, and contained a complete NBD. More importantly, the TMD-TMD interface of *Vc*MsbA is also widely regarded as being more likely to represent one of the physiological conformations of MsbA and is consistent with available biochemical data [25]. In the TMD of the MsbA structures, there are a number of helices that contribute to the so-called intracellular domain (ICD1, ICD2 and ICD3); ICD1 is located in the sequence between TM2 and TM3, ICD2 between TM4 and TM5, and ICD3 after TM6. Some of them are physical extensions of TM helices. These ICD helices can also be identified in the P-gp sequence, for both the C- and N-terminal half, and in the LmrA sequence (Fig. 4). In the ABCG subfamily and possibly in Pdr5, the ICD helices are only partially conserved, as illustrated in Fig. 4. In ABCG2, the sequence between TM2 and TM3 mirrors the ICD1 in MsbA; however, the ICD2 appears to be missing and ICD3 is most likely present in the linker region between the NBD and TMD. Since the ICD1s are observed physically coupling to the NBD and represent a conserved feature in all known ABC transporter structures, such a feature should be preserved in the modeling of ABCG2. The sequence segment of ICD1 (449–478) was subject to secondary structure prediction by various programs, giving a helical content in the range between 32 and 83%. These tests gave us confidence to further align the ICD1 sequence to the three helical motifs in the ICD1 of the template as ICD1a, ICD1b and ICD1c (Fig. 4).

Following a similar procedure to modeling of the NBD, the ABCG2 TM model was built initially for the monomer based on the *Vc*MsbA coordinates. Subsequently, a symmetric dimer

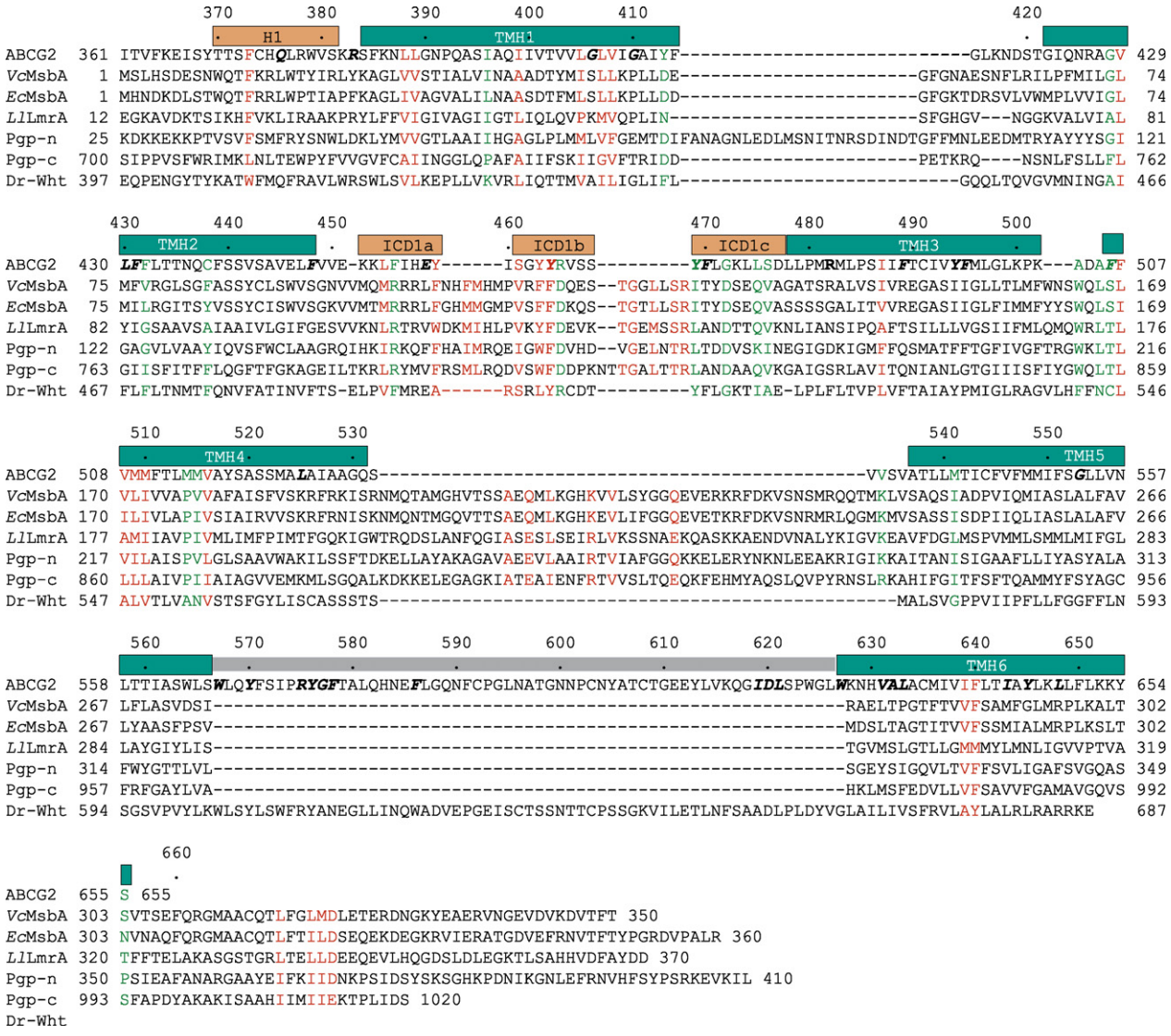

Fig. 4. Structure-based sequence alignment of the TM domains. Seven sequences are aligned: the human ABCG2; *Vc*MsbA, the MsbA of the *V. cholera; Ec*MsbA, the MsbA from *E. coli*; *Ll*LmrA, the *L. lactis* LmrA; Pgp-n, the N-terminal TM domain of the human P-gp; Pgp-c, the C-terminal TM domain of P-gp; Dr-Wht, the white protein from *Drosophila melanogaster*. The modeled TM helices are indicated by green boxes above the sequence and are as labeled. The orange boxes are ICDs. The gray bar indicates a missing part in the modeled structure. Highly conserved residues appear in red and conserved residues are in green. At positions in the TMD sequence where mutations are known, the residues are highlighted in bold–italic type.

was assembled. In this preliminary TM model (initial model), we emphasized that the residue G553 in TM5 should be located at the dimer interface based on homology with mutations in the Drosophila white protein [50] and on the fact that the side chains of residues R482 must be facing the interior cavity because mutations at this position showed a gain-of-function effect, suggesting a role in substrate binding or translocation [19,21,51].

3.6. Model-based site-directed mutagenesis and model refinement

Residues in a TM helix are grouped into three categories: lipid-facing, interior cavity-facing, and hydrophobic core. The lipophilic index (hydrophobicity index) for each category of residues in the TM helix can be estimated based on available empirical amino acid hydrophobicity scales [38,52,53]. Such analyses with known structures demonstrated that for integral membrane proteins, lipid-facing residues are much more hydrophobic than those residing in the protein's hydrophobic core [54]. We calculated the lipophilic indices for the three categories of residues in TM helices of the initial ABCG2 model, using the lipophilic propensity scales developed by Adamian et al. [38], and compared the resulting indices to those for the TM helices of *Vc*MsbA and *St*MsbA (Table 3). As expected for the two MsbA proteins with known structures, there is a strong similarity in the distribution of the lipophilic indices for the three categories of residues in each TM helix with the lipid-facing residues being the most hydrophobic. Except for the TM2, all TM helices in the ABCG2 initial model

displayed distributions that were comparable to those of known structures (Table 3).

In the initial model, TM2 was one of the TM helices involving dimer contact and the TM2 residue F431, highly conserved in the ABCG subfamily of transporters, was identified at the dimer interface interacting with V556 of TM5 of the neighboring subunit. However, the calculated lipophilic indices (−0.42, −1.27, −0.54) for the three categories of residues in the helix showed a large deviation from values typical for TM helices (Table 3). We reasoned that if F431 were important for dimer contact, a mutation designed to disrupt the interaction at this position would impair the function of the transporter. Otherwise, residue F431 should be modeled to face the lipid bilayer to meet the hydrophobicity requirement. To test these possibilities, we made two mutants, F431S and V556S, and probed cell surface expression with the ABCG2 specific monoclonal antibody 5D3 [39] (Fig. 5A). Two clones were tested for each mutant and both were found to

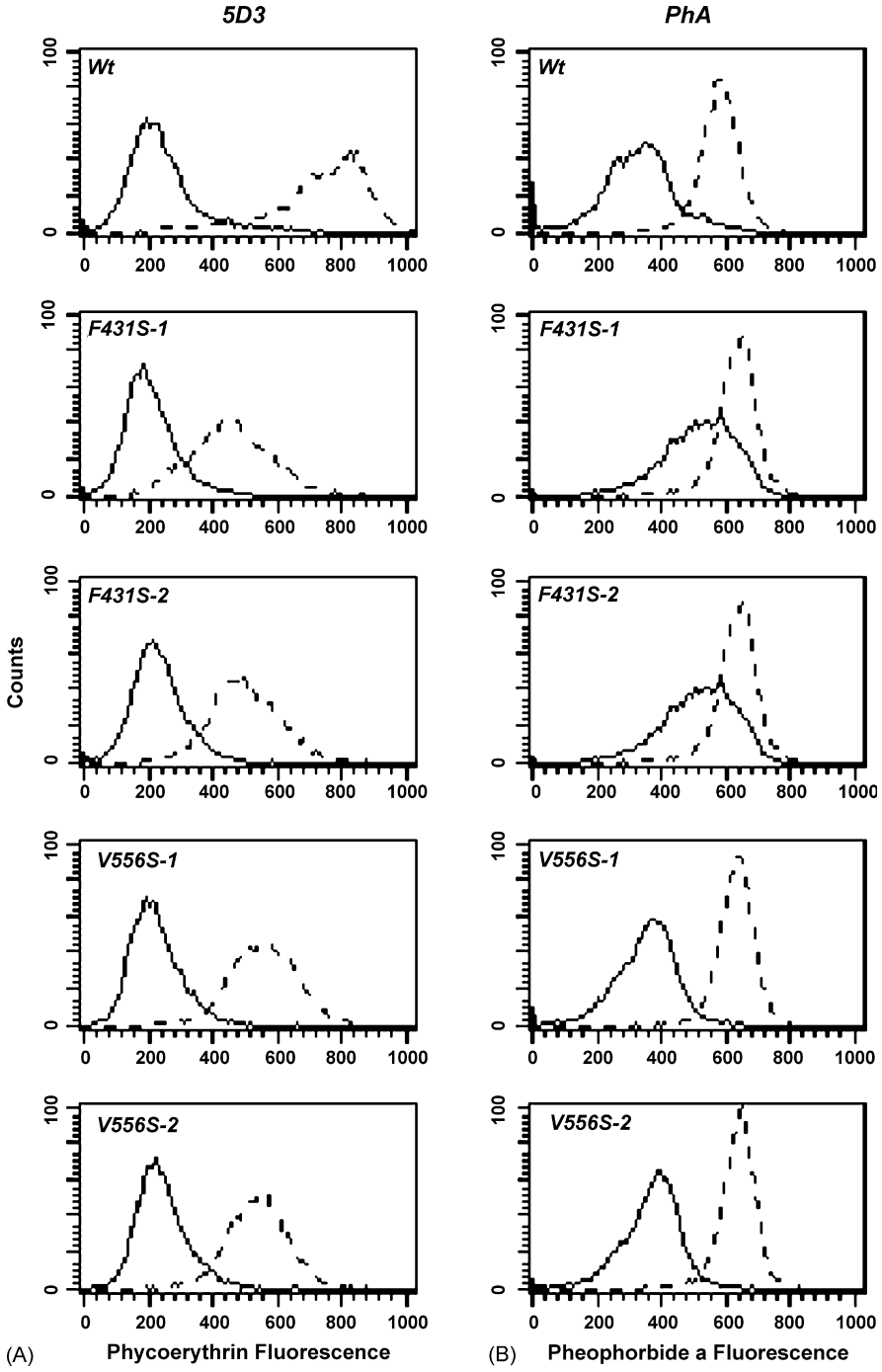

Fig. 5. Flow cytometry analysis of ABCG2 mutants F431S and V556S. (A) Using the 5D3 surface antibody, the F431S and V556S mutant proteins are detectable on the cell surface of HEK293 cells. Two clones for each mutant are shown. Levels of surface expression are compared to the wild-type (Wt) transfectant. ABCG2 (- - -), negative control (—). (B) Transport capacity for pheophorbide a (PhA) for the same cell lines. Accumulations without the ABCG2 inhibitor FTC (—) and with FTC (- - -) are shown.

appear on the cell surface, even though the level of cell-surface expression was not as high as for the wild type. The mutant proteins were active, as judged by pheophorbide a (PhA) retention in the presence of the ABCG2 specific inhibitor FTC (Fig. 5B) [39]. Therefore, both mutants were capable of forming functional dimers, suggesting that residue F431 is not located at the dimer interface as the initial model predicted. To comply with the mutagenesis data, we shifted the sequence alignment for TM2 upwards by two residues, making F431 face the lipid bilayer. Once this was done, the refined distribution of the lipophilic indices of TM2 appeared consistent with those for other TM helices (Table 3).

The refinement led to a final TM model that was consistent with most mutagenesis data available in the literature (Table 4) as well as with the above-mentioned new information. In particular, the residue V556 in the final model faces G498 in the TM3 helix of the neighboring subunit, explaining the phenotype for the V556S mutant because it faces a small residue. Despite an overall low similarity in the final sequence alignment (20% in Table 1), we are quite confident in this model, as it incorporates mutagenesis results and is in agreement with distributions of calculated hydrophobicity indices between the model and experimental structures. The final TMD model for the monomer is given in Fig. 3C and that for the dimer in Fig. 3D, which was generated by applying a two-fold symmetry operation based on that of the *Vc*MsbA dimer. The final TMD model has an rms deviation of 1.05 Å for 206 Cα atoms when superimposed to the *Vc*MsbA transmembrane domain.

3.7. Structure of the dimeric ABCG2 TM domain and the dimer interface

Our final model for the ABCG2 TMD starts at the residue T370 and ends at S655 with a gap between the TM5 and TM6 (Fig. 4). The monomeric TMD is rather compact with the six TM helices roughly positioned parallel to each other (Fig. 3C). When the dimer forms, the two subunits enclose a large, central cavity, representing the most prominent feature of the TM model. Residues from TM3 and TM5 contribute most to the cavity wall, with a few residues from TM1, TM6, and TM4 but none from TM2. The dimer interface of the TMD, on the other hand, is formed mostly by residues from TM2 and TM3 of one monomer and by those from TM5 of its symmetry mate. Residues from TM1 and TM6 near the extracellular side are also involved in the interface formation due to the cone-shaped TMD. No residues from TM4 are found at the dimer interface. If the monomeric TMD is viewed as a right-parallelpiped embedded in the membrane bilayer (Fig. 3D), its orientation is slightly tilted with respect to the normal of the membrane plane and there is a $\sim 30^{\circ}$ angle between the two subunits when the two monomers are packed against each other.

The GXXXG sequence motif in the ABCG2 has been investigated as a potential dimerization motif by site-directed mutagenesis (Table 4, [17]). Leucine substitutions, for example, resulted in impaired ATP hydrolysis and substrate transport, as well as increased degradation. Therefore, the mutant proteins are likely to be improperly folded or assembled. As one of the most frequently occurring sequence motifs between interacting TM helices [55], the GXXXG motif in ABCG2 was suggested to form a symmetrical interaction in the TM dimer [17]. In the current TMD model, however, this motif is situated near the extracellular side of the membrane, presumably forming part of the exit passage for substrates and potentially interacting with residues from TM5 of the symmetry-related subunit. The 3-D model suggests that the possibility for the two GXXXG motifs between the symmetry mates to engage in a dimeric interaction is small. As the GXXXG motif is not conserved in the ABC transporter family and only partially conserved in the ABCG subfamily proteins (Fig. 4), it is possible that the ABCG subfamily proteins use a mechanism that requires transient or dynamic interaction between the GXXXG motifs during a catalytic cycle. Whether the GXXXG motif really involves dimer interaction requires further experimental verification. Our current model, however, is consistent with the observed mutational phenotypes for the GXXXG motif. Mutants with smaller substitutions such as G406/410A appeared normal in transport activity (Table 4), whereas larger substitutions such as G406/410L could cause structural perturbation to the interface, blocking substrate passage and/or interfering with closure of the passage, and therefore rendering the mutants defective in drug transport.

Mutation at G553 (TM5) led to ER localization of ABCG2 in mammalian cells [56], while mutations in the corresponding residue in the *Drosophila* white protein have been presumed to disrupt the white-brown heterodimer formation (Table 4) [50]. In our model, G553 is located at the dimer interface facing the side chain of Y494 from the neighboring subunit. Introducing a larger or charged residue at this position (G553L and G553E) would result in a clash of amino acid side chains and disrupt the dimer formation both in ABCG2 and in its *Drosophila* orthologue.

It has been hypothesized that residue R482 in TM3 is likely to interact with substrates based on the effect of R482G/T mutations [19]; both are gain-of-function mutants, conferring resistance to a broader range of substrates than the wild-type transporter including rhodamine 123 and doxorubicin [19]. These results have been extended by mutagenesis studies showing that replacement of arginine with virtually any residue that was not positively charged led to a similar gain-of-function or change-of-function [57]. Interestingly, no single nucleotide polymorphisms (SNPs) were identified at this position, and methotrexate, for example, was found only to be transported by the wild-type protein [58]. We speculated that R482 must be part of the substrate-binding pocket in the structure. In our model, R482 is placed at the beginning of the TM3 on the cytoplasmic side of the membrane, and its side chain points toward the interior of the cavity, forming part of the surface of the central TM cavity and to a cluster of positively charged residues (R482, K652, K452, K543). The mutation would obviously enlarge the cavity and disrupt the positively charged cluster, which may be the reason for the wider substrate specificity.

3.8. Modeling of the complete dimeric ABCG2 and interactions between the NBD and TMD

Since homology and knowledge-based modeling were employed for building the NBD and the TMD, respectively, three degrees of freedom need to be fixed in order to make a complete ABCG2 model: there are two possible ways of orienting the two-fold axis of NBD to align with that of the TMD; there is a rotational freedom of the NBD about the two-fold axis, and a translational freedom along the two-fold axis (distance) between the NBD and TMD. First, in aligning the two two-fold axes, one possible orientation was eliminated, as the Q motif in the NBD (Fig. 2) must be situated close to the TMD [27]. To fix the rotational and translational degrees of freedom, we analyzed the four known structures of full-length ABC transporters and found a highly conserved contact

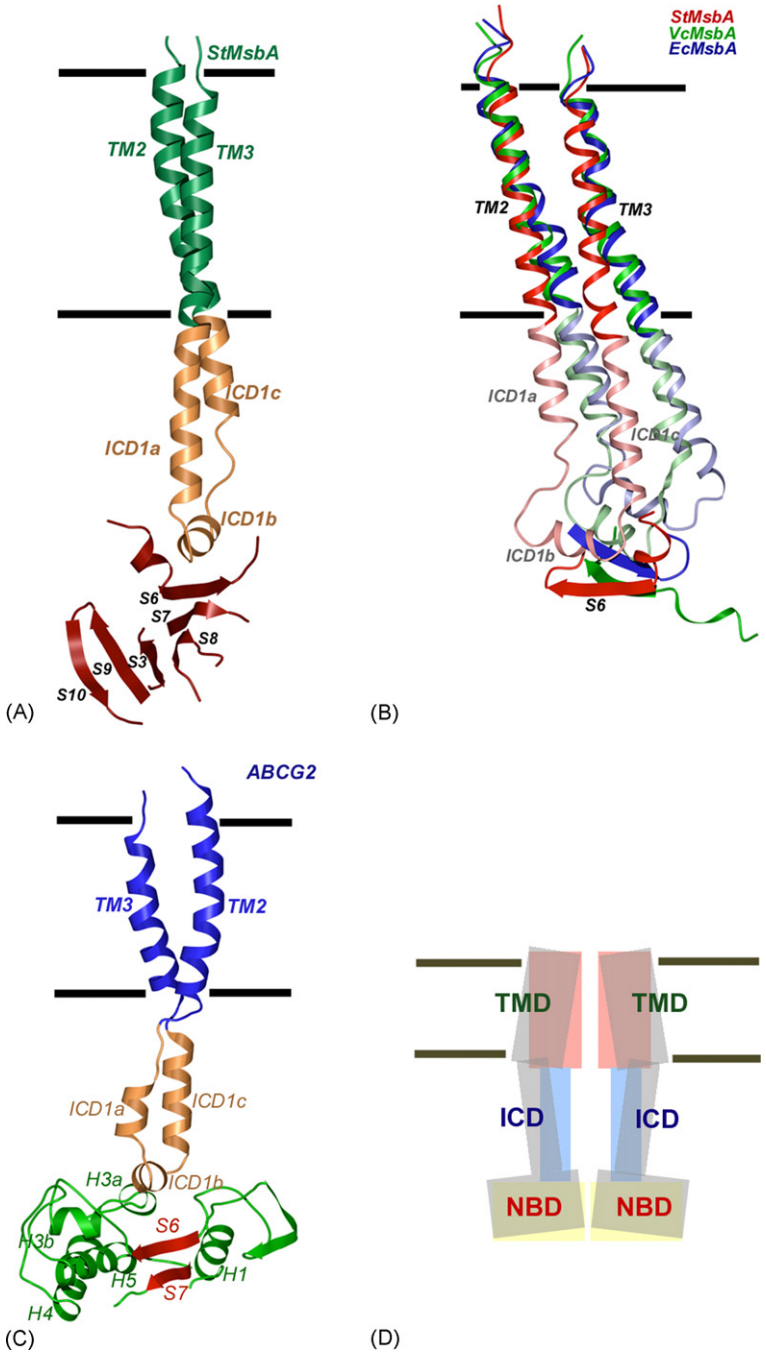

Fig. 6. Interactions between the TMD and the NBD in ABC transporters. (A) In *St*MsbA, the ICD1 helices (coral) of TMD are extensions of the helices from TM2 and TM3 (green) and are in contact with the terminal β-strand S6, which is one edge of the central β-sheet in the NBD. (B) Structure superposition of TM2 and TM3 from *St*MsbA (red), *Vt*MsbA (green) and *Ec*MsbA (blue). (C) The interface between the TMD and the NBD in the ABCG2 model. The interaction between ICD1 and S6 is preserved. (D) A conceptualized model for the role of ICDs in coupling the conformational changes between TMD and NBD. In one conformation, the TMDs are shown in pink, ICDs in blue and NBDs in yellow. In another conformation, they are shown in gray. The linkage between TMD and NBD via ICD is conserved.

between the middle helix (ICD1b) of the ICD1 in the TMD and the terminal β-strand (strand S6) of the central β-sheet in the NBD. The β-strand S6 is located just before the Q-loop in sequence and conceivably is able to serve as an important constraint for fixing the rotational and translational degrees of freedom for the final assembly (Fig. 6). In all the MsbA structures, the interactions between the ICD1b helix and β-strand S6 are hydrophobic and the shortest distance between Cα atoms of the ICD1b and S6 is roughly 10 Å (Fig. 6A). We also observed in the three MsbA structures (Fig. 6B) that the ICD1 motifs are able to make small orientational adjustments with respect to TM2 and TM3. The same interaction is also observed in the BtuCD structure. Therefore, by twisting the ICD1 motif with respect to TM2 and TM3, this conserved interaction between TM and NBD was preserved in the final assembly and used as a critical constraint for modeling the full-length ABCG2 (Fig. 6C). The observed conservation of the linkage in available structures between TMD and NBD appears to suggest that the ICD1 motif may act as a soft-link engaging in amplifying, attenuating or relaying conformational changes in the NBDs as a result of ATP hydrolysis or in TMDs upon substrate binding (Fig. 6D).

4. Discussion

The final model starts at residue Y44, has one missing fragment of nearly 80 residues between the NBD and the TMD (from N288 to T370) and another between TM5 and TM6 (from W567 to L626), and ends at S655. As it is, the model is already quite informative. The dimeric ABCG2 is about 100 Å in height and 55 Å in width with a shape similar to MsbA (Fig. 7A). The electrostatic surface representation of the model shows a general hydrophobic TM, a positively charged ICD and a slightly negatively charged NBD (Fig. 6B). The 12 TM helices arrange in such a way that a cavity with the shape of an elliptical cone, viewed perpendicular to membrane plane, is formed at the center of the TMD. At the base of the cone, the longer dimension is close to 20 Å and the shorter axis is about 10 Å. The cavity near the extracellular side of the membrane is sealed off but open to the intracellular side facing the NBDs. The model presented here represents a closed conformation of the ABCG2 protein because the NBD model is based on the dimeric structure of *E. coli* MalK in a closed conformation and the ICD1 motifs were twisted in order to make contacts with the S6 strands of the NBDs. Future mutagenesis experiments will concern residues forming the dimer interface between the TM domains and those making contacts between the transmembrane and ATP-binding domains.

To date, the crystal structures of four full-length bacterial ABC transporters and a number of isolated NBDs have been determined. These structures have inspired homology modeling for a few important eukaryotic and prokaryotic ABC transporters including the human multidrug resistance transporter P-gp [59], the bacterial antibiotic resistance transporter LmrA from *L. lactis* [60] and the human multidrug resistance protein MRP1 [61], for which experimental structural information have been difficult to obtain. A common feature

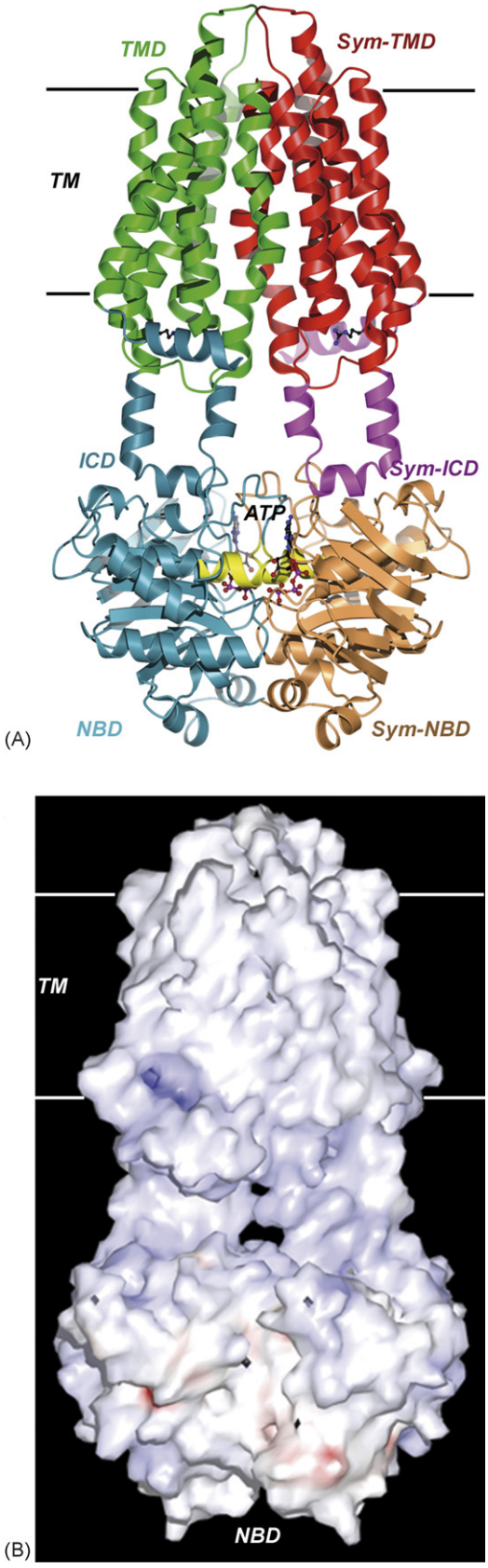

Fig. 7. Structure of the dimeric model of ABCG2. (A) Ribbon diagram of the complete ABCG2 dimeric model. (B) Surface electrostatic rendition of the ABCG2 dimer produced in GRASP [64].

in these studies is the existence of reasonably high degree of sequence identity or similarity to the model template, permitting homology modeling for the full-length molecules. In contrast, transporters in the ABCG subfamily have lower sequence similarity, especially for the TMD, making homology modeling impossible. Additionally, the reversed domain organization and the lack of some conserved structural motifs such as the ICD2 and ICD3 in the ABCG subfamily suggests that a ABCG2 model should not be a simple repeat of those already reported in the literature. Indeed, significant differences exist not only in the way that our ABCG2 model was built but also in structural details of various models, even though other models did use the same initial template. These differences are expected to become more obvious as future experimental data are further incorporated into respective models.

As stressed at the beginning of this report, the purpose of creating a model is to fill a knowledge vacuum created by the lack of structural data for the ABCG subfamily proteins; it is by no means a substitute for a real experimental structure at atomic resolution, nor is it expected to be a final model. Since it is only a model, a realistic evaluation of its quality, reliability and applicability should be provided. As it is, our ABCG2 model fits all experimental data, including structural, genetic, and biochemical studies, available to us. It also satisfies standard stereochemical parameters derived from atomic resolution structures. We are quite confident in the general features of our model in terms of overall folding, the nucleotide binding, the TMD cavity and various interfaces. However, we are less certain about the amino acid side chain positions and their exact neighbors. Currently, we are making a number of mutations based on the model to further improve its quality and accuracy.

Our ABCG2 model is a symmetric homodimer including two ATP moieties and two Mg^{2+} ions. We are aware of a body of published work demonstrating the asymmetric nature of the two NBD domains during a catalytic cycle for well-studied transporters such as P-gp. However, the lack of knowledge both in the causality of the asymmetry at the molecular level for ABCG2 and in the subtlety of modeling conformational changes did not permit asymmetric modeling of the two NBDs.

Acknowledgements

The authors wish to thank Drs. Michael M. Gottesman and Suresh V. Ambudkar of LCB for their critical comments, and George Leiman for editorial assistance. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

[1] M. Dean, A. Rzhetsky, R. Allikmets, The human ATP-binding cassette (ABC) transporter superfamily, Genome Res. 11 (2001) 1156–1166.
[2] H. Nikaido, J. Hall, Overview of bacterial ABC transporters, Meth. Enzymol. 292 (1998) 3–20.
[3] R.L. Juliano, V.A. Ling, A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants, Biochim. Biophys. Acta 455 (1976) 152–162.
[4] K. Ueda, C. Kardarelli, M.M. Gottesman, I. Pastan, Expression of a full-length cDNA for the human MDR1 gene confers resistance to colchicine, doxorubicin, and vinblastine, Proc. Natl. Acad. Sci. USA 84 (1987) 3004–3008.
[5] S.P. Cole, G. Bhardwaj, J.H. Gerlach, J.E. Mackie, C.E. Grant, K.C. Almquist, A.J. Stewart, E.U. Kurz, A.M. Duncan, R.G. Deeley, Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line, Science 258 (1992) 1650–1654.
[6] K. Miyake, L. Mickley, T. Litman, Z. Zhan, R. Robey, B. Cristensen, M. Brangi, L. Greenberger, M. Dean, T. Fojo, S.E. Bates, Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone resistance cells: Demonstration of homology to ABC transport genes, Cancer Res. 59 (1999) 8–13.
[7] L.A. Doyle, W. Yang, L.V. Abruzzo, T. Krogmann, Y. Gao, A.K. Rishi, D.D. Ross, A multidrug resistance transporter from human MCF-7 brest cancer cells, Proc. Natl. Acad. Sci. USA 95 (1998) 15665–15670.
[8] J.M. Rommens, M.C. Iannuzzi, B. Kerem, M.L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J.L. Cole, D. Kennedy, N. Hidaka, et al., Identification of the cystic fibrosis gene: chromosome walking and jumping, Science 245 (1989) 1059–1065.
[9] J.R. Riordan, J.M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J.L. Chou, et al., Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA, Science 245 (1989) 1066–1073.
[10] B. Kerem, J.M. Rommens, J.A. Buchanan, D. Markiewicz, T.K. Cox, A. Chakravarti, M. Buchwald, L.C. Tsui, Identification of the cystic fibrosis gene: genetic analysis, Science 245 (1989) 1073–1080.
[11] R. Allikmets, B. Gerrard, A. Hutchinson, M. Dean, Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database, Hum. Mol. Genet. 5 (1996) 1649–1655.
[12] R. Allikmets, L.M. Schriml, A. Hutchinson, V. Romano-Spica, M. Dean, A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance, Cancer Res. 58 (1998) 5337–5339.
[13] T. Litman, T.E. Druley, W.D. Stein, S.E. Bates, From MDR to MXR: New understanding of multidrug resistance systems, their properties and clinical significance, CMLS Cell Mol. Life Sci. 58 (2001) 931–959.
[14] E. Balzi, M. Wang, S. Laterme, L. van Dyck, A. Goffeau, PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDR1, J. Biol. Chem. 269 (1994) 2206–2214.
[15] A. Bhatia, H.J. Schafer, C.A. Hrycyna, Oligomerization of the human ABC transporter ABCG2: evaluation of the native protein and chimeric dimers, Biochemistry 44 (2005) 10893–10904.
[16] T. Litman, U. Jensen, A. Hansen, K. Covitz, Z. Zhan, P. Fetsch, A. Abati, P.R. Hansen, T. Horn, T. Skovsgaard, S.E. Bates, Use of peptide antibodies to probe for the mitoxantrone resistance associated protein MXR/BCRP/ABCP/ABCG2, Biochim. Biophys. Acta 20 (2002) 6–16.
[17] O. Polgar, R.W. Robey, K. Morisaki, M. Dean, C. Michejda, Z.E. Sauna, S.V. Ambudkar, N. Tarasova, S.E. Bates, Mutational analysis of ABCG2: role of the GXXXG Motif, Biochemistry 43 (2004) 9448–9456.
[18] O. Alqawi, S. Bates, E. Georges, Arginine482 to threonine mutationn in the breast cancer resistance protein ABCG2 inhibits rhodamine 123 transport while increasing binding, Biochem. J. 382 (2004) 711–716.
[19] Y. Honjo, C.A. Hrycyna, Q.-W. Yan, W.Y. Medina-Perez, R.W. Robey, A. van de Laar, T. Litman, M. Dean, S.E. Bates, Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells, Cancer Res. 62 (2001) 6635–6639.
[20] U. Henriksen, U. Gether, T. Litman, Effect of Walker A mutation (K86M) on oligomerization and surface targeting of the multidrug resistance transpsorter ABCG2, J. Cell Sci. 118 (2005) 1417–1426.
[21] M. Miwa, S. Tsukahara, E. Ishikawa, S. Asada, Y. Imai, Y. Sugimoto, Single aminno acid substitutions in the tramsmembrane domains of breast resistance protein (BCRP) alter cross resistance patterns in transfectants, Int. J. Cancer 107 (2003) 757–763.
[22] Y. Honjo, K. Morisaki, L.M. Huff, R.W. Robey, M. Dean, S.E. Bates, Single-nucleotide polymorphism (SNP) analysis in the ABC half-

transporter ABCG2 (MXR/BCRP/ABCP1), Cancer Biol. Ther. 1 (2002) 696–702.
[23] Y. Imai, M. Nakane, K. Kage, S. Tsukahara, E. Ishikawa, T. Tsuruo, Y. Miki, Y. Sugimoto, C421A polymorphism in the human breast cancer resistance protein gene is associated with low expression of Q141K protein and low-level drug resistance, Mol. Cancer Therapeutics 1 (2002) 611–616.
[24] G. Chang, C.B. roth, Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters, Science 293 (2001) 1793–1800.
[25] G. Chang, Structure of MsbA from *Vibrio cholera*: a multidrug resistance ABC transporter homolog in a closed conformation, J. Mol. Biol. 330 (2003) 419–430.
[26] C.L. Reyes, G. Chang, Structure of the ABC transporter MsbA in complex with ADP·vanadate and lipopolysaccharide, Science 308 (2005) 1028–1031.
[27] K.P. Locher, A.T. Lee, D.C. Rees, The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism, Science 296 (2002) 1091–1098.
[28] A.L. Davidson, J. Chen, ATP-binding cassette transporters in bacteria, Annu. Rev. Biochem. 73 (2004) 241–268.
[29] M.C. Peitsch, ProMOd and Swiss-model: internet-based tools for automated comparative protein modeling, Biochem. Soc. Trans. 24 (1996) 274–279.
[30] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for binding protein models in electron density maps and the location of errors in these models, Acta Crystallogr. A47 (1991) 110–119.
[31] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kanstleve, J. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr. D54 (1998) 905–921.
[32] T. Hirokawa, S. Boon-Chieng, S. Mitaku, SOSUI: classification and sedondary structure prediction system for membrane proteins, Bioinformatics 14 (1998) 378–379.
[33] B. Rost, PHD: predicting one-dimensional protein structure by profile-based neural networks, Meth. Enzymol. 266 (1996) 525–539.
[34] G.E. Tusnady, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics 17 (2001) 849–850.
[35] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[36] G. von Heijne, The distribution of positively charged residues in bacterial inner membrane protein correlates with the trans-membrane topology, EMBO J. 5 (1986) 3021–3027.
[37] G. von Heijne, Membrane proteins: from sequence to structure, Annu. Rev. Biophys. Biomol. Struct. 23 (1994) 167–192.
[38] L. Adamian, V. Nanda, W.F. DeGrado, J. Liang, Empirical lipid propensities of amino acid residues in multispan alpha helical membrane proteins, Protein: Struct. Funct. Bioinform. 59 (2005) 486–509.
[39] R.W. Robey, K. Steadman, O. Polgar, K. Morisaki, M. Blayney, P. Mistry, S.E. Bates, Pheophorbide a is a specific probe for ABCG2 function and inhibition, Cancer Res. 64 (2004) 1242–1246.
[40] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[41] J. Chen, G. Lu, J. Lin, A.L. Davidson, F.A. Quiocho, A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle, Mol. Cell 12 (2003) 651–661.
[42] K. Kage, S. Tsukahara, T. Sugiyama, S. Asada, E. Ishikawa, T. Tsuruo, Y. Sugimoto, Dominant-negative inhibition of breast cancer resistance protein as drug efflux pump through the inhibition of S–S dependent homodimerization, Int. J. Cancer 97 (2002) 626–630.
[43] C. Ozvegy, A. Varadi, B. Sarkadi, Characterization of drug transport, ATP hydrolysis, and nucleotide trapping by the human ABCG2 multidrug transporter. Modulation of substrate specificity by a point mutation, J. Biol. Chem. 277 (2002) 47980–47990.
[44] K. Morisaki, R.W. Robey, C. Ozvegy-Lacka, Y. Honjo, O. Polgar, K. Steadman, B. Sarkadi, S.E. Bates, Single nucleotide polymorphisms modify the transporter activity of ABCG2, Cancer Chemother. Pharmacol. 56 (2005) 161–172.
[45] C. Kondo, H. Suzuki, M. Itoda, S. Ozawa, J. Sawada, D. Kobayashi, I. Ieiri, K. Mine, K. Ohtsubo, Y. Sugiyama, Functional analysis of SNPs variants of BCRP/ABCG2, Pharm Res. 21 (2004) 1895–1903.
[46] D. Kobayashi, I. Ieiri, T. Hirota, H. Takane, S. Maegawa, J. Kigawa, H. Suzuki, E. Namba, M. Oshimura, N. Terakawa, K. Otsubo, K. Mine, Y. Sugiyama, Functional assessment of ABCG2 (BCRP) gene polymorphisms to protein expression in human placenta, Drug Metab. Dispos. 33 (2005) 94–101.
[47] S. Mizuarai, N. Aozasa, H. Kotani, Single nucleotide polymorphisms result in impaired membrane localization and reduced ATPase activity in multidrug transporter ABCG2, Int. J. Cancer 109 (2004) 238–246.
[48] A. Sparreboom, W.J. loos, H. Burger, T.M. Sissung, J. Yerweij, W.D. Figg, K. Nooter, H. Gelderblom, Effect of ABCG2 genotype on the oral bioavailability of fTopotecan, Cancer Biol Ther. 4 (2005) 650–658.
[49] A. Sparreboom, H. Gelderblom, S. Marsh, R. Ahluwalia, R. Obach, P. Principe, C. Twelves, J. Verweij, H.L. McLeod, Diflometecan pharmacokinetics in relation to ABCG2 421C > A genotype, Clin. Pharmacol. Ther. 76 (2004) 38–44.
[50] S.M. Mackenzie, M.R. Brooker, T.R. Gill, G.B. Cox, A.J. Howells, G.D. Ewart, Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration, Biochim. Biophys. Acta 1419 (1999) 173–185.
[51] T. Janvilisri, S. Shahi, H. Venter, L. Balakrishnan, H.W. van Veen, Arginine-482 is not essential for transport of antibiotics, primary bile acids and unconjugated sterols by the human breast cancer resistance protein (ABCG2), Biochem. J. 385 (2005) 419–426.
[52] J. Kyte, R.F. Doolittle, A simple method for displaying the hydrophathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[53] S. Jayasinghe, K. Hristova, S.H. White, Energetics, stability, and prediction of transmembrane helices, J. Mol. Biol. 312 (2001) 927–934.
[54] D.C. Rees, L. DeAntonio, D. Eisenberg, Hydrophobic organization of membrane proteins, Science 245 (1989) 510–513.
[55] W.P. Russ, D.M. Engelman, The GxxxG motif: a framework for transmembrane helix–helix association, J. Mol. Biol. 296 (2000) 911–919.
[56] O. Polgar, C. Ozvegy-Laczka, R.W. Robey, K. Morisaki, M. Okada, A. Tamaki, G. Koblos, N.B. Elkind, Y. Ward, M. Dean, B. Sakadi, S.E. Bates, Mutational studies of G553 in TM5 of ABCG2: a residue potentially involved in dimerization, Biochemistry 45 (2006) 5251–5260.
[57] K.F.K. Ejendal, N.K. Diop, L.C. Schweiger, C.A. Hrycyna, The nature of amino acid 482 of human ABCG2 affects substrate transport and ATP hydrolysis but not substrate binding, Protein Sci. 15 (2006) 1597–1607.
[58] Z.S. Chen, K. Lee, S. Walther, R.B. Raftogianis, M. Kuwano, H. Zeng, G.D. Kruh, Analysis of methotrexate and folate transport by multidrug resistance protein 4 (ABCC4): MRP4 is a component of the methotrexate efflux system, Cancer Res. 62 (2002) 3144–3150.
[59] D.R. Stenham, J.D. Campbell, M.S.P. Sansom, C.F. Higgins, I.D. Kerr, K.J. Linton, An atomic detail model for the human ATP binding cassett transporter P-glycoprotein derived from disulfide cross-linking and homology modeling, FASEB J. 17 (2003) 2287–2289.
[60] K. Pleban, A. Macciarulo, G. Costantino, R. Pellicciari, P. Chiba, G.F. Ecker, Homology model of the multidrug transporter LmrA from *Lactococcus lactis*, Bioorg. Med. Chem. Lett. 14 (2004) 5823–5826.
[61] J.D. Campbell, K. Koike, C. Moreau, M.S. Sansom, R.G. Deeley, S.P.C. Cole, Molecular modeling correctly predicts the functional importance of Phe594 in transmembrane helix 11 of the multidrug resistance protein MRP1 (ABCC1), J. Biol. Chem. 279 (2004) 463–468.
[62] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Crystallogr. 24 (1991) 946–950.
[63] R.M. Esnouf, An extensively modified version of Molscript that includes greatly enhanced coloring capabilities, J. Mol. Graph. 15 (1997) 133–138.
[64] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons., Proteins: Struct. Funct. Genet. 11 (1991) 281–296.
[65] N.K. Diop, C.A. Hrycyna, N-linked glycosylation of the human ABC transpoerter ABCG2 on asparaginge 596 is not essential for expression, transport activity, or trafficking to the plama membrane, Biochemistry 44 (2005) 5420–5429.