JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

The Effect of Environment on the Structure of a Membrane Protein: P-Glycoprotein under Physiological Conditions

Megan L. O'Mara[†] and Alan E. Mark*[,†,‡]

[†]School of Chemistry and Molecular Biosciences (SCMB), University of Queensland, Brisbane, QLD 4072, Australia

[‡]Institute for Molecular Bioscience (IMB), University of Queensland, Brisbane, QLD 4072, Australia

Supporting Information

ABSTRACT: The stability of the crystal structure of the multidrug transporter P-glycoprotein proposed by Aller et al. (PDBid 3G5U) has been examined under different environmental conditions using molecular dynamics. We show that in the presence of the detergent cholate, the structure of P-glycoprotein solved at pH 7.5 is stable. However, when incorporated into a cholesterol-enriched POPC membrane in the presence of 150 mM NaCl, the structure rapidly deforms. Only when the simulation conditions closely matched the experimental conditions under which P-glycoprotein is transport active was a stable conformation obtained. Specifically, the presence of Mg^{2+}, which bound to distinct sites in the nucleotide binding domains (NBDs), and the double protonation of the catalytic histidines (His583 and His1228) and His149 were required. While the structure obtained in a membrane environment under these conditions is very similar to the crystal structure of Aller et al., there are several key differences. The NBDs are in direct contact, reminiscent of the open state of MalK. The angle between the transmembrane domains is also increased, resulting in an outward motion of the intracellular loops. Notably, the structures obtained from the simulations provide a better match to a range of experimental cross-linking data than does the original 3G5U-a crystal structure. This work highlights the effect small changes in environmental conditions can have of the conformation of a membrane protein and the importance of representing the experimental conditions appropriately in modeling studies.

INTRODUCTION

The ABC transporter P-glycoprotein is one of the principal multidrug transporters in humans.[2] P-glycoprotein (P-gp) modulates the pharmacokinetic properties of many clinically important drugs, influencing their rate of cellular uptake and clearance and their tissue distribution. P-gp has broad substrate specificity and has been implicated in chemotherapy resistance in cancer. Despite intensive research over 30 years,[3] the mechanism by which P-gp transports drugs and endogenous substrates is still poorly understood. P-gp is a full-length ABC transporter. It is 1280 amino acids in length and is composed of two heterodimers fused together into a single polypeptide chain. Each heterodimer consists of an N-terminal transmembrane domain (TMD) and a C-terminal nucleotide binding domain (NBD). The two half-transporters are connected by a linker sequence approximately 60 amino acids in length.

In 2009, a series of structures of P-gp were published with and without ligand. These structures, which were solved at resolutions between 3.8 and 4.4 Å,[1] were crystallized in the presence of the detergent sodium cholate but without nucleotides or Mg^{2+}, both of which are known to be required for activity. The structure of nucleotide free P-gp proposed by Chang and co-workers (PDBid 3G5U) based on X-ray crystallographic data spans 92% of the protein sequence and displays the characteristic ABC exporter domain-swapping motif first observed in Sav1866.[4] In both the Sav1866 and P-gp structures, two of the long helical intracellular loops (ICLs) that extend from each transmembrane domain (TMD) contact the adjacent nucleotide binding domain (NBD), while one ICL contacts the opposite NBD.[1,4] However, in contrast to the nucleotide free structures of other well-characterized ABC transporters such as the maltose transporter system,[5−7] the NBDs in P-gp were widely splayed and separated by a distance of approximately 3.0 nm. In addition, Aller et al. were unable to soak a range of substrates into the crystals or to cocrystallize known substrates with P-gp under the conditions used for crystallization. The P-gp inhibitors cocrystallized in the two ligand-containing structures were cyclic peptides that had not been characterized previously and differ significantly from other physiologically important P-gp substrates or inhibitors.

The crystallization of integral membrane proteins such as P-gp is difficult. The protein has to be removed from its native membrane environment, purified, and stabilized by a detergent. A variety of conditions can be used to obtain crystals of sufficient quality to determine the structure of an integral membrane protein, and there is continuing debate as to whether or not the structures obtained in a nonmembrane environment necessarily correspond to a physiologically relevant state. For example, in the crystal structure of the voltage gated potassium channel, KvAP, the voltage sensor paddles became partially detached from the channel pore during crystallization.[8] While the structures of each KvAP domain have been validated, it was subsequently shown that the structure of the pore and paddle complex represented a nonphysiological conformation due to the presence of a Fab tag used to promote crystallization.[9−11]

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: March 29, 2012

dx.doi.org/10.1021/ct300254y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

In the case of P-gp, the structure was solved in the presence of the ionic detergent sodium cholate, which was required to maintain P-gp in a nonaggregated solubilized state and to facilitate the process of crystallization. While the presence of detergent is clearly required to crystallize most membrane proteins, it is also known that the choice of detergent can affect the stability and also the conformation of the protein.[12]

In this paper, the effect of environment on the conformational properties of nucleotide-free P-gp has been investigated. It was found that the presence of cholate stabilizes the structure of P-gp observed crystallographically. The proposed structure was, however, not stable when directly incorporated into a membrane environment. Only under conditions where P-gp is physiologically active, namely, in the presence of Mg^{2+} and when the catalytic histidines (His583 and His1228) and His149 were positively charged, was a stable structure obtained. The structure observed to be stable in the simulations was found to be very similar to that proposed by Chang and co-workers. Nevertheless, there were differences in the relative orientation of the NBDs and the relative positions of the TMDs and ICLs. The stable P-gp conformation derived from the MD simulations was found to be more compatible with a range of experimental cross-linking and cysteine mutagenisis data than the 3G5U crystal structure. The differences between this structure and the original structure proposed by Chang and co-workers are described below.

MATERIALS AND METHODS

Simulation Details. All MD simulations were performed using the GROMACS (Groningen Machine for Chemical Simulation) package, version 3.3.3,[13] in conjunction with the GROMOS 53A6 force field.[14] The simple point charge (SPC) water model[15] was used to describe the solvent water. The initial configuration of P-gp was taken from the structure of P-gp proposed by Aller et al. (PDBid 3G5U molecule a).[1] Full details of how the simulations were performed are provided as Supporting Information.

Before the simulations were performed, the pK_a values of all ionizable groups were predicted using the PROPKA server,[16,17] and the initial protonation states of the relevant groups at a given pH were assigned accordingly. P-gp was crystallized at pH 7.5.[1] On the basis of the pK_a values predicted using the PROPKA server,[16,17] all histidine residues were predicted to be predominately singly protonated at pH 7.5. In general, transport assays are performed at a lower pH of 6.8 to 7.0.[18,19] At a pH of 6.8, the protonation states of several residues is less certain. In particular, the catalytic histidines, His583 and His1228 as well as His149 (from ICL1) could be either doubly or singly protonated. His149 (from ICL1) was predicted to have a pK_a of 7.1. On the basis of the original crystal structure, the catalytic histidines are predicted to be singly protonated at pH 6.8. However, as the catalytic histidines His583 and His1228 interact directly with Glu552 and Glu1197, respectively, they have commonly been assumed to be doubly protonated in previous studies of similar transporters.[20,21] Full details are provided as Supporting Information.

P-Glycoprotein in the Presence of Sodium Cholate. To determine whether the conformation of P-gp observed crystallographically was stabilized by the presence of the detergent, simulations of P-gp were performed in a solution of 75 mM NaCl and containing 130 molecules of NaCholate (as Na^+ and $C_{24}H_{39}O_5^-$). Parameters for cholate were developed using the Automated Topology Builder (ATB, http://compbio.biosci.uq.edu.au/atb/), as described by Malde et al.[22] All histidine residues were considered to be singly protonated as is appropriate for pH 7.5, and eight Cl^- counterions were added to neutralize the charge on the system. The system was equilibrated progressively over 3 ns. New velocities were then assigned and the system simulated for 30 ns with position restraints applied to all backbone atoms using a force constant of 50 $kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$. The system was then simulated for a further 40 ns without restraints. Three independent simulations with different sets of starting velocities were performed. Full details are provided as Supporting Information.

Membrane-Embedded P-Glycoprotein Simulations. To determine whether the proposed structure of P-gp was stable under physiologically relevant conditions, a series of simulations were performed on membrane-bound P-gp in different environments. P-gp was embedded in a cholesterol-enriched POPC (2-oleoyl-1-palmitoyl-*sn*-glycero-3-phosphocholine) bilayer,[23,24] consisting of 268 POPC and 29 cholesterol molecules.[25] Parameters for cholesterol were obtained from the ATB.[22] Simulations were performed under four sets of conditions: in the presence of 150 mM NaCl with all histidine residues singly protonated (series 1); in the presence of 150 mM NaCl with His583 and His1228 doubly protonated and all other histidine residues singly protonated (series 2); in the presence of 150 mM NaCl and 1.5 mM $MgCl_2$ with His583 and His1228 doubly protonated and all other histidine residues singly protonated (series 3); in the presence of 150 mM NaCl and 1.5 mM $MgCl_2$[19] with His583, His1228, and His149 doubly protonated (series 4). In each case, the appropriate number of counterions was included to neutralize the charge on the system. To initiate the simulations, the system was carefully equilibrated while position restraints on the protein were gradually relaxed over 10 ns. New velocities were then assigned and the system simulated without restraints. Series 1, 2, and 3 were each simulated three times for 30 ns. Series 4 was simulated five times for 60 ns. Full details are provided as Supporting Information.

Analysis. *Solvent Accessible Surface Area Analysis.* The average solvent accessible surface area of P-gp in solution with the detergent sodium cholate, and the TM helices of P-gp in series 4, was calculated using the method of Shrake and Rupley[26] and a probe of radius 0.14 nm. The contribution per atom to the solvent accessible surface of the overall system (P-gp and cholate) was determined and the total surface area estimated from the sum of the individual atomic contributions. The initial solvent accessible surface area was taken to be the solvent accessible surface area of P-gp in the crystallographic structure.

Cholate Binding to the Nucleotide Binding Domains. The number of cholate molecules lying between the nucleotide binding domains was estimated as the average number of cholate molecules found within 0.6 nm of the Walker A and Walker B motifs of both NBD1 and NBD2. This was calculated as the minimum distance between the center of mass of any atom in the cholate molecule and any of the atoms of the Walker A or B motif.

Cholesterol Binding to the Transmembrane Domains. The number of cholesterol molecules within 0.4 nm of the protein Cα was calculated over the five 60 ns simulations of series 4. This was calculated as the minimum distance between the center of mass of any atom in the cholesterol molecule and the center of mass of the Cα.

Table 1. A Comparison of the Inter-Residue Distances Obtained from Simulations of Membrane-Bound P-gp (series 4) and the 3G5U Crystal Structures with a Range of Cross-Linking Data Obtained Experimentally[a]

cross-linked residues (mouse P-gp numbering)			distance between Cα's (nm)							
			3G5U		MD simulations					
TM helices	1	2	A	B	min	max	average (5 × 60 ns)	calcd. from MTS spacer	temp (K)	ref
2 and 11	113	951	1.1	1.1	0.9	1.5	1.1 ± 0.1	1.2–2.0[b]	310	42
2 and 11	113	952	1.3	1.3	1.0	1.6	1.2 ± 0.1	1.2–2.0[b]	310	42
2 and 11	113	953	1.1	1.1	0.7	1.5	1.1 ± 0.1	1.2–2.0[b]	310	42
2 and 11	114	951	0.8	0.8	0.7	1.5	1.0 ± 0.1	1.2–2.0[b]	310	42
2 and 11	114	952	1.0	1.1	1.0	1.7	1.3 ± 0.1	1.2–2.0[b]	310	42
2 and 11	114	953	1.0	1.0	0.7	1.6	1.2 ± 0.2	1.2–2.0[b]	310	42
2 and 11	115	951	1.1	1.1	1.0	1.8	1.3 ± 0.1	1.2–2.0[b]	310	42
2 and 11	115	952	1.4	1.4	1.3	2.0	1.6 ± 0.1	1.2–2.0[b]	310	42
2 and 11	115	953	1.4	1.4	1.1	1.9	1.5 ± 0.2	1.2–2.0[b]	310	42
2 and 11	129	931	0.7	0.7	0.4	0.9	0.6 ± 0.1	1.2–2.0[b]	310	37
2 and 11	129	935	0.5	0.5	0.5	1.2	0.7 ± 0.2	1.2–2.0[b]	277, 295, 310	37
2 and 11	133	931	0.5	0.5	0.5	1.3	0.7 ± 0.2	1.2–2.0[b]	277, 295, 310	37
2 and 11	133	935	0.9	0.9	0.9	1.9	1.2 ± 0.2	1.2–2.0[b]	310	37
3 and 9	171	816	2.7	2.6	2.6	5.2	3.9 ± 0.4	1.3[e]	273	36
4 and 10	218	864	3.0	3.1	2.9	4.4	3.3 ± 0.3	1.4–2.6[c]	310	43, 44
4 and 10	218	868	3.0	3.2	2.9	4.4	3.5 ± 0.2	1.4–2.6[c]	310	43, 44
4 and 12	223	989	2.3	2.3	2.0	3.2	2.5 ± 0.2	0.9[d]	310	38
4 and 12	227	989	2.1	2.1	1.8	2.9	2.2 ± 0.2	0.9[d]	294, 310	38
4 and 12	228	989	2.0	2.0	1.5	2.8	2.2 ± 0.2	0.9[d]	294, 310	38
4 and 12	229	989	1.7	1.6	1.2	2.5	1.7 ± 0.2	0.9[d]	277, 294, 310	38
4 and 12	231	989	2.0	2.0	1.5	2.8	2.0 ± 0.2	0.9[d]	277, 294, 310	38
4 and 12	232	989	1.8	1.8	1.2	2.6	1.7 ± 0.3	0.9[d]	277, 294, 310	38
4 and 12	291	989	1.6	1.6	0.8	1.9	1.2 ± 0.2	0.9[d]	294, 310	38
4 and 12	295	989	1.3	1.4	0.6	2.1	1.2 ± 0.3	0.9[d]	277, 294, 310	38
5 and 8	302	864	2.9	3.0	2.4	3.7	3.0 ± 0.2	1.7–2.6[c]	310	43, 44
5 and 8	302	868	2.8	3.0	2.2	3.8	2.9 ± 0.3	1.7–2.6[c]	310	43, 44
5 and 8	313	749	1.3	1.3	1.2	2.0	1.6 ± 0.2	1.2–2.0[b]	310	42
5 and 8	313	750	1.3	1.3	1.2	1.9	1.5 ± 0.1	1.2–2.0[b]	310	42
5 and 8	313	751	1.0	1.0	0.9	1.5	1.2 ± 0.1	1.2–2.0[b]	310	42
5 and 8	314	749	1.3	1.4	1.2	2.1	1.6 ± 0.1	1.2–2.0[b]	310	42
5 and 8	314	750	1.2	1.3	1.1	2.0	1.5 ± 0.2	1.2–2.0[b]	310	42
5 and 8	314	751	0.9	1.0	0.9	1.7	1.2 ± 0.1	1.2–2.0[b]	310	42
5 and 8	315	749	1.1	1.1	0.9	2.1	1.4 ± 0.2	1.2–2.0[b]	310	42
5 and 8	315	750	1.0	1.1	1.0	2.1	1.3 ± 0.2	1.2–2.0[b]	310	42
5 and 8	315	751	0.9	0.9	0.7	1.9	1.1 ± 0.2	1.2–2.0[b]	310	42
5 and 11	302	941	3.1	3.2	2.3	3.7	3.0 ± 0.3	1.7–2.6[c]	310	43, 44
5 and 12	302	978	1.9	1.9	1.5	2.7	2.0 ± 0.2	1.7–2.6[c]	310	43, 44
5 and 12	302	980	2.3	2.2	1.0	3.1	1.9 ± 0.5	1.7–2.6[c]	310	43, 44
6 and 10	328	852	2.4	2.4	2.1	3.6	2.7 ± 0.3	1.2–2.0[b]	277	44
6 and 10	335	864	2.3	2.3	1.9	3.6	2.6 ± 0.2	1.7–2.6[c]	310	43, 44
6 and 10	335	868	2.3	2.3	2.0	3.6	2.6 ± 0.3	1.7–2.6[c]	310	43, 44
6 and 10	346	870	2.3	2.3	2.1	3.8	3.1 ± 0.2	0.9[d]	294, 310	38
6 and 10	346	871	2.0	2.0	1.8	3.5	2.7 ± 0.3	0.9[d]	310	38
6 and 10	346	872	1.8	1.8	1.6	3.7	2.7 ± 0.3	0.9[d]	294, 310	38
6 and 11	335	938	2.4	2.4	2.1	3.3	2.6 ± 0.3	2.6[c]	310	43, 44
6 and 11	335	941	2.3	2.3	2.0	3.1	2.5 ± 0.2	2.3–2.6[c]	310	43, 44
6 and 11	346	935	3.3	3.2	2.1	3.6	3.0 ± 0.2	0.9[d]	294, 310	38
6 and 12	335	978	1.5	1.4	0.9	2.3	1.6 ± 0.3	2.0–2.6[c]	310	43, 44
6 and 12	335	981	1.6	1.5	1.0	2.8	1.7 ± 0.5	2.3–2.6[c]	310	43, 44
6 and 12	346	989	1.5	1.3	1.3	3.0	2.3 ± 0.3	0.9[d]	310	45
6 and 12	339	982	1.5	1.3	0.6	2.4	1.4 ± 0.4	0.9[d]	310	45
6 and 12	342	985	1.5	1.3	0.9	2.8	1.9 ± 0.4	0.9[d]	310	45

[a]***NOTE:*** The distance between Cα atoms has been estimated as follows: 2 × (length of cysteine side chain + distance to spacer) + length of spacer. [b]Precise reducing agent not provided. Reducing agents used in this study were N,N′-*o*-phenylenedimaleimide, N,N′-*p*-phenylenedimaleimide, and 1,6-bismaleimidohexane. The spacer span lengths range from ∼0.3 to ∼1.0 nm. [c]Cross-linking was attempted with each of the following spacers: M2M (approximate cross-linking span: 0.52 nm), M3M (∼0.65 nm), M4M (∼0.78 nm), M5M (∼0.91 nm), M6M (∼1.04 nm), M8M (∼1.30 nm), M11M (∼1.69 nm), M14M (∼2.08 nm), and M17M (∼2.47 nm). The distance range given is calculated from the span of the largest and smallest

dx.doi.org/10.1021/ct300254y | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

Table 1. continued

spacer cross-linked. [d]Reducing agent is copper phenanthroline. Distance between the S atoms is ∼0.37 nm. [e]Spacer used was M1M, which has a span of ∼0.4 nm.

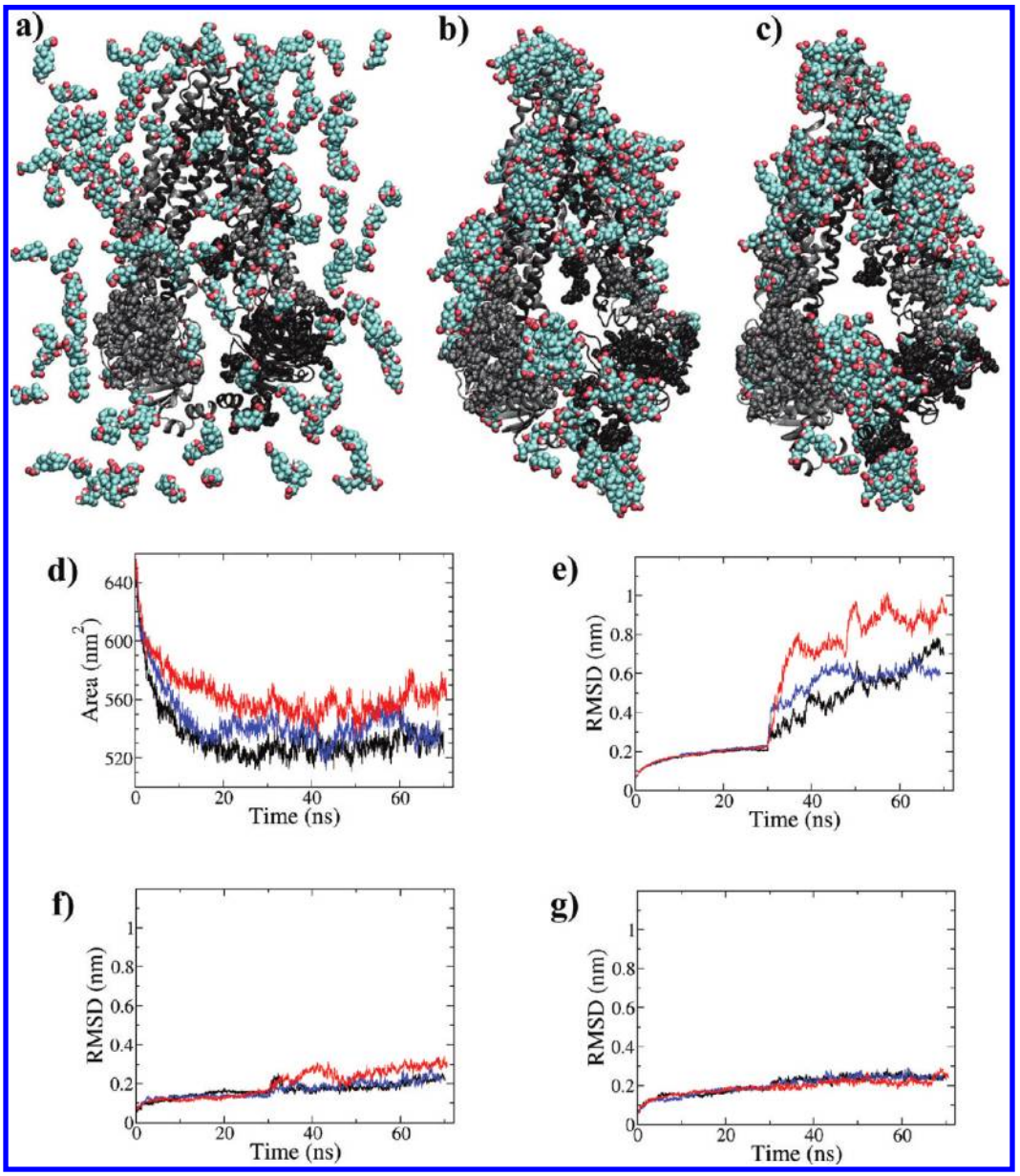

Figure 1. The binding of cholate to P-glycoprotein (P-gp). (a) The starting configuration (PDB 3G5U-a), (b) the configuration after 30 ns of MD simulation with position restraints, and (c) after 30 ns of unrestrained MD. The N- and C-terminal half-transporters of P-gp are shown in silver and dark gray, respectively. Residues involved in crystal contacts are shown in a space-filling representation. The remainder of the structure is shown in a ribbon representation. Cholate is shown in cyan using a CPK representation. (d) The total solvent accessible surface area of P-gp as a function of time. (e−g) The Cα RMSD as a function of time for (e) the entire molecule, (f) the RecA-like subdomain of NBD1, and (g) the RecA-like subdomain of NBD2. The individual runs are shown in different colors.

Separation of the Nucleotide Binding Domains. The distance between the ATP binding sites in the nucleotide binding domains was measured as the distance between the centers of mass of the Walker B motif from one NBD and the center of mass of the signature motif from the opposite NBD.

Inter-Residue Distances. The distance between specific pairs of residues was determined in order to compare the simulations to a range of cross-linking studies. The distance between two residues was taken as the distance between the centers of the corresponding Cα atoms. The distance was calculated for each of the configurations stored. The maximum, minimum, and average values obtained for the 300 ns of the combined trajectory are reported in Table 1. The equivalent distances for the two P-gp molecules in the asymmetric unit of the 3G5U crystal structure, denoted 3G5U A and B, are also given in Table 1 for comparison.

Root Mean Squared Deviation (RMSD). To compare the relative difference between trajectories or clusters, the RMSD was calculated using the method of Maiorov and Crippin[27] after

dx.doi.org/10.1021/ct300254y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

first fitting the trajectory frames to a reference structure or domain.

Root Mean Squared Fluctuation (RMSF). In order to determine whether each simulation had converged, the RMSF was determined by calculating the standard deviation in the positional coordinates of the Cα of each stable secondary structural element of P-gp over a 5.0 ns window after fitting each trajectory to the crystal structure.

Cluster Analysis. To determine the relative populations of specific conformations the trajectories were clustered using the method of Daura et al.[28,29] In this work, two conformations were considered neighbors if the backbone RMSD between the two conformations was <0.4 nm. Cluster analysis was performed on the five individual trajectories of series 4 and also the combined trajectory of series 4. To eliminate the possibility of the clusters being biased by the fact that all simulations were initiated from the same configuration, the cluster analysis was repeated discarding the first 0, 5, 10, and 20 ns of each trajectory. The coordinates of the membrane-embedded central structure of the dominant/most populated cluster of the combined trajectory are available for download as part of the Supporting Information and from the Automated Topology Builder Web site.[22]

RESULTS AND DISCUSSION

P-gp in the Crystallographic Solution. P-gp was crystallized in the presence of the ionic detergent NaCholate at pH 7.5, in a magnesium-free solution, and in the absence of nucleotides.[1] The role of detergent when determining the structures of integral membrane proteins is a matter of continuing debate. Detergents are needed to maintain the protein in a soluble form for crystallization but also can affect the structural properties of the protein directly.[12] Furthermore, as the detergent generally binds to the protein nonspecifically, crystals formed in the presence of detergent are often less well ordered than detergent-free crystals.[12] In the crystal structure of P-gp, no individual detergent molecules could be resolved. In fact, the protein itself could only be resolved to a resolution of 3.8 Å.[1] To investigate the role of cholate in determining the structure observed crystallographically, three 30 ns simulations were performed in which cholate was allowed to bind spontaneously to the crystal conformation of P-gp while it was positionally restrained. Then, 40 ns of unrestrained MD was performed. In all simulations, cholate initially formed micelles composed of 3 or 4 molecules, corresponding to the experimental aggregation number for cholate.[30] These micelles rapidly bound to the P-gp structure, forming larger aggregates on the surface. In each of the simulations, all cholate molecules had bound to the protein within 10 ns. As shown in Figure 1, the anionic cholate molecules heavily coated the TMDs and ICLs. Cholate was also associated with the NBDs. It coated the NBD:ICL interface and formed a large aggregate that bridged between the two NBDs, interacting with positively charged residues in the NBDs and ICLs. Cholate entered the region between the NBDs within the first 2.5 ns and remained there throughout each of the simulations. On average, 2.4 ± 1.2 molecules of cholate bound between the NBDs, within 0.6 nm of the Walker A and Walker B motifs, maintaining the separation of the two NBDs. Figure 1 shows cholate also aggregated in the gaps between the TMDs blocking the membrane-embedded portals through which calcein-AM[31] and other hydrophobic compounds are believed to enter the translocation pore. The presence of cholate in these portals may explain why verapamil and other compounds could not be soaked into the P-gp crystal.[1]

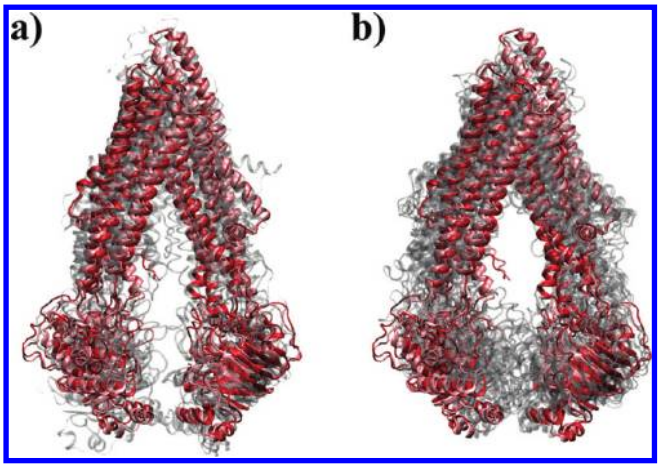

Figure 2. Superposition of the two P-gp crystal structures in the 3G5U asymmetric unit (red and pink) with the conformation of P-gp from (a) each of the three simulations after 30 ns of unrestrained MD in the presence of cholate and (b) the five predominant conformations from series 4 of the membrane embedded simulations. The simulated conformations of P-gp are shown in silver.

In Figure 1, residues that were involved in crystal contacts, defined as a residue in which any atom lay within 0.5 nm of an atom in another copy of P-gp in the same or another unit cell, are shown in a space filling representation. As evident from Figure 1, cholate did not adhere to the regions of P-gp that are involved in crystal contacts. This suggests that the specific crystal form obtained by Aller et al. may have been heavily influenced by the presence of cholate. The solvent accessible surface area (SASA) of P-gp was 653 nm^2 at the start of the simulation. Due to the adsorption of cholate, this rapidly decreased, reaching a minimum of between 530 and 570 nm^2 depending on the simulation. This minimum was reached within the initial 30 ns of simulation, during which time the conformation of the protein was positionally restrained. The SASA then gradually increased as the cholate clustered more tightly around the TMDs and at the interface between the ICLs and NBDs. As can be seen in Figure 1d, which shows the SASA as a function of time for each of the three simulations, the final SASA after 70 ns is similar (541 ± 19 nm^2) in each case. The Cα RMSD was calculated across all atoms of P-gp in the presence of cholate for each of the three simulations. After 40 ns of unrestrained simulations, the maximum Cα RMSD measured with respect to molecule 3G5U-a of the asymmetric unit varied between 0.59 and 0.89 nm across all three simulations, as shown in Figure 1e. In comparison, the Cα RMSD between the two molecules (3G5U-a and 3G5U-b) of the 3G5U asymmetric unit was 0.33 nm. Figure 2a shows the superposition of the 3G5U-a and 3G5U-b (red and pink, respectively) and the three P-gp conformations after 30 ns of unrestrained simulation (silver). Visually, it can be seen that the differences between the structures arise primarily from variations in the position of the ICLs and NBDs. In particular, small changes in the relative orientation of the two NBDs heavily influence the Cα RMSD of the entire protein. The NBDs of all ABC transporters, including P-gp, consist of two subdomains: the RecA-like subdomain, containing the Walker A and Walker B motifs, and the helical subdomain, which

dx.doi.org/10.1021/ct300254y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

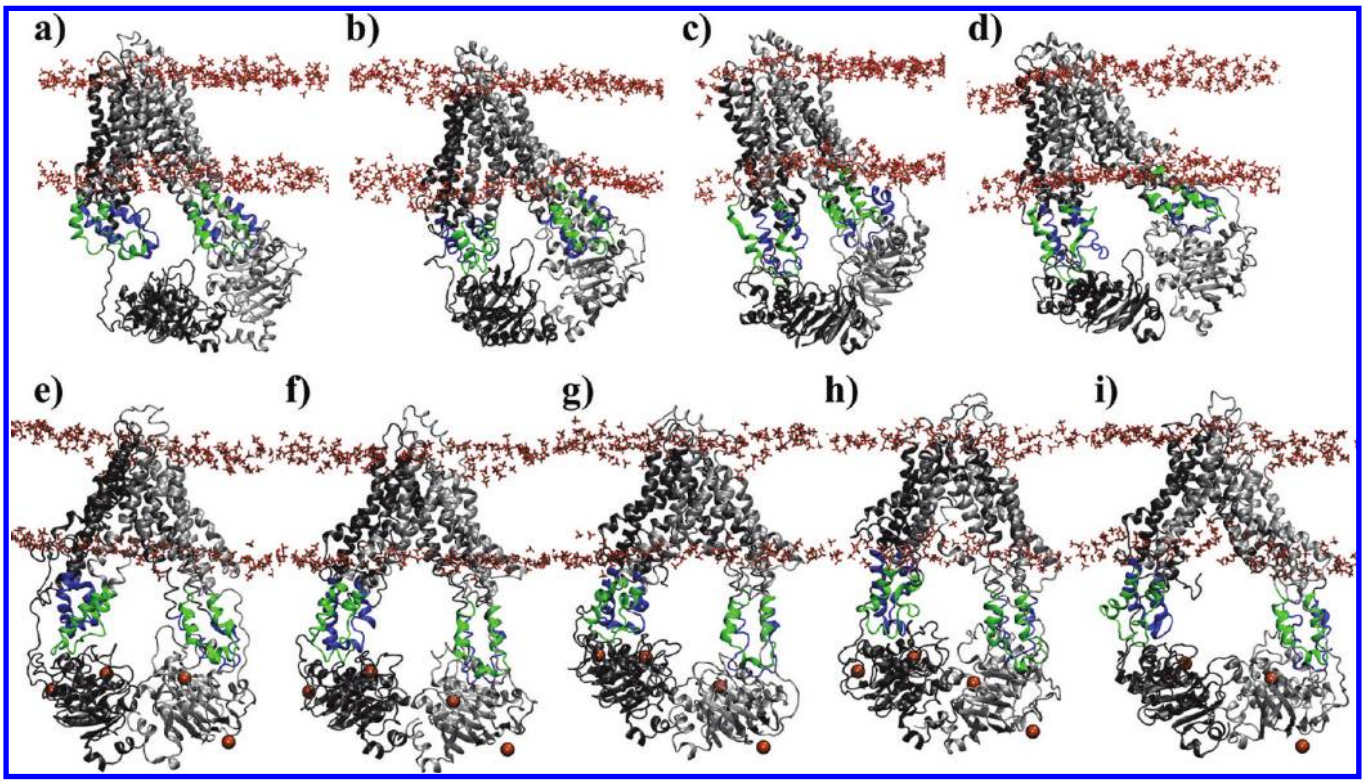

Figure 3. The effect of pH and the presence of $MgCl_2$ on the structural stability of membrane bound P-gp. (a) In the absence of $MgCl_2$ and with all histidines singly protonated. (b–d) In the absence of $MgCl_2$ and with His149 singly protonated, taken from the final conformation of each of the three simulations. (e–i) In the presence of 1.5 mM $MgCl_2$ (His583, 1228, and 149 protonated), taken from the dominant conformation in each of the five simulations. The N- and C-terminal intracellular loops (ICLs) are shown as green and blue ribbons, respectively. The N- and C-terminal half-transporters of P-gp are shown in silver and dark gray, respectively. The POPC headgroups are shown in CPK. The four Mg^{2+}'s that bind to the NBDs are shown in orange.

contains the consensus signature motif for ABC transporters. In the presence of ATP, the two subdomains from each NBD come together to cooperatively bind one molecule of ATP between the Walker A and B motifs of one NBD and the signature motif of the opposite NBD. The binding of ATP constrains the relative orientation of both the subdomains and the NBDs themselves. The RecA-like subdomains of both NBD1 and NBD2 were very stable, having a Cα RMSD of just 0.23 ± 0.04 nm and 0.23 ± 0.01 nm, respectively, after 40 ns of unrestrained simulation (Figure 1f and g). In all of the unrestrained simulations of P-gp coated with cholate, the distance between the NBDs, as measured from the center of mass of the Walker B motif of one NBD and the signature motif of the opposite NBD, fluctuated between 2.5 and 5.2 nm before converging to around 3.8 nm after 40 ns of unrestrained simulation. A plot of the distance between the center of mass of the Walker B motifs and the opposing signature motifs is provided as Supporting Information (Figure S1). In the 3G5U-a crystal structure, the distance between the Walker B motif of NBD2 and the signature motif of NBD1 was 3.3 nm, while the distance between the Walker B motif of NBD1 and the NBD2 signature motif was 3.7 nm.

P-gp in a Membrane Environment. Experimentally, functionally active P-gp can be obtained by reconstituting purified P-gp into liposomes. Such a membrane environment is very different from the conditions under which P-gp was crystallized. As mentioned earlier, P-gp was crystallized at pH 7.5, whereas transport and ATPase assays are normally performed close to pH 6.8.[18,32] The presence of 1.5 mM $MgCl_2$ in reconstitution experiments is also required for P-gp to hydrolyze ATP.

To determine whether the structure of P-gp obtained crystallographically was stable within a membrane environment, the protein was embedded in a cholesterol-enriched POPC bilayer. Phosphatidylcholines (PC) such as POPC and DPPC are a major component of eukaryotic cell membranes, and P-gp is functionally active when reconstituted into purified POPC or DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine) liposomes.[33] Experimentally, the addition of cholesterol to these liposomes improves the efficacy of drug-mediated ATPase activity in P-gp.[33] Four series of simulations were performed. In series 1, all histidines were singly protonated. In series 2, His583 and His1228 were doubly protonated (charge +1). His149 was singly protonated (neutral). In series 3, His583 and His1228 were doubly protonated, His149 was singly protonated, and the equivalent of 1.5 mM $MgCl_2$ was added to the system. In series 4, His583, His1228, and His149 were doubly protonated and again the system contained 1.5 mM $MgCl_2$. In each case, the system was equilibrated for 10 ns, during which time the conformation of P-gp was restrained. Then, the restraints were removed, and either three (series 1, 2, and 3) or five (series 4) independent simulations were performed using different starting velocities.

In both series 1 and series 2, the structure of the protein became heavily distorted within 30 ns. Most importantly, the protein progressively tilted with respect to the plane of the membrane during the simulation, with the intracellular loops of TM wing 1 becoming embedded in the membrane. In the case of series 1, NBD2 became decoupled from the ICLs, as shown

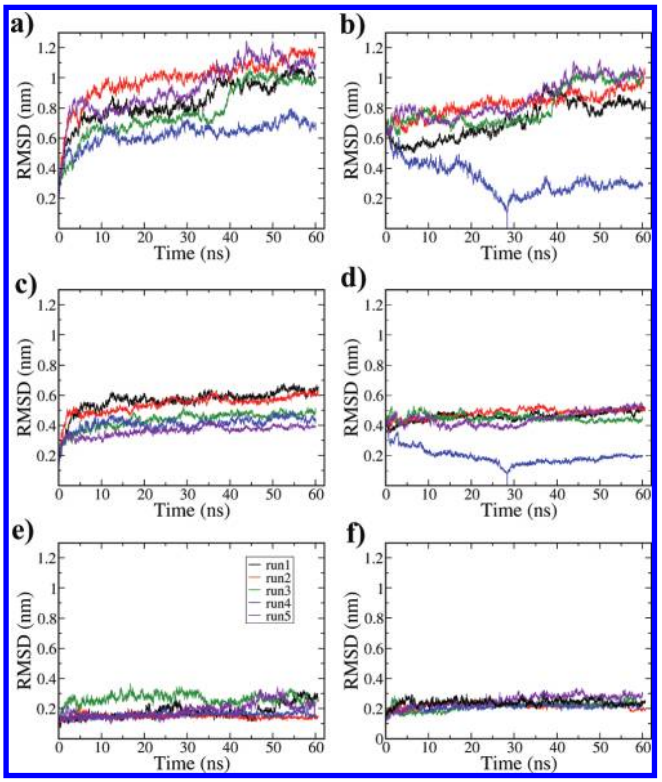

Figure 4. Cα RMSD of membrane-embedded P-gp as a function of time. The RMSD calculated for all Cα in the structure, with respect to (a) the starting crystal structure and (b) the most populated cluster of the combined trajectories. The Cα RMSD of the major elements of the secondary structure in the membrane-embedded regions of TM helix 1 (residue 45−82), 2 (112−142), 4 (212−232), 5 (282−312), 7 (709−734), 8 (749−774), 10 (854−879), and 11 (934−959), calculated with respect to (c) the starting crystal structure and (d) the most populated cluster of the combined trajectories. The Cα RMSD of the RecA-like subdomain of (e) NBD1 and (f) NBD2 calculated with respect to the 3G5U-a crystal structure.

in Figure 3a. As protonation of the catalytic histidines is believed to be required for activity, and NBD2 became decoupled from the rest of the protein, these simulations will not be discussed further. In the case of series 2, the Cα RMSD with respect to the initial structure reached 0.6 nm within 2 ns. In all three simulations, an interface formed between NBD1 and NBD2. In two of the three simulations, this was associated with the disruption of either the interface between ICL1 and NBD1 or the interface between ICL4 and NBD2. The extent of the tilt (30° to 45°) is evident in Figure 3b−d, which shows the final structure after 30 ns of unrestrained simulation for each of the three simulations performed in series 2. The final structures are shown as the proteins continued to diverge throughout the simulations. Note, the tilted structures shown in Figure 3b−d are very similar to the heavily distorted structures of the P-gp homologue, Sav1866, obtained by Oliveira et al. under comparable conditions and without Mg^{2+}.[21]

The simulations in series 3 were identical to those in series 2 except that 1.5 mM $MgCl_2$ was added to the system. Mg^{2+} is known to bind to the Walker B motifs and is required for P-gp to hydrolyze ATP. In each of the three independent simulations, four Mg^{2+}'s bound spontaneously to distinct sites in the NBDs within 6 ns. A single Mg^{2+} bound to the magnesium-binding site within the Walker B motif of each NBD (residues Asp551 and Glu552 in NBD1 and Asp1196 and Glu1197 in NBD2). The binding of Mg^{2+} to this location has been observed in crystal structures of other ABC transporters.[6,34] A second Mg^{2+} bound to a previously uncharacterized site on the exterior surface of each NBD, in the vicinity of Glu389 in NBD1 and of Glu1096 in NBD2. The binding of Mg^{2+} effectively increased the net charge on each NBD. The binding of Mg^{2+} significantly increased the structural stability of the membrane embedded protein but did not completely eliminate the tilting and deformation of the protein (results not shown).

His149 was predicted to have a pK_a of 7.1 and might be either singly or doubly protonated at pH 6.8. In series 4,

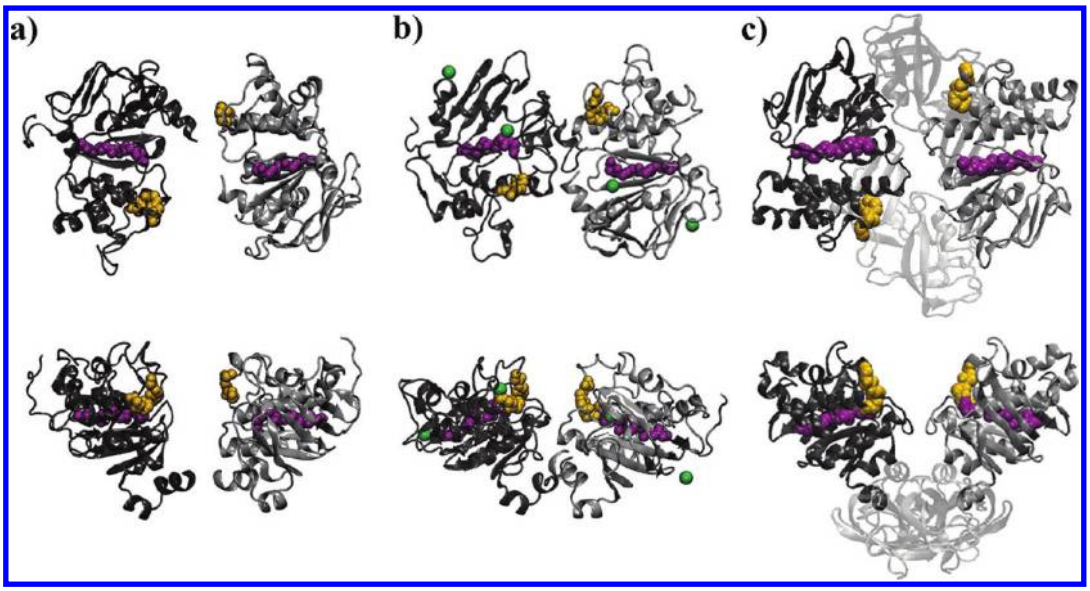

Figure 5. Comparison of the P-glycoprotein NBD and Walker B orientation. Top view (upper panel) and side view (lower panel) of (a) the P-gp crystal structure, (b) the stable membrane-embedded P-gp conformation with Mg^{2+}, and (c) the crystal structure of the MalK open conformation (the C-terminal regulatory domain is transparent). The Walker B motif is shown in purple. The signature loops are shown in gold, and Mg^{2+} is shown in green spacefill. The N- and C-terminal NBDs are shown in silver and dark gray ribbons, respectively.

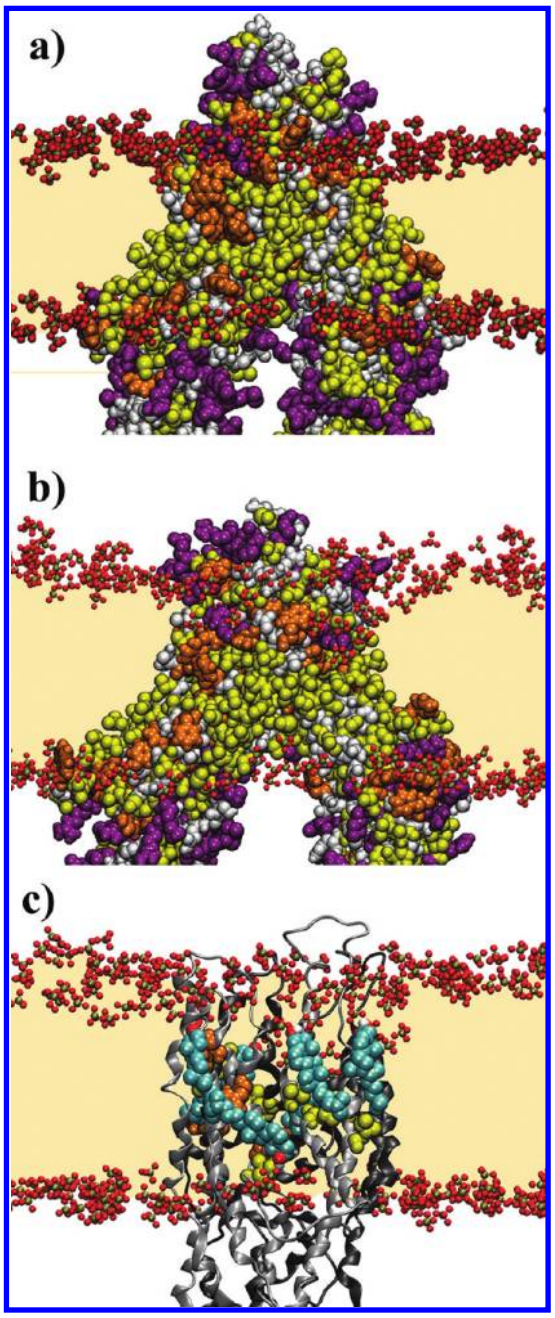

Figure 6. The interaction of P-gp with the cholesterol-enriched POPC bilayer. A front view of (a) the 3G5U-a crystal structure and (b) the central structure of the combined trajectories. (c) A side view of run 4 of series 4 after 60 ns showing the location of cholesterol (cyan spacefill). Protein residues within 0.4 nm of a cholesterol molecule are shown in spacefill. The N- and C-terminal TMDs of P-gp are silver and dark gray ribbons, respectively. Charged and polar residues are shaded purple and white, respectively. Aromatic and nonpolar residues are shaded orange and yellow, respectively. The POPC headgroups are shown as small CPK spheres, and the tail region is colored beige.

His583, His1228, and His149 were doubly protonated, and 1.5 mM $MgCl_2$ was present in the system. In contrast to the simulations of membrane bound P-gp performed in the absence of Mg^{2+}, or with His149 singly protonated, the structure of the protein did not tilt in the membrane. In addition, the states sampled in all five simulations were similar. Figure 3e−i show the central structure of the dominant conformation from each of the five simulations as determined by cluster analysis using the method of Daura et al.[28,29] Note the conformation shown in Figure 3e was also the most common conformation obtained after combining all five trajectories. This was irrespective of whether all of the configurations were included in the cluster analysis or whether the first 5, 10, or 20 ns of each of the trajectories was treated as equilibration and removed from the analysis. As can been seen in Figure 3e−i, the conformations that were predominantly sampled during all five simulations were very similar to the X-ray structure of P-gp proposed by Chang and co-workers. The primary difference between the structures sampled in series 4 and the crystal structure is the inward movement of the two NBDs, leading to the formation of direct contacts between the NBDs. At the same time, the angle between the TMDs increases, resulting in an outward motion of the ICLs. This can be seen from Figure 2b, which shows the superposition of the two molecules of the 3G5U crystal structure and the dominant conformation from each of the five simulations of series 4.

Table 2. Arginine Mutations That Alter the Rate of Drug Transport[41]

inhibiting mutations		stimulating mutations	
residue	TM helix	residue	TM helix
58	1	133	2
127	2	195	3
190	3	228	4
223	4	295	5
300	5	298	5
712	7	302	5
771	8	332	6
829	9	335	6
866	10	339	6
940	11	762	8
984	12	766	8
		864	10
		868	10
		869	10
		938	11
		941	11
		942	11
		978	12
		979	12
		982	12
		986	12

To characterize the variation between the simulations and the structure proposed by Aller et al., the RSMD along each of the five trajectories of series 4 was calculated with respect to the starting crystal structure and with respect to the central structure of the combined trajectories. The RMSD was calculated on the basis of (i) all Cα atoms, (ii) considering only those amino acids involved in the major elements of the secondary structure in the membrane embedded regions, and (iii) only residues within the RecA-like subdomain of the NBDs. Before the variation in the RSMD is discussed, it should be noted that P-gp is a large multidomain protein containing 1280 amino acids. In addition, P-gp is highly elongated. In large elongated proteins such as P-gp, the RMSD can be very sensitive to changes in the relative positions of the different domains as well as to fluctuations in exposed loops. These can alter the overall fit of one molecule relative to the other, amplifying the effect of changes at the extremities of the molecule.[27] P-gp is also highly flexible. EM studies in the

dx.doi.org/10.1021/ct300254y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

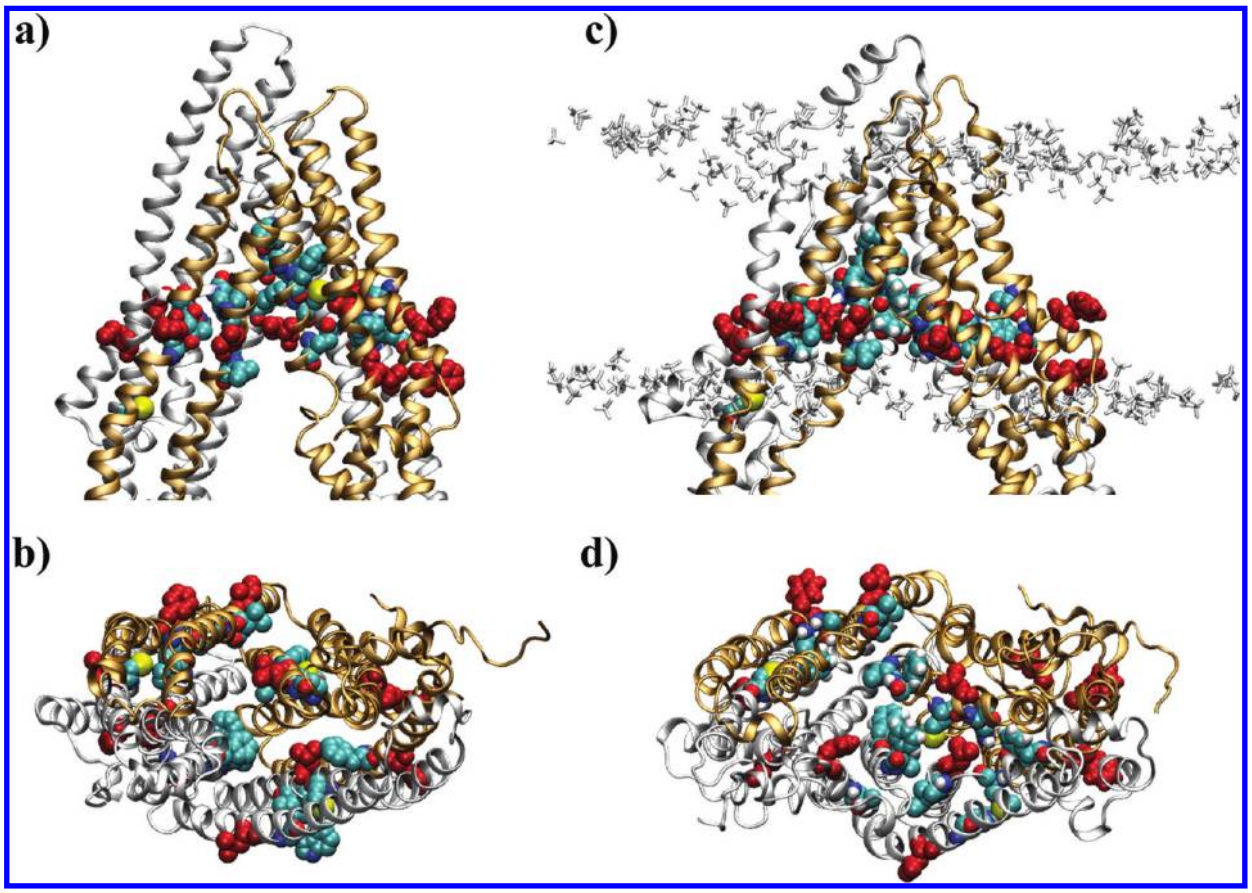

Figure 7. Residues implicated in drug transport mapped onto the 3G5U-a crystal structure and the central conformation from the dominant stable cluster of the combined trajectories. The front view and cytoplasmic view through the membrane of (a and b) 3G5U-a and (c and d) the central conformation from the dominant cluster showing the residues experimentally identified as increasing (CPK spacefill) or decreasing (red spacefill) the drug transport rate. TMD1 is silver, and TMD2 is bronze.

presence and absence of different nucleotides have shown that the P-gp transport cycle is associated with large changes in the relative positions of the NBDs and the TMDs.[35] The sensitivity of the RMSD to changes in the relative positions on the NBDs and the TMDs is reflected in the fact that the Cα RMSD between the two molecules of P-gp in the asymmetric unit of 3G5U is 0.33 nm. Figure 4a shows the Cα RMSD of P-gp with respect to the starting crystal structure (3G5U-a) for each of the five simulations of series 4. As can be seen from Figure 4a, the RMSD rose rapidly during the first 10 ns, reaching values of between 0.6 and 1.0 nm. The RSMD continued to increase throughout the remaining 50 ns, albeit slowly, reaching values of between 0.7 and 1.2 nm after 60 ns, depending on the simulation. That the variation in the conformations of the flexible loops makes a major contribution to the overall RMSD is evident in Figure 4c, which shows the Cα RMSD with respect to the starting crystal structure considering only those amino acids involved in the major elements of secondary structure in the membrane-embedded regions of TM helices 1 (residue 45−82), 2 (112−142), 4 (212−232), 5 (282−312), 7 (709−734), 8 (749−774), 10 (854−879), and 11 (934−959). In this case, the RMSD begins to plateau within the first 5 ns, rising only slightly during the remainder of the simulation. The RMSD after 60 ns varied between 0.4 and 0.7 nm. Considering the individual domains, the RMSD is even smaller: Figure 4e and f show the Cα RMSD with respect to the starting crystal structure for the RecA-like subdomains of NBD1 (residues 387 to 480 and 547 to 622, Figure 4e) and NBD2 (residues 1029 to 1114 and 1194 to 1269, Figure 4f). As can been seen, both domains are stable with the RMSD fluctuating between 0.15 and 0.30 nm throughout the five 60 ns simulations.

The Cα RMSD was also calculated with respect to the central structure of the most populated cluster. The most populated cluster varied from 20 to 28% of all the configurations in the combined trajectories, depending on whether the first 5, 10, or 20 ns of each of the trajectories was removed from the analysis, clearly indicating that simulations were converging to a structure different from the initial crystal structure. This structure, shown in Figure 3e, corresponds to the structure after 28 ns of simulation from run 4 in series 4. Figure 4b shows the Cα RMSD of P-gp with respect to this central structure for each of the five simulations. As can be seen in Figure 4b, the RMSD is in general less than that observed in Figure 4a, indicating that the conformations sampled during the simulations were closer to the structure shown in Figure 3e than to the starting crystal structure. In particular, after 60 ns of simulation, the RMSD in Figure 4b is less than that in Figure 4a, varying between 0.3 and 1.1 nm. Note the RMSD drops to zero at 28 ns for run 4 as this structure corresponds to the central structure shown in Figure 3e. The same effect is more pronounced when only those amino acids involved in the major elements of the secondary structure in the membrane-embedded regions of TM helices 1, 2, 4, 5, 7, 8, 10, and 11 are considered. In this case, the RMSD remains almost constant throughout the simulations and is on average lower by approximately 0.1 nm. The final RMSD after 60 ns varied

between 0.2 and 0.5 nm. That the structures obtained in series 4 were stable is also reflected in the root mean squared positional fluctuations (RMSF), which decreased markedly after the initial 10 ns of simulation. The RMSF of each Cα was calculated for 5 ns windows throughout the simulations. The average Cα atom RMSF for residues found in elements of the secondary structure decreased from approximately 0.4 at the beginning of the simulation (0−5 ns) to 0.1 nm toward the end of the simulation (50−55 ns). Plots of the fluctuations as a function of the residue number for periods 0−5 ns and 50−55 ns for each of the five simulations are provided as Supporting Information (Figure S2).

As noted above, Figure 2b shows the superposition of each of the central structures from the dominant conformations of the five simulations (gray) with 3G5U-a (red) and 3G5U-b (pink). The main difference between the structures sampled during the simulations and the starting crystal structure arises from the orientation and relative positions of the NBDs. In 3G5U-a, the vectors projecting along the Walker B motifs as measured from the Cα of Ile547 to Glu552 and Leu1192 to Glu1197 are almost parallel, intersecting at an angle of 165°. When averaged over the predominant conformation observed in each of the five simulations, the angle of intersection of the Walker B vectors was 121 ± 16°. The angle of intersection of the Walker B motifs, calculated with respect to the central structure of the most populated cluster of the combined trajectories (Figure 3e), was 127°. The change in the relative orientation of the NBDs was also associated with a decrease in the distance between the two NBDs. The distance between the centers of mass of the NBD1 Walker B motif and NBD2 signature motif in 3G5U-a was 3.9 nm. In the simulations, this distance varied between 2.3 and 4.4 nm. When averaged over the course of the five simulations, the value was 3.2 ± 0.7 nm, while the average value obtained by considering the central structure of the five most populated clusters was 3.1 ± 0.5 nm. The distance between the NBD2 Walker B motif and NBD1 signature motif in the 3G5U-a crystal structure was 3.6 nm. In the simulations, this distance varied between 2.2 and 4.0 nm. Averaged over the course of the five simulations, the value was 3.0 ± 0.3 nm. When averaged over the central structure of the five most populated clusters, the value was also 3.0 ± 0.3 nm. While the variation between the simulations might seem large, P-gp is highly flexible and undergoes significant conformational changes during the transport cycle. In fact, the position and orientation of the NBDs varies significantly between the two molecules of P-gp in the asymmetric unit of 3G5U.[1] Whereas the distances between the centers of mass of the Walker B and signature motifs in 3G5U-a are 3.9 and 3.6 nm, the corresponding distances in 3G5U-b are 4.5 and 4.0 nm. This highlights the importance of packing forces in determining the relative positions of the NBDs in the crystal.

The reorientation of the NBDs and the formation of a NBD:NBD interface observed in the simulations is reminiscent of the open state of the MalK NBDs (PDBid 1Q1E). In the MalK structure, the angle of intersection of the Walker B vectors is 138°, close to that of the most populated cluster of the combined trajectories. Figure 5 shows the orientation of the NBDs and Walker B motifs in the P-gp crystal structure (Figure 5a), the central structure of the most populated cluster of the combined trajectories (Figure 5b), and in the MalK open state dimer (Figure 5c). Note that in the case of MalK, the NBDs have an additional C-terminal regulatory domain which brings the base of the NBDs into direct contact.[5] Despite not having an equivalent regulatory domain, the interface between the NBDs in P-gp still occurs primarily in the C-terminal region mediated by Thr 559 and Glu560 of NBD1 and Ser1067, Ser1068, and Gly1069 of NBD2.

The reorientation of the NBDs is coupled to the splaying of the ICLs. To quantify the increase in the separation between the ICLs, the distance between the Cα atoms of pairs of residues in the opposing ICLs was calculated. In the simulations, the distance between residue pairs close to the NBDs increased by approximately 1.0 nm, while the distance between residues pairs in the membrane-proximal regions of the ICLs remained close to the 3G5U-a structure. As the Mg^{2+} binds to the NBDs and the catalytic histidines lie within the NBDs, only the protonation of His149, which lies in ICL1, directly affects the splaying of the ICLs. Although there was a change in the relative positions of the NBDs, Figure 4e and f demonstrate that the domains themselves were stable.

Comparison of the Simulations to the Experimental Data. To verify that the conformation of membrane-bound P-gp obtained is physiologically reasonable, the distances between Cα atoms of residue pairs in the TMDs and ICLs were compared to distances inferred from chemical cross-linking studies of the corresponding residues of nucleotide free human P-gp. In these cross-linking studies, specific pairs of residues in the TMDs and ICLs were mutated to cysteine. A series of chemical agents was then used to covalently link the two sulfur atoms. A spacer of a given length was considered compatible with the physiologically relevant conformation if P-gp retained the ability to hydrolyze ATP and, thus, transport substrates. Such cross-linking studies can be used to infer the minimum distance between the Cα atoms of the two residues that is compatible with ATPase and transport activity. Specifically, to form a cross-link, the maximum distance spanned by a given agent must be greater than the minimum distance between the two residues that is compatible with ATPase activity. In order to examine whether the simulations were compatible with the cross-linking data, the five 60 ns trajectories were concatenated and the minimum, maximum, and average distances between pairs of residues were determined as described in the Materials and Methods. The results are presented in Table 1 arranged on the basis of the order of the TM helices. Also given in Table 1 are the equivalent distances between pairs of residues in the 3G5U crystal structure for both molecules in the asymmetric unit. In total, 52 pairs of residues were examined experimentally (Table 1). For 30 of these pairs, the distance between the Cα atoms in the crystal structure 3G5U (columns 4 and 5 of Table 1) is less than or equal to the maximum distance allowed by the cross-linking agent (column 9). However, in 22 cases, the distance between the Cα atoms in the crystal structure is more than 0.1 nm greater than the maximum distance that can be spanned by the cross-linking agent. This suggests that the structure proposed by Aller et al. cannot fully explain the cross-linking data unless some allowance is made for thermal motion. Considering the average distance obtained from the simulations (column 8), the same 30 pairs of residues would be compatible with the cross-linking data. In the other 22 cases, the average distance observed in the simulations is once again more than 0.1 nm greater than the maximum distance that can be spanned by the cross-linking agent. However, during the simulations the structure of P-gp fluctuates, allowing pairs of residues to approach more closely. Considering the minimum distance sampled during the simulations (column 6), 37 pairs are compatible with the cross-linking data. In fact, in all cases, the

minimum distance observed in the simulations was between 0.0 and 1.3 nm shorter than the distance in the crystal structure. Thus, even for the 15 cases in which the minimum distance sampled in the simulations is still greater than the maximum distance allowed by the cross-linking agent, the simulations are more compatible with the cross-linking data than the starting crystal structures. This suggests that the simulations of membrane bound P-gp in which His583, His1228, and His149 are doubly protonated and two Mg^{2+} are bound to each NBD better represent the cross-linking data than does the 3G5U crystal structure. The highest correspondence to the cross-linking data was observed for pairs spanning TM helices 2 and 11 and TM helices 5 and 8. Residue pairs spanning TM helices 4 and 10 and TM helices 6 and 10 showed the largest deviation from the cross-linking data. For all 13 pairs of residues in TM helices 2 and 11, both the minimum and the average distances obtained in the simulations, and the distance observed in the crystal structure, were less than or equal to the maximum distance compatible with the presence of the cross-linking agent.

Several pairs deserve special note. In the case of residues 171 and 816, which are located in the ICL extensions of TM helices 3 and 9, respectively, neither the crystal structure nor the simulated data were consistent with the formation of a cross-link. Experimentally, 171 and 816 could only be cross-linked at 273 K.[36] In addition, cross-linking these residues resulted in a 3-fold increase in the basal rate of ATP hydrolysis.[36] Thus, the conformation adopted by the protein in the presence of this cross-link is not the same as the resting state of nucleotide free P-gp. In this case, the average distance sampled during the simulation (3.9 nm) is significantly larger than the distance observed in the crystal structure (2.6 nm). Two pairs of residues spanned TM helices 4 and 10. In neither case was either the crystal structure or the simulation compatible with the cross-link. Of the eight pairs spanning TM helices 4 and 12, none are satisfied by the crystal structure, while two pairs are satisfied by the minimum distance in the simulations. More importantly, the average distance is significantly lower than that found in the crystal structure in several cases, suggesting that the orientation of TM helix 4 adopted in the simulations is more appropriate than that in the crystal structure. Likewise, for eight of the 14 pairs linking TM helix 6 with TM helices 10, 11, and 12, the 3G5U crystal structure is not compatible with the formation of a cross-link, while only six cannot be accounted for in the simulations. In this case, configurations are sampled during the simulations in which the distance between the residues is up to 1.3 nm lower than that observed in the crystal structure. This suggests that the orientation of TM helix 6 in the crystal structure is inappropriate. Finally, we note that the majority of the experimental studies were performed at 310 K, while the simulations were performed at 300 K. Experimentally, the cross-linking efficiency was found to be highly temperature dependent.[37,38] Some cross-links only formed at low or high temperatures. This suggests that structural variations and/or thermal fluctuations are required for specific pairs of residues to be cross-linked.

The Effect of the Membrane on the Conformation of the TMDs. The crystal structure of P-gp was placed in the membrane such that the bilayer was positioned in the middle of the hydrophobic belt of the P-gp TMDs. Figure 6 shows the structure as initially embedded into the membrane (Figure 6a), the central structure of the combined trajectories (Figure 6b), and the positions of various cholesterol molecules after 60 ns of simulation for run 4 of series 4 (Figure 6c). From Figure 6a, it can be seen that when the crystal structure was initially embedded in the membrane, clusters of polar and charged residues were located within the span of the membrane, while groups of nonpolar residues were exposed to aqueous solution. In addition, several aromatic groups projected toward the center of the bilayer. During the simulations, the reorientation of the NBDs and the increase in distance between the ICLs was also associated with an increase in the interaction of the aromatic residues in the TMDs with the membrane−water interface and the interaction of nonpolar residues with the membrane. In the 3G5U-a crystal structure, the internal cavity of the TMD pore is large and accessible to water. In contrast, cysteine mutagenesis studies suggest that residues located near the center of the membrane in the pore-lining helices TM 6 and TM 12 are largely inaccessible to water-soluble probes.[39,40] In all five simulations, water was excluded from the TMD pore beyond the inner leaflet of the bilayer. In order to quantify the local environment of residues located within the span of the membrane, the solvent accessible surface area (SASA) of the membrane-embedded residues of each TM helix was calculated, taking into account the presence of the lipid bilayer. For residues in the 12 TM helices, the average SASA per residue throughout the simulations was less than or equal to the SASA of the corresponding residues in the membrane-embedded 3G5U-a crystal structure. Plots of the SASA for each of the membrane-embedded regions of the TMDs are provided as Supporting Information (Figures S4 and S5). Several polar residues from TM helices 3, 4, and 6 contribute to the surface of the proposed substrate uptake portal located between TM helix 3 (residues 172 to 180) and 4 (residues 233 to 242). The average SASA of these residues throughout the combined 300 ns remained the same as in the 3G5U-a crystal structure. The average SASA of residues near the extracellular lipid−water interface of TM helices 1, 2, 4, 5, 7, 9, and 11, which are in direct contact with the bilayer, also remained essentially the same as that of the 3G5U-a crystal structure. Between four and eight cholesterol molecules bound to the TMDs during the five 60 ns simulations of series 4. Figure 6c shows that cholesterol binds nonspecifically to the aromatic residues Phe and Trp and aliphatic residues Val, Ile, and Leu facing the lipid milieu. As can be seen in Figure 6c, the long axes of cholesterol molecules are aligned with the tilt of the transmembrane helices, rather than perpendicular to the membrane normal.

Arginine mutagenesis studies on the TMDs and ICLs of P-gp have identified clusters of residues that when mutated, either simulate or inhibit the ability of P-gp to transport drugs.[41] These residues are listed in Table 2. As can be seen from Figure 7a and b, these residues are dispersed throughout the membrane embedded regions in the 3G5U-a structure. Figure 7c and d show the spatial distribution of these residues for the dominant stable cluster from the combined trajectories of the membrane-bound simulations of P-gp. In contrast to the crystal structure, the spatial distribution of these residues in the MD simulations is more restricted. The residues form a distinct band contained within the intracellular lipid leaflet.

In summary, the conformation in which an integral membrane protein crystallizes can be critically dependent on the experimental conditions used. In the case of P-gp, the MD simulations of the 3G5U-a crystal structure in the presence of the detergent cholate suggest that the structure observed crystallographically was influenced by the presence of the detergent cholate. Namely, the aggregation of cholate between

the NBDs stabilizes a conformation in which the NBDs are widely separated. When P-gp was embedded in a membrane environment, a stable conformation was only obtained when the simulation conditions matched closely the experimental conditions under which P-gp is transport active. Specifically, the presence of Mg^{2+} was required. In addition, the catalytic histidines, His583 and His1228, along with His149 were required to be doubly protonated. Mg^{2+} bound to the magnesium binding sites in the Walker B motif and to an additional site on the exterior surface of each NBD. The conformation of membrane-embedded P-gp differs from that of the crystallographic conformation in a number of key areas. In particular, the NBDs pincer inward forming a C-terminal NBD:NBD interface. The relative orientation of the P-gp NBDs and angle of intersection of the Walker B motifs is similar to that of the MalK open state crystal structure. As with all simulation studies, the extent to which the structure has relaxed within the membrane environment and truly represents the physiological structure is limited by both the force field and the time scales that could be investigated. Nevertheless, the structure obtained in the simulations provides a better overall match to a wide range of cross-linking data as compared to the original crystal structure. In addition, water is excluded from the center of the TM pore in the simulated conformation, again in better agreement with experimental results than the 3G5U-a crystal structure. More generally, the work highlights how even small changes in the environmental conditions can affect the results obtained in theoretical as well as experimental studies that the structures of integral membrane proteins solved under conditions that are far from those under which they are physiologically active must be treated with caution.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The equilibrated structure of membrane embedded P-gp along with a detailed description of the methodology used to perform the simulations is provided. Also provided is a figure showing the distance between the Walker B and signature motifs of the NBDs as a function of time for the simulations of P-gp in the presence of cholate. In addition, plots of the Cα RMSF as a function of residue number and the SASA of the TMDs as a function of simulation time are given for the series 4 simulations (membrane embedded P-gp). This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: a.e.mark@uq.edu.au.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. Richard Callaghan, Anthony M. George, and Christopher A. McDevitt for their discussions and insights. This work was supported by the Australian Research Council (DP110100327) and the Merit Allocation Scheme on the NCI National Facility at the ANU. M.L.O. is supported by a University of Queensland Postdoctoral Fellowship.

■ DEDICATION

We thank Wilfred van Gunsteren for support, encouragement, and friendship over many years.

■ REFERENCES

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

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

(3) Juliano, R. L.; Ling, V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta* **1976**, *455*, 12−162.

(4) Dawson, R. J.; Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **2006**, *443*, 180−5.

(5) Chen, J.; Lu, G.; Lin, J.; Davidson, A. L.; Quiocho, F. A. A Tweezers-like motion of the ATP-Binding cassette dimer in an ABC transport cycle. *Mol. Cell* **2003**, *12*, 651−661.

(6) Lu, G.; Westbrooks, J. M.; Davidson, A. L.; Chen, J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17969−74.

(7) Orelle, C.; Alvarez, F. J. D.; Oldham, M. L.; Orelle, A.; Wiley, T. E.; Chen, J.; Davidson, A. L. Dynamics of α-helical subdomain rotation in the intact maltose ATP-binding cassette transporter. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20293−20298.

(8) Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R. X-ray structure of the KvAP potassium channel voltage sensor in complex with an Fab. *Nature* **2003**, *423*, 33−41.

(9) Bezanilla, F. The voltage-sensor structure in a voltage-gated channel. *Trends Biochem. Sci.* **2005**, *30*, 166−168.

(10) Tieleman, D. P.; Robertson, K. M.; Maccallum, J. L.; Monticelli, L. Computer simulations of voltage-gated potassium channel KvAP. *Int. J. Quantum Chem.* **2004**, *100*, 1071−1078.

(11) Monticelli, L.; Robertson, K. M.; MacCallum, J. L.; Tieleman, D. P. Computer simulation of the KvAP voltage-gated potassium channel: steered molecular dynamics of the voltage sensor. *FEBS Lett.* **2004**, *564*, 325−332.

(12) Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Struct. Biol. Membr. Proteins* **2007**, *41*, 388−397.

(13) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(14) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(15) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. *Intermol. Forces* **1981**, *11*, 331−338.

(16) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765−783.

(17) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 710−721.

(18) Sauna, Z. E.; Kim, I.-W.; Nandigama, K.; Kopp, S.; Chiba, P.; Ambudkar, S. V. Catalytic cycle of ATP hydrolysis by P-glycoprotein: evidence for formation of the E.S reaction intermediate with ATP-gamma-S, a nonhydrolyzable analogue of ATP. *Biochemistry* **2007**, *46*, 13787−13799.

(19) Storm, J.; O'Mara, M. L.; Crowley, E. H.; Peall, J.; Tieleman, D. P.; Kerr, I. D.; Callaghan, R. Residue G346 in transmembrane segment six is involved in inter-domain communication in P-Glycoprotein. *Biochemistry* **2007**, *46*, 9899−9910.

(20) Jones, P. M.; George, A. M. Molecular-dynamics simulations of the ATP/apo state of a multidrug ATP-Binding Cassette Transporter provide a structural and mechanistic basis for the asymmetric occluded state. *Biophys. J.* **2011**, *100*, 3025−3034.

(21) Oliveira, A. S.; Baptista, A. M.; Soares, C. M. Conformational changes induced by ATP-hydrolysis in an ABC transporter: a

molecular dynamics study of the Sav1866 exporter. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1977−1990.
(22) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated force field Topology Builder (ATB) and repository: version 1.0. *J. Chem. Theory Comput.* **2011**, *12*, 4026−4037.
(23) Poger, D.; van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **2009**, *31*, 1117−1125.
(24) Poger, D.; Mark, A. E. On the validation of molecular dynamics simulations of saturated and cis-monounsaturated phosphatidylcholine lipid bilayers: a comparison with experiment. *J. Chem. Theory Comput.* **2010**, *6*, 325−336.
(25) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475−88.
(26) Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms - lysozyme and insulin. *J. Mol. Biol.* **1973**, *79*, 351−371.
(27) Maiorov, V. N.; Crippin, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular protiens. *J. Mol. Biol.* **1994**, *235*, 625−634.
(28) Daura, X.; van Gunsteren, W. F.; Mark, A. E. Folding-unfolding thermodynamics of α-heptapeptide from equilibrium simulations. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269−280.
(29) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(30) Hao, L.; Lu, R.; Leaist, D. G.; Poulin, P. R. Aggregation number of aqueous sodium cholate micelles from mutual diffusion measurements. *J. Sol. Chem.* **1996**, *26*, 113−125.
(31) Homolya, L.; Holló, Z.; Germann, U. A.; Pastan, I.; Gottesman, M. M.; Sarkadi, B. Fluorescent cellular indicators are extruded by the multidrug resistance protein. *J. Biol. Chem.* **1993**, *268*, 21493−21496.
(32) Rothnie, A.; Storm, J.; Campbell, J.; Linton, K. J.; Kerr, I. D.; Callaghan, R. The topography of transmembrane segment six is altered during the catalytic cycle of P-glycoprotein. *J. Biol. Chem.* **2004**, *279*, 34913−34921.
(33) Modok, S.; Heyward, C.; Callaghan, R. P-glycoprotein retains function when reconstituted into a sphingolipid- and cholesterol-rich environment. *J. Lipid Res.* **2004**, *45*, 1910−1918.
(34) Oldham, M. L.; Chen, J. Crystal structure of the maltose transporter in a pre-translocation intermediate state. *Science* **2011**, *332*, 1202−1205.
(35) Rosenberg, M. F.; Kamis, A. B.; Callaghan, R.; Higgins, C. F.; Ford, R. C. Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding. *J. Biol. Chem.* **2003**, *278*, 8294−8299.
(36) Loo, T. W.; Bartlett, M. C.; Clarke, D. M. Human P-glycoprotein is active when the two halves are clamped together in the closed conformation. *Biochem. Biophys. Res. Commun.* **2010**, *395*, 436−440.
(37) Loo, T. W.; Bartlett, M. C.; Clarke, D. M. Val^{133} and Cys^{137} in Transmembrane Segment 2 Are Close to Arg^{935} and Gly^{939} in Transmembrane Segment 11 of Human P-glycoprotein. *J. Biol. Chem.* **2004**, *279* (18), 18232−18238.
(38) Loo, T. W.; Clarke, D. M. The Packing of the Transmembrane Segments of Human Multidrug Resistance P-glycoprotein Is Revealed by Disulfide Cross-linking Analysis. *J. Biol. Chem.* **2000**, *275* (8), 5253−5256.
(39) Crowley, E.; O'Mara, M. L.; Kerr, I. D.; Callaghan, R. Transmembrane helix 12 plays a pivotal role in coupling energy provision and drug binding in ABCB1. *FEBS J.* **2010**, *277*, 3974−3985.
(40) Storm, J.; Modok, S.; O'Mara, M. L.; Tieleman, D. P.; Kerr, I. D.; Callaghan, R. Cytosolic region of TM6 in P-glycoprotein: topographical analysis and functional perturbation by site directed labeling. *Biochemistry* **2008**, *47*, 3615−3624.
(41) Loo, T. W.; Bartlett, M. C.; Clarke, D. M. Identification of residues in the drug translocation pathway of the human multidrug resistance P-glycoprotein by arginine mutagenesis. *J. Biol. Chem.* **2009**, *284*, 24074−24087.
(42) Stenham, D.; Campbell, J. D.; Sansom, M. S. P.; Higgins, C. F.; Kerr, I. D.; Linton, K. J. An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulphide cross-linking and homology modeling. *FASEB J.* **2003**, *17*, 2287−2289.
(43) Loo, T. W.; Clarke, D. M. Determining the dimensions of the drug-binding domain of human P-glycoprotein using thiol cross-linking compounds as molecular rulers. *J. Biol. Chem.* **2001**, *276*, 36877−36880.
(44) Loo, T. W.; Clarke, D. M. Vanadate trapping of nucleotide at the ATP-binding sites of human multidrug resistance P-glycoprotein exposes different residues to the drug-binding site. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3511−3516.
(45) Loo, T. W.; Clarke, D. M. Drug-stimulated ATPase Activity of Human P-glycoprotein Requires Movement between Transmembrane Segments 6 and 12. *J. Biol. Chem.* **1997**, *272* (84), 20986−20989.