

RESEARCH ARTICLE

# The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment

Karmen Condic-Jurkic<sup>1,2</sup>, Nandhitha Subramanian<sup>1,2</sup>, Alan E. Mark<sup>2</sup>, Megan L. O'Mara<sup>1\*</sup>

<sup>1</sup> Research School of Chemistry, The Australian National University, Canberra, ACT 2601, Australia,

<sup>2</sup> School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, ACT 4072, Australia

\* [megan.o'mara@anu.edu.au](mailto:megan.o'mara@anu.edu.au)



## Abstract

Despite decades of research, the mechanism of action of the ABC multidrug transporter P-glycoprotein (P-gp) remains elusive. Due to experimental limitations, many researchers have turned to molecular dynamics simulation studies in order to investigate different aspects of P-gp function. However, such studies are challenging and caution is required when interpreting the results. P-gp is highly flexible and the time scale on which it can be simulated is limited. There is also uncertainty regarding the accuracy of the various crystal structures available, let alone the structure of the protein in a physiologically relevant environment. In this study, three alternative structural models of mouse P-gp (3G5U, 4KSB, 4M1M), all resolved to 3.8 Å, were used to initiate sets of simulations of P-gp in a membrane environment in order to determine: a) the sensitivity of the results to differences in the starting configuration; and b) the extent to which converged results could be expected on the times scales commonly simulated for this system. The simulations suggest that the arrangement of the nucleotide binding domains (NBDs) observed in the crystal structures is not stable in a membrane environment. In all simulations, the NBDs rapidly associated (within 10 ns) and changes within the transmembrane helices were observed. The secondary structure within the transmembrane domain was best preserved in the 4M1M model under the simulation conditions used. However, the extent to which replicate simulations diverged on a 100 to 200 ns timescale meant that it was not possible to draw definitive conclusions as to which structure overall was most stable, or to obtain converged and reliable results for any of the properties examined. The work brings into question the reliability of conclusions made in regard to the nature of specific interactions inferred from previous simulation studies on this system involving similar sampling times. It also highlights the need to demonstrate the statistical significance of any results obtained in simulations of large flexible proteins, especially where the initial structure is uncertain.

## OPEN ACCESS

**Citation:** Condic-Jurkic K, Subramanian N, Mark AE, O'Mara ML (2018) The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. *PLoS ONE* 13(1): e0191882. <https://doi.org/10.1371/journal.pone.0191882>

**Editor:** Claudio M Soares, Universidade Nova de Lisboa Instituto de Tecnologia Quimica e Biologica, PORTUGAL

**Received:** November 15, 2017

**Accepted:** January 13, 2018

**Published:** January 25, 2018

**Copyright:** © 2018 Condic-Jurkic et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All simulation protocols, trajectories and input files and movies have been made available on Figshare. doi:[10.6084/m9.figshare.4806544.v4](https://doi.org/10.6084/m9.figshare.4806544.v4).

**Funding:** This work was supported by grants from the Australian Research Council (DE120101550 to MLO, DP130102153 to AEM and MLO), the National Health and Medical Research Council (APP1049685 to MLO and AEM), and with the assistance of resources and services from the

National Computational Infrastructure (NCI), which is supported by the Australian Government.

**Competing interests:** The authors have declared that no competing interests exist.

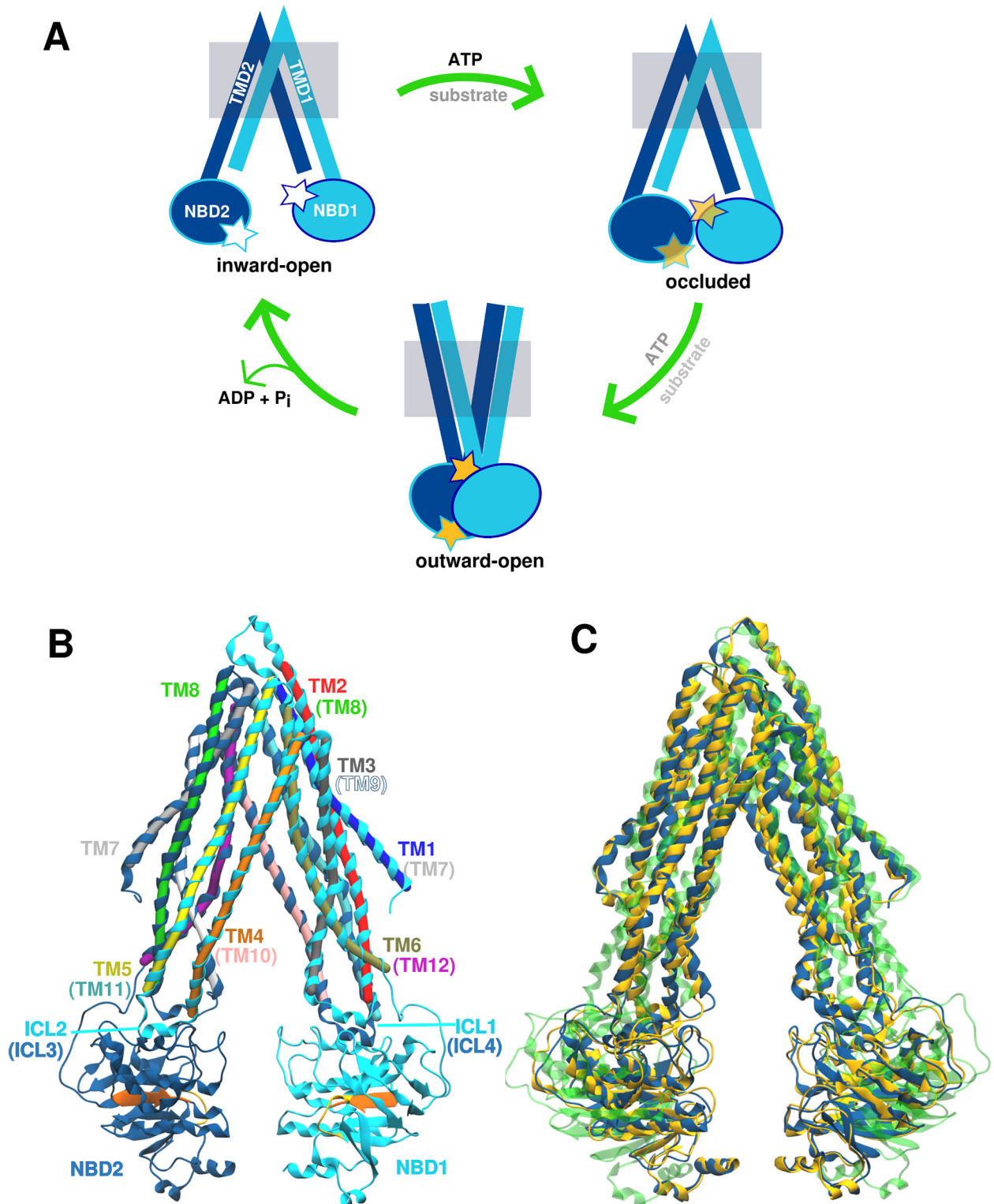
## Introduction

P-glycoprotein is a member of a large family of membrane-bound ATP-Binding Cassette (ABC) transporters, which utilise ATP hydrolysis to induce a conformational change to transport a variety of substrates across the membrane [1–3]. P-glycoprotein (P-gp) has been implicated in multidrug resistance of cancer cells, and is able to efflux a diverse range of neutral and cationic xenobiotics, varying in size and chemical properties [4–6]. Despite being one of the most comprehensively studied ABC transporters, many aspects of the P-gp transport mechanism remain elusive, including how P-gp binds such a wide range of substrates, and how ATP hydrolysis drives the conformational change that ultimately expels substrate from the cell [7, 8].

The initial elucidation of homologous bacterial ABC transporter crystal structures [9–11], solved in different conformations, has made a major contribution to our understanding of the P-gp transport mechanism. These structures have provided a basis for construction of a putative transport cycle (Fig 1A). In this cycle, substrate binds to the transmembrane pore when the transporter is in an inward-facing (IF) conformation, in which the transmembrane pore is occluded to the extracellular environment and open to the cytoplasm. The conformational changes associated with ATP binding and hydrolysis lead to an outward-facing (OF) state in which the pore is open to the extracellular space, while the cytoplasmic entrance is occluded.

In addition to bacterial transporters, a number of crystal structures of homologous eukaryotic ABC transporters have been solved [12–14], including several different structures of P-glycoprotein from mouse [15–18] and *C. elegans* [19]. The structure of human P-gp has not been solved thus far, but it shares 87% sequence identity with its murine counterpart. An example of an inward-open crystallographic conformation of mouse P-gp is shown in Fig 1B (PDB ID: 4M1M [16]), which shares a common architecture with the rest of the B-subfamily of ABC transporters and has two (pseudo)symmetrical halves with domain swapping features (Fig 1B). Each half consists of one transmembrane domain (TMD) containing six transmembrane (TM) helices and one nucleotide binding domain (NBD). Each NBD contains the highly conserved Walker A and signature motif sequences that interact to form two catalytic ATP binding sites upon head-to-tail dimerization of the NBDs. This interaction has been proposed to result in a conformational change to the outward-open conformation [20]. With the exception of two bacterial transporters, Sav1866 [10] and MsbA [11], which crystallised in an outward-open conformation, all available structures of ABC exporters are in an inward-open conformation. The conformation of P-gp under physiological conditions is still an open debate as membrane protein purification usually involves extraction from the native membrane into detergent micelles or artificial lipid bilayers, which may result in improper folding, denaturation, or loss of activity [21]. The choice of detergent can influence both the protein conformation [22, 23] as well as crystallisation outcomes [24], such as the internal order of the crystal. Furthermore, crystal packing may select for conformations that readily form a crystal lattice, putting a strain on the structure and challenging the assumption that the crystal structure is always closely related to a biologically active conformation [25, 26].

Molecular dynamics (MD) simulations are increasingly used to explore protein structure and dynamics at atomic resolution [27, 28]. In such simulations, proteins can sample different conformations under given conditions, and provide insights into structure-function relationships. A large number of MD studies have been performed on mouse P-gp, as well as on homology models of human P-gp based either on mouse P-gp or homologous ABC transporters. The various aims of these studies were to gain a better understanding of the protein dynamics in the membrane environment [29–34], examine the effect of ATP and substrate binding [35–38], or to identify key residues involved in substrate binding [39, 40]. None of the



**Fig 1. Putative transport cycle of P-glycoprotein.** (A) The transport cycle is assumed to start from the inward-open conformation, followed by the occluded state assumed to form upon ATP and/or substrate binding. Finally, the transporter adopts the outward-open conformation driven by ATP hydrolysis and subsequently, the substrate is released in the extracellular space. (B) The inward-open conformation of crystallised mouse P-gp structure (4M1M) contains 12 transmembrane helices (TM1-12) and two nucleotide binding domains (NBDs). Helices TM1-6 and NBD1 are shown in light blue, while helices TM7-12 and NBD2 are given in dark blue shade. The conserved residues involved in ATP binding and hydrolysis,

Walker A and signature motifs, are highlighted in orange and yellow, respectively (see S1 Fig for the back view). (C) Overlap of the three crystal structure models of mouse P-gp solved at 3.8 Å: 3G5U (yellow), 4KSB (green) and 4M1M (blue). Differences in the secondary structure assignment of TM helices between 3G5U and 4M1M are visible, as well as wider separation between NBDs found in 4KSB model.

<https://doi.org/10.1371/journal.pone.0191882.g001>

studies reported any spontaneous conformational change from the inward- to outward-facing conformation, although several targeted MD simulations were used to investigate potential transition pathways between these two conformations [41, 42]. Despite such studies, the molecular mechanism of P-gp remains poorly understood. The difficulty is that both simulations and indeed different experiments yield apparently contradictory results. In fact, one of the few findings that have emerged from both the simulations and experimental data is that the structure of P-gp is highly flexible. To date, MD simulations, EPR and NMR studies of P-gp all suggest that it is best represented by a dynamic ensemble of structures, rather than a series of distinct conformational snapshots [43–49]. The conformational dynamics within P-gp are expected due to the large-scale conformational changes required for the proposed transport cycle, although findings such as the unfolding of certain helices were reported in several MD studies. The observed unfolding has been attributed either to the relatively high Gly/Pro content of the TMDs [46], or the instability of the crystal structure under the simulated conditions [32]. However, these findings also might have been simply simulation artefacts.

There are more than 20 crystal structures of mouse P-gp currently deposited in Protein Data Bank [15–18, 50]. These have been solved under varying crystallisation conditions, with dataset resolutions ranging from 3.4–4.4 Å. While these lower-resolution datasets (3–4 Å) can provide information about secondary structure motifs, the positions of the side-chains and any ligands are difficult to assign precisely. This leads to increased uncertainty regarding the atomic positions [51], and potentially to local structural instabilities during MD simulations [52, 53]. This is especially relevant in the context of the original 2009 mouse P-gp structure (PDB ID: 3G5U), solved to 3.8 Å resolution [15], which was subjected to revision after the publication of a *C. elegans* P-gp crystal structure (resolved to 3.4 Å) revealed discrepancies between the two models [19]. These discoveries led to the re-refinement of the original diffraction dataset [16], resulting in an alternate structural model of P-gp (PDB ID: 4M1M) with better structure refinement parameters, summarised in Table 1. The most notable difference between the original 3G5U model and the revised 4M1M model is the change in the registry of four of the twelve transmembrane helices (TM), namely TM3, TM4, TM5, and TM12 (Fig 1B and 1C). A

**Table 1. Refinement statistics for the three P-glycoprotein crystallographic models [15–17].**

Crystallographic refinement statistic	3G5U	4M1M	4KSB
Resolution range high (Å)	3.8	3.8	3.8
Resolution range low (Å)	19.98	24.89	92.56
Number of reflections	41131	41385	21753
R value (working set)	0.306	0.267	0.357
Free R value	0.346	0.267	0.357
Free R value test set	4203	4203	1062
Estimated coordinate error (maximum likelihood)	-	0.46	0.82
Phase error (degrees, maximum-likelihood)	-	27.47	38.35
Luzzati plot	0.86	-	-
SigmaA	0.82	-	-
Ramachandran outliers (asymmetric unit)	594	92	79
Atomic coordinates in asymmetric unit	18352	18378	9171

<https://doi.org/10.1371/journal.pone.0191882.t001>



comprehensive description of the differences between 3G5U and 4M1M P-gp can be found in Li et al [16].

A third structural model of the *apo* P-gp (PDB ID: 4KSB), also resolved to 3.8 Å, was published around the same time as the 4M1M model, but derived from an independent set of crystallographic data [17]. The 4KSB model was crystallised using a different protocol to that of 3G5U/4M1M and was subjected to reductive methylation to promote crystallisation. The structure refinement parameters for the 4KSB model were close to that of the 4M1M model, however a wider separation between the nucleotide binding domains (NBDs) was proposed in the 4KSB structural model (Fig 1C). Changes in the helical registry were again noted in TM3, TM4, TM5, and TM12, when compared to the original 3G5U structure. The 4KSB model also includes a registry shift in the same four TM helices, however it is notable that the amino acid assignment in TM12 differs to that of 4M1M model. This results in three unique structural models of *apo* P-gp at 3.8 Å resolution (Fig 1C). Ideally, MD simulations initiated from any one of these crystallographic models should sample a similar region of conformational space, provided the MD simulations were conducted under the same conditions and of sufficient duration. However, such conformational changes may not be sampled on currently accessible time scales, which range from nanoseconds to microseconds [54].

The majority of the simulations published to date have been based on trajectories in the order of 100 ns, using the 3G5U crystal structure as the starting conformation. The question is, given the inherent flexibility of P-gp and the uncertainties in the structure of P-gp, can the results of such simulations be considered reliable? Here we examine the conformational space sampled by simulations initiated from the three P-gp structures corresponding to PDB entries 3G5U, 4M1M and 4KSB, each of which contain differences in their residue assignment. We compare the results of 9 independent 200 ns MD simulations initiated from the three different 3.8 Å resolution P-gp structures embedded in a membrane environment, following a standard simulation protocols for membrane proteins. The protein dynamics, structural integrity and backbone fluctuations of each systems were examined across all 9 simulations. The aim is to determine which of the features extracted from the simulations, if any, are observed in all three models, and which starting configuration could be considered most reliable under the simulation conditions examined.

## Results

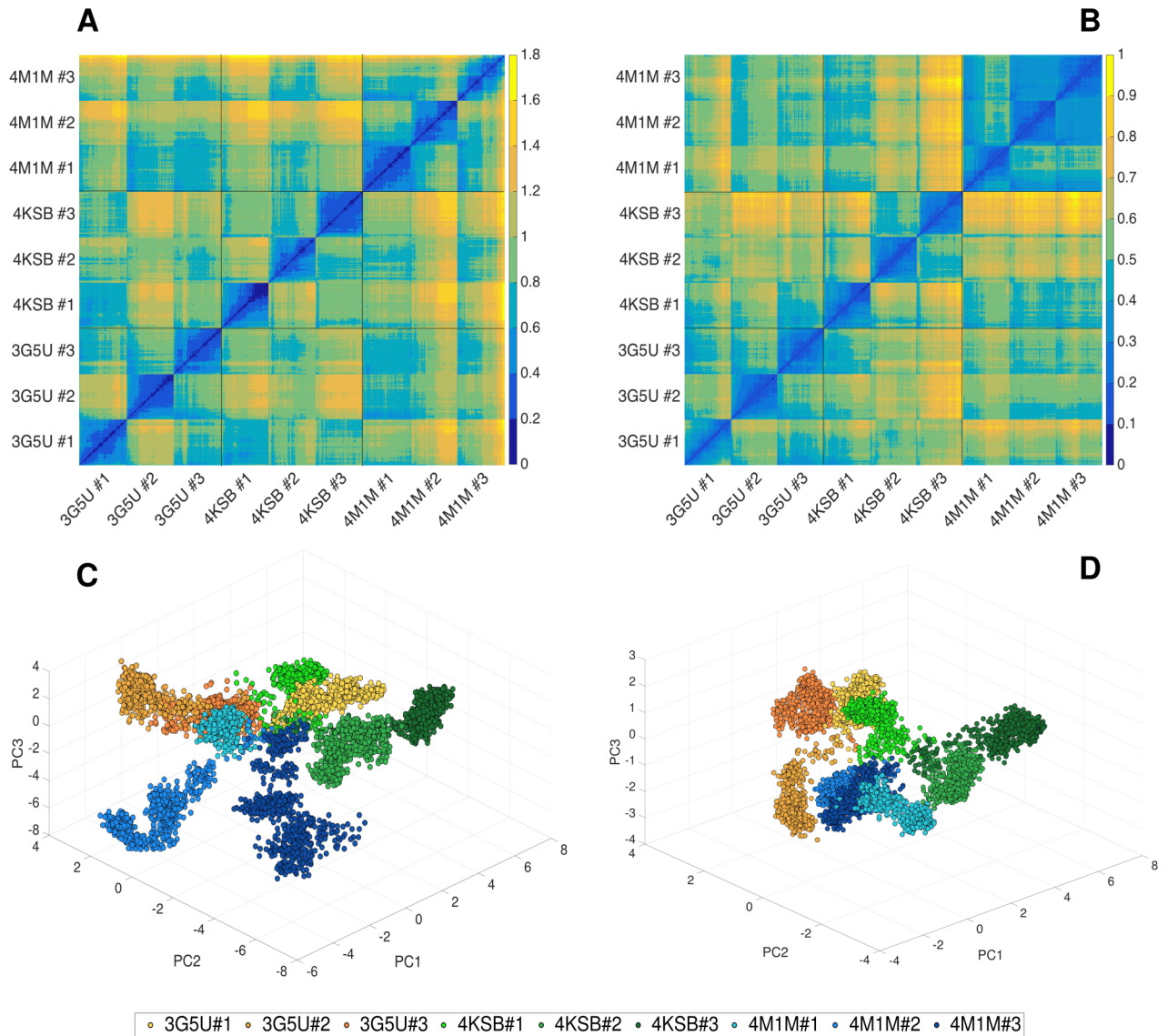
### Protein dynamics

A set of three replicate simulations of 200 ns duration were performed for each of the membrane-embedded, solvated P-gp systems corresponding to the three alternative structural models (PDB ID: 3G5U, 4M1M, and 4KSB). The simulations did not include the linker (~ 60 residues), which was unresolved in all three crystal structures. The root mean squared deviations (RMSD) of the backbone atoms from the positions in the starting crystal structure were calculated for each simulation (S2 Fig). In each of the nine simulations, the backbone RMSDs showed a rapid increase in the first 10 ns, followed by more gradual rise over the course of the simulation. The final backbone RMSDs ranged from 0.7 nm to 1.0 nm across the nine simulations. Visual inspection showed that these large deviations were the result of systematic conformational changes. Most notably, in all simulations, the separation between the NBDs decreased over time in comparison to the crystal structure. The average backbone RMSD of the individual NBDs ranged between 0.3–0.4 nm (S3 Fig), mostly due to variations in the disordered loops. There were no significant differences in the structural stability of the NBDs of each of the P-gp structural models. The average backbone RMSD of the TMD was higher than that of the isolated globular NBDs, ranging between 0.4 and 0.5 nm with respect to the initial

starting structure over the nine simulations. The disparity between the RMSDs measured for the entire system, and each domain is consistent with the conformational rearrangement of the domains observed visually (S3 Fig). In all three P-gp systems, this conformational change involves the pivoting of the NBDs inward to form a protein-protein interface and a narrowing of the  $\Lambda$ -shaped conformation of the TMD in the 4KSB systems.

The backbone RMSD values measured for each replica using the respective starting crystal structures as a reference are comparable, suggesting that the extent of structural change is similar in all three systems. To determine whether all systems converge to a similar conformation, the pairwise RMSD was calculated across the 9 simulation trajectories. Fig 2A shows the backbone RMSD between pairs of structures corresponding to each frame of the nine trajectories presented as a 2D heat map. The diagonal dark blue line represents the backbone RMSD of a structure to itself and thus has an RMSD of 0 nm. Pairs of structures with the greatest backbone RMSD are in yellow. The heat map shown in Fig 2A reveals that for replicas of the same system, the pairwise RMSDs of conformations range from 0.4–1.4 nm, indicating a variety of conformations are adopted by P-gp in the different replicas. A similar range of pairwise RMSDs was measured between conformations from three the different systems. Here the minimum pairwise RMSD was 0.6 nm, and the maximum was 1.8 nm. The overlap of the final three conformations obtained for each system is shown in Fig 3, illustrating the degree of conformational diversity. The main conformational differences can be attributed to the movement of the NBDs. Due to the differences in TMDs between the three structural models, the pairwise backbone RMSDs were calculated for the TMD only to examine the similarity of these domains between three systems. Fig 2B shows a heat map corresponding to just the TMDs. The TMD in the 4M1M system shows less variation (lower RMSD) between replicates compared to either 3G5U or 4KSB.

Similar results were obtained from a principle component (PC) analysis performed on the concatenated trajectory containing all 9 simulations generated from the three crystal structures. The major contribution to the first principal component (PC1), which is the eigenvector containing the largest variance and accounts for 34% of protein motion, originates from the motion of the NBDs (S4 Fig), while the greatest fluctuations of the TMD were found in the intracellular loops ICL1 and ICL4 connecting TM2-TM3 and TM10-TM11, respectively. Principal components corresponding to the second (PC2) and third (PC3) largest variance have a similar intensity in the TMD region, while the NBD contributions are reduced (S4 Fig). The fraction of variance corresponding to PC2 and PC3 is 15% and 10%, respectively, for both the total protein motion and the TMD only. The contributions of all 10 calculated PCs are given in S5 Fig, including the values corresponding to the PCs calculated for each replica separately. Pairwise comparisons between the principal components obtained from each replica and from the concatenated trajectory are provided in S1, S2 and S3 Tables for the 3G5U, 4KSB and 4M1M models, respectively. Plotting this concatenated trajectory as a function of the first three principal components (Fig 2C) reveals that the distribution of P-gp conformations differs not only between the simulations launched from the different crystal structures, but also between replicas of the same system. Trajectories initiated from 3G5U are shown in orange, 4KSB in green and 4M1M in blue tones. The data points corresponding to the same trajectory are broadly clustered together, and there is little overlap between any of the trajectories in this space, also confirmed by the low covariance overlap calculated for each pair of all the replicas (S4 Table). This result suggests that the region of conformational space explored by the protein in each simulation is unique. This supports the view that P-gp is a very flexible and dynamic system. However, it is not possible to distinguish between the intrinsic dynamics and other factors that might influence the observed motion, such as the local quality of the model, the removal of crystal packing forces, change in the environment, etc.

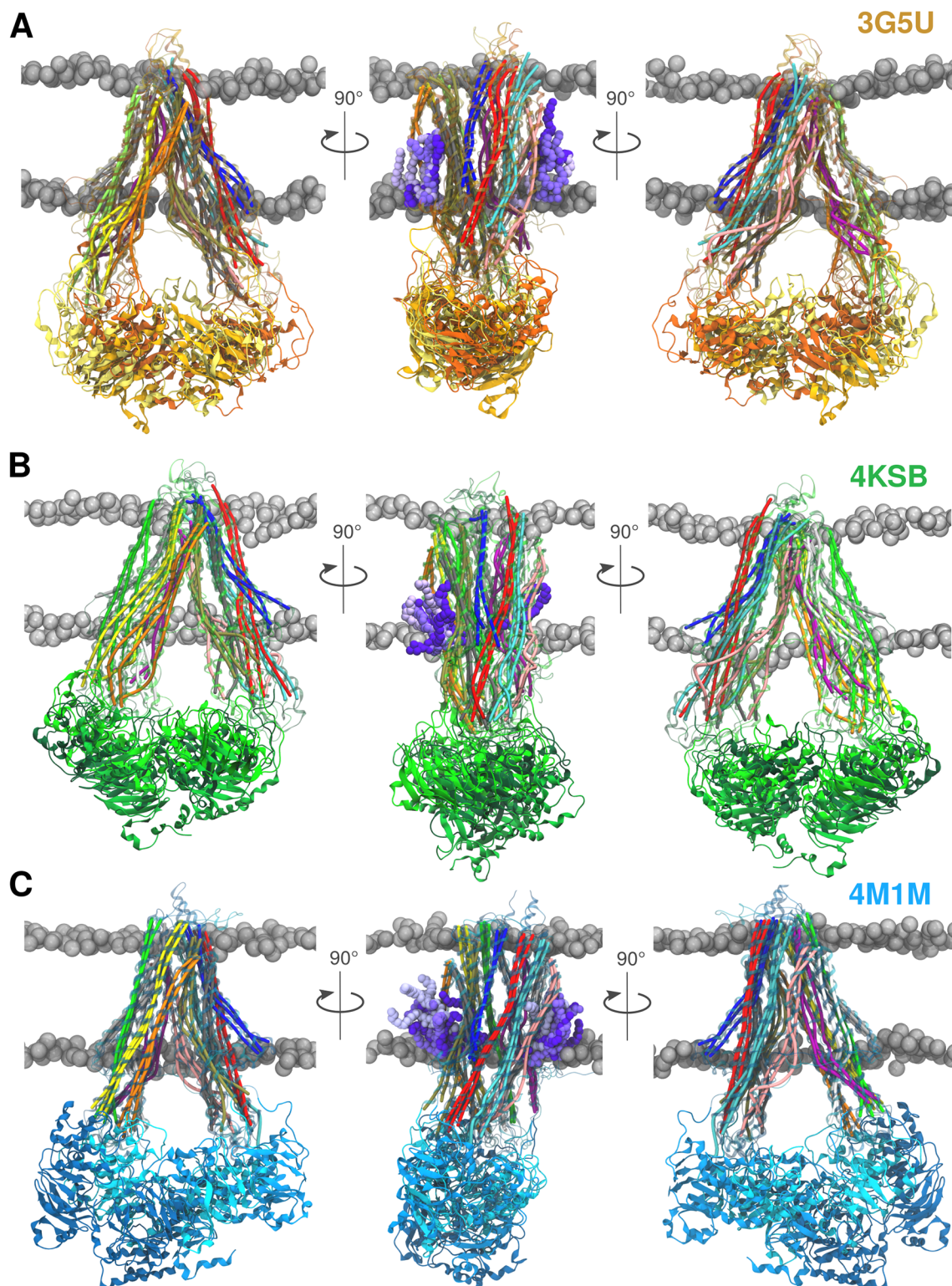


**Fig 2. Pairwise RMSD and principal component analysis (PCA).** Heatmap representing pairwise RMSD [nm] calculated for the backbone atoms of P-gp conformations sampled in all 9 trajectories started from 3G5U, 4KSB and 4M1M for (A) the entire protein and (B) the TMD coordinate subset only. Black gridlines separate the three systems, while each replica (3G5U#1-3, 4KSB#1-3, 4M1M#1-3) is visible as a square formation around diagonal line (diagonal corresponds to the RMSD of a structure to itself, which is 0 nm). Conformational space sampled in each simulation as a function of the first three principal components (PC1, PC2, PC3) obtained from the analysis of the concatenated 9 trajectories for (C) the entire protein and (D) the TMD only, respectively. Each replica is shown in a different shade of yellow (3G5U), green (4KSB) or blue (4M1M). The TMD residues used in the pairwise RMSD and PC analysis were Tyr41-Ala358 (TMD1) and Leu684-Phe990 (TMD2).

<https://doi.org/10.1371/journal.pone.0191882.g002>

PC analysis was also repeated using just the coordinates of the TMDs to remove the effects of the NBDs, which undergo large-scale movements, to understand the extent and the location of the structural changes of the TMDs in more detail. The results (Fig 2B and 2D) again show marked differences between the simulations initiated using each of the three structural models. There are also marked differences between the three replicas initiated using 3G5U and 4KSB. In contrast, the three replicas initiated using 4M1M yielded very similar results in agreement with the pairwise RMSD calculations. It is important to note that the variance in the TMDs is





**Fig 3. Final snapshots.** Overlap of the final P-gp conformations after 200 ns obtained from three replicas performed for each system. Each replica is coloured in a different shade of yellow for 3G5U (A), green for 4KSB (B), and blue for the 4M1M (C) system, where the lightest shade corresponds to replica #1 and the darkest to replica #3. The left and right panels show the TMD1 and the TMD2 at the front, respectively. The middle panel gives the side-view with lipids interacting with the TM4/6 (left side) and the TM10/12 (right side) portal. The POPC molecules are shown in different shades of violet and again, the lightest shade corresponds to replica #1. The protein

is shown in cartoon representation and the TM helices are represented as cylinders that follow the helix axis (*bendices*) to highlight the change in helical geometry [55]. The POPC/cholesterol bilayer is depicted using the VDW sphere representation of phosphate atoms (grey).

<https://doi.org/10.1371/journal.pone.0191882.g003>

lower compared to the entire protein, as expected (S4 Fig). Finally, the RMS fluctuations were measured for the C $\alpha$  atoms in the entire protein and TMDs only (S6 Fig), showing similar patterns between three systems—the greatest mobility corresponds to NBDs, the intracellular loops (ICL1-4) and the extracellular loop connecting TM1 and TM2. The RMS fluctuations for the entire protein are comparable between the three systems, but the smallest TMDs fluctuations were found in 4M1M system. All this suggests that the 4M1M TMDs are more structurally stable in the simulations than either 3G5U or 4KSB.

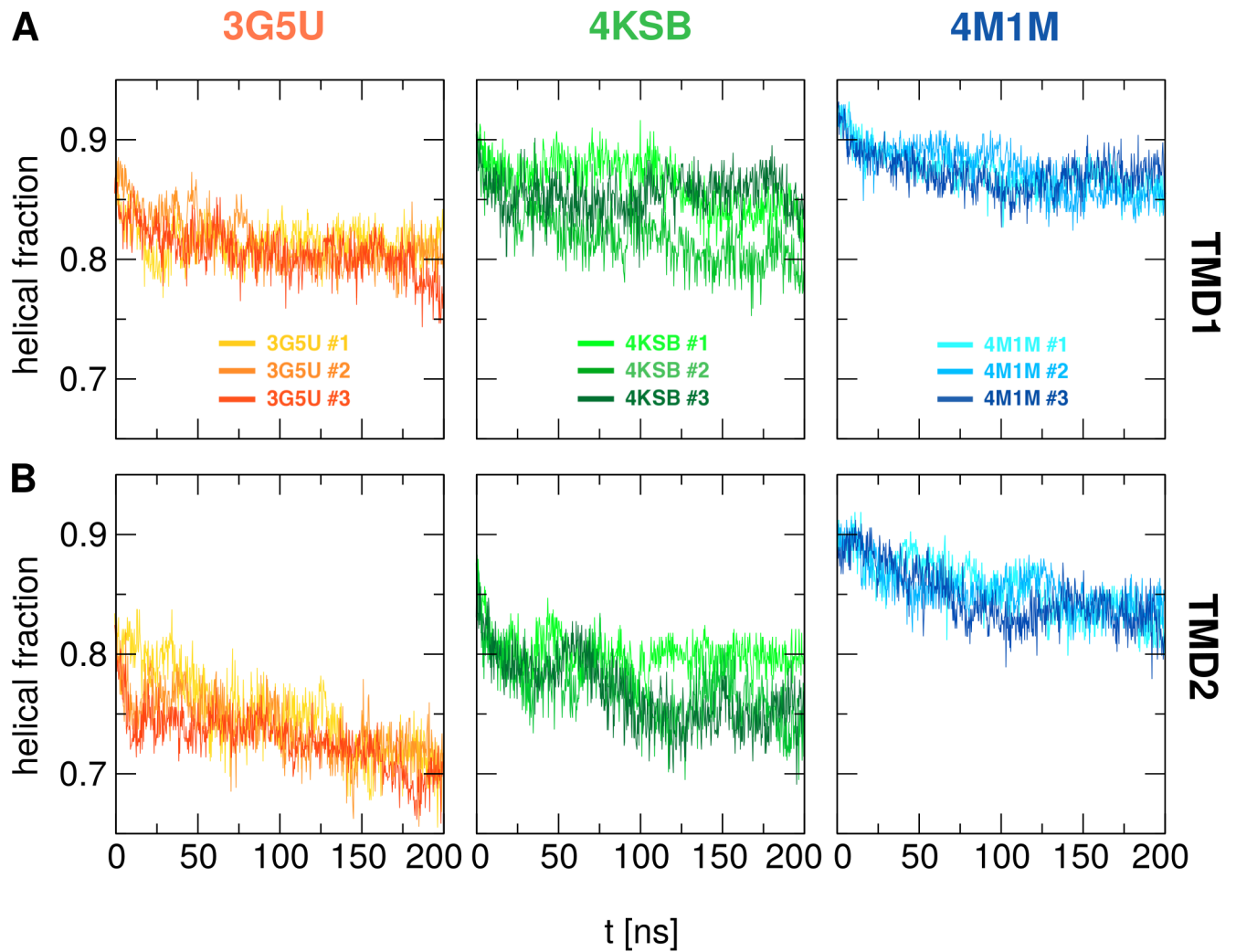
### Transmembrane domains

The most notable differences between the amino acid assignment in the 3G5U, 4M1M, and 4KSB structural models are in the TM helices 3, 4, 5 and 12. In addition to these differences, 4KSB features a wider angle between the TMD helices (Fig 1C). For each simulation, the backbone RMSD calculated with respect to its starting crystal structure was 0.4–0.5 nm (S2 and S3 Figs), suggesting that the scale of the structural changes within the TMDs is comparable for all three systems during each 200 ns simulation. However, the nature of these conformational changes was different in each simulation. The secondary structure content of TMD1 (TM1-6) and TMD2 (TM7-12) was assessed using the DSSP algorithm and the fraction of helical content was calculated as a function of time, shown in Fig 4. The average values for each run are given in Table 2.

The results show that 4M1M has the highest percentage of helical content, both in the starting crystal structure and in the resulting simulations (Table 2). The ranking of the structures based on helical content is 4M1M > 4KSB > 3G5U, while TMD1 has a higher percentage of helical content than TMD2. A certain degree of unfolding was found in all three systems, but it was especially pronounced in the 3G5U system (Table 2). The DSSP plots as function of time reveals loss of helicity in TM4, TM5, TM10, TM11, TM12 and the intracellular coupling loops (ICLs) that form an interface with the NBDs, ICL3 and ICL4 in particular (S7 Fig). The unfolding of the coupling loops was also observed to a lesser extent in the simulations initiated from the 4KSB model, while in the 4M1M simulations the helical motifs in the ICLs were maintained. The degree of unfolding increased progressively over the 200 ns simulation, although the largest drop in helical content occurs in the first 50 ns (Fig 4). The loss of secondary structure resulted in the deformation of the TM helices of both 3G5U and 4KSB and each simulation resulted with a different TMD conformation. The range of TMD conformations is illustrated with the overlay of the final snapshots for the three replicas for 3G5U, 4KSB and 4M1M in Fig 3. The overlay of the three final conformations of the 4M1M (Fig 3C) system shows that the TMDs adopt a consistent conformation with a well-defined shape, while the overlays of the 3G5U and 4KSB conformations show a range of lateral helical displacements.

Despite the relative stability of the 4M1M TMDs, localised regions like the ICLs exhibit a high degree of mobility, as measured by the root mean square fluctuations (RMSF) of C $\alpha$  atoms (S6 Fig) and shown on the secondary structure plots (S7 Fig). Increased mobility was observed in the 4M1M simulations within the portal helices (TM4/6, TM10/12). These helices connect the lipid bilayer to the TM cavity and are implicated as potential entry points for substrate uptake. The protrusion of the tails of different lipids into both the TM4/6 and TM10/12 portals was repeatedly observed in the 4M1M simulations. However, the lipid tails protruded deeper into the protein cavity through the TM4/6 portal (Fig 3). Similar lipid tail protrusions





**Fig 4. Helical content of transmembrane domain.** The fraction of helical content in TMD1 (A) and TMD2 (B) as function of time for each replica started from 3G5U, 4KSB and 4M1M models calculated from the secondary structure analysis performed using the DSSP algorithm.

<https://doi.org/10.1371/journal.pone.0191882.g004>

**Table 2. Calculated average helical content (and standard deviation) from DSSP analysis of TMD1 (TM1-6) and TMD2 (TM7-12) for each replica of simulated systems 3G5U, 4KSB and 4M1M.** The amount of helical content was also averaged over all three simulations. TMD1 spans residues Trp44-Asp366, while TMD2 spans residues Trp704-Thr1011.

Helicity	TMD1			TMD2		
	3G5U	4KSB	4M1M	3G5U	4KSB	4M1M
Crystal structure	0.87	0.90	0.92	0.83	0.85	0.90
Replica #1	0.82 ± 0.02	0.86 ± 0.02	0.87 ± 0.02	0.75 ± 0.04	0.80 ± 0.02	0.86 ± 0.02
Replica #2	0.82 ± 0.02	0.82 ± 0.02	0.87 ± 0.02	0.74 ± 0.03	0.77 ± 0.03	0.86 ± 0.02
Replica #3	0.81 ± 0.02	0.85 ± 0.02	0.87 ± 0.02	0.73 ± 0.02	0.77 ± 0.03	0.85 ± 0.02
Total	0.81 ± 0.02	0.85 ± 0.02	0.87 ± 0.02	0.74 ± 0.03	0.78 ± 0.02	0.85 ± 0.02

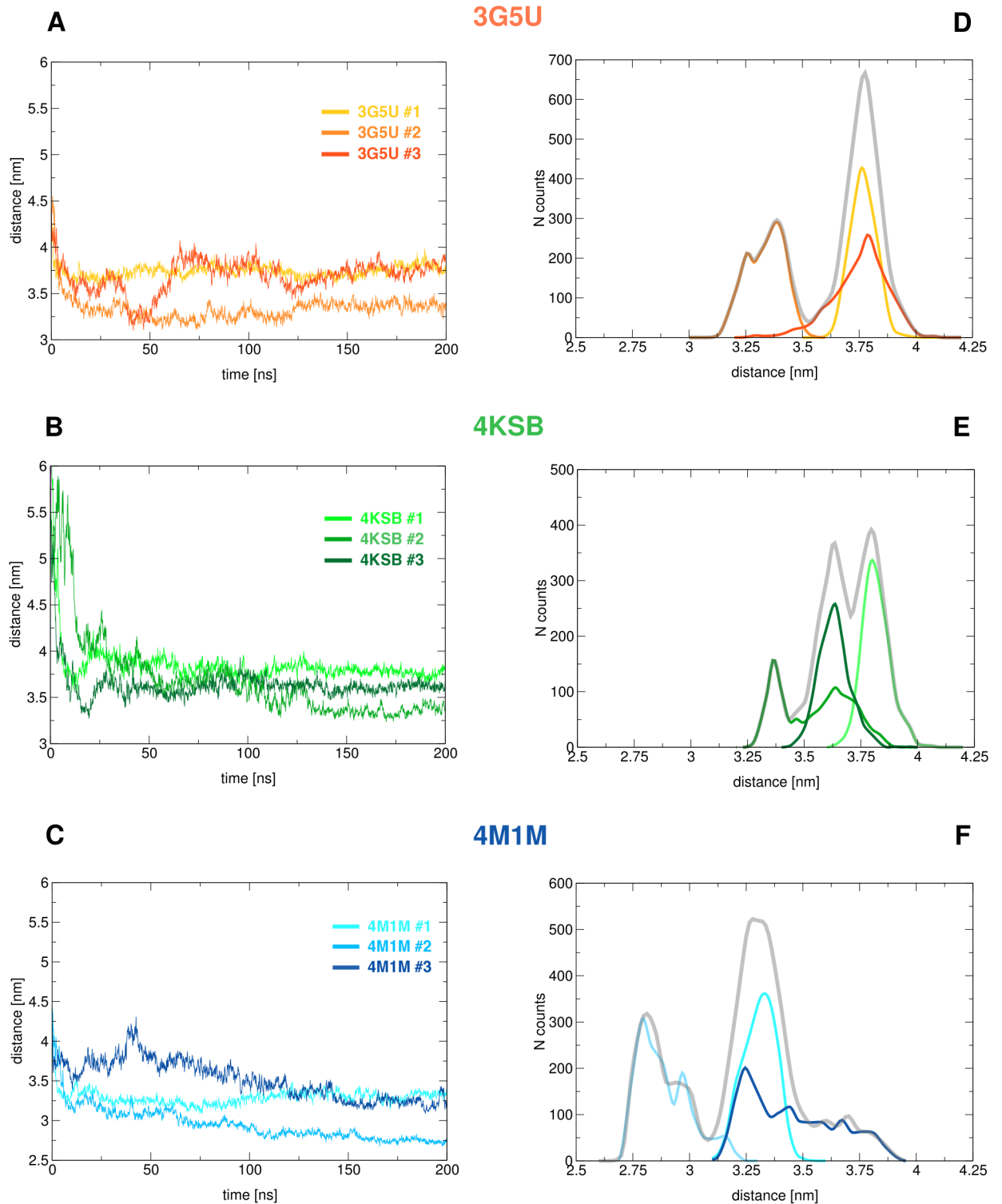
<https://doi.org/10.1371/journal.pone.0191882.t002>

through the TM4/6 portal were reported in a previous study using the 3G5U structure [46]. In the simulations initiated using 3G5U and 4KSB in this work, lipids were observed to partly protrude into the TM10/12 portal, but no lipid protruded into the TM4/6 portal. This may be a consequence of the placement of TM6 directly between TM3 and TM4 in the 3G5U and 4KSB structural models. This effectively occludes the entrance to the TM cavity in the starting conformation. The occlusion of the TM4/6 portal was also noted in the study by Ferreira *et al.*, in which the 3G5U model was used to examine the energetics of drug entry through the TM10/12 portal [40]. On the contrary, a similar study of potential drug entry pathways based on the 4M1M structure identified TM4/6 as the most energetically favourable uptake pathway [56]. Differences in the protein-lipid interactions observed in the 3G5U and 4KSB systems may be attributed to the loss of secondary structure in the TM4/6 and TM10/12 portal helices during the simulations, which directly affects the protein-lipid interface.

### Nucleotide binding domains

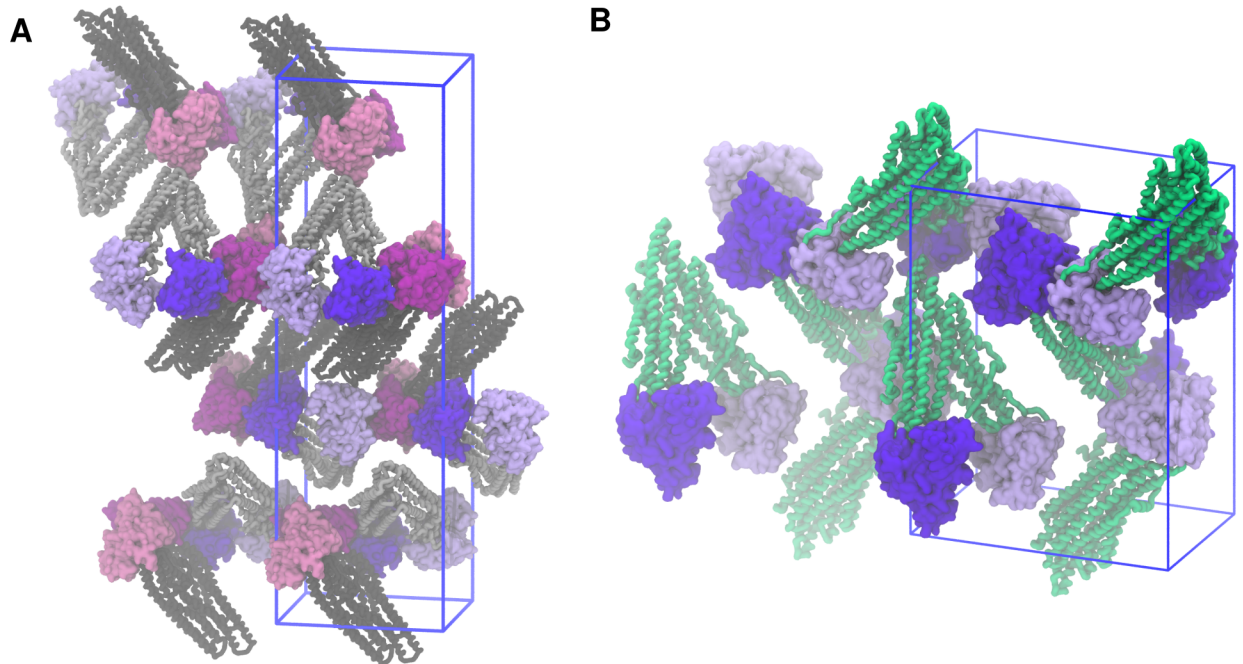
In the simulations performed in this study, all of which were unbiased, there was a rapid association of the nucleotide binding domains. This occurred within the first 10 ns in all cases (Fig 5A–5C). The most pronounced structural change was observed in the 4KSB system, which has the largest inter-NBD separation of the three structural models. After the NBDs had associated, they remained in close contact throughout the simulation. This is consistent with previous MD studies based on the 3G5U structural model [29, 30, 32]. We note, however, that Wen *et al.* proposed that larger distances are also sampled, based on a combination of MD and DEER experiments [46]. The nature of the NBD1:NBD2 interface in the different simulations is determined by the relative orientation of the NBDs at the time of contact, and the dimer interface is slightly different in each replica of each system. This is illustrated by the inter-NBD distance distribution in Fig 5D–5F. Analysis of the distances between the Walker A motif and Signature motif of opposite NBDs, which form the two ATP binding sites, shows that the association of the NBDs is asymmetric in all three structural models (S8 Fig). This is consistent with previous reports of a non-specific, asymmetric association of the NBDs in the presence and absence of ATP [29, 30, 38, 57]. Fig 3 shows the range of NBD1:NBD2 conformations obtained for each replica initiated using each of the three structural models, after 200 ns of simulation. This conformational diversity suggests that the spread of distances between labelled residues, as observed in DEER experiments, could also arise from conformations in which the NBDs remain in close contact, and do not necessarily imply there is a change in the inter-NBD separation. Other experimental studies involving fluorescence resonance energy transfer (FRET) efficiency measurements [44], intra-NBD cross-linking [58], and electron microscopy projections of membrane-embedded P-gp [59] provide evidence for the NBDs coming into very close contact during the transport cycle. For example, in the case of a homodimeric ABC transporter TM287/288 from *T. maritima* crystallised in the presence and absence of AMP-PNP, the NBDs in both crystal structures were found to be in contact, which was in agreement with supplementary results obtained by DEER measurements [12, 13]. The native conformation of NBDs under physiological conditions perhaps remains inconclusive, but the very dynamic nature of these domains is indisputable.

That the formation of a NBD1:NBD2 contact interface was strongly favoured, regardless of whether the starting structure was 3G5U, 4M1M or 4KSB, suggests that the crystallographic conformation could be the result of a combination of the crystallisation conditions and crystal packing forces. Previous simulations have shown that the aggregation of detergent between the NBDs stabilises the splayed 3G5U conformation [29]. Recent crystallographic studies of the P-gp homologue, PglK, provide strong evidence that the crystallization detergent and lattice



**Fig 5. Distances between the NBDs.** (A-C) Distances measured between the centres of mass (COM) of the nucleotide binding domains during each simulation starting from 3G5U (A), 4KSB (B), and 4M1M (C); (D-F) Distribution of the inter-NBD distances obtained for each replica independently and for all three replicas (grey), starting from 3G5U (D), 4KSB (E), and 4M1M (F).

<https://doi.org/10.1371/journal.pone.0191882.g005>



**Fig 6. Crystal packing.** (A) Orientation of P-gp molecules in the crystal lattice from which the 3G5U and 4M1M structural models were derived. The asymmetric unit contains two P-gp molecules (the NBDs highlighted in violet and magenta, respectively). (B) The crystal lattice representation of the 4KSB model, where the asymmetric unit contains a single P-gp molecule. Both crystal lattices belong to the P21 21 21 space group and the unit cell is represented with the blue wireframe.

<https://doi.org/10.1371/journal.pone.0191882.g006>

interactions influence the inter-NBD distance observed in the corresponding protein crystal structures [60]. An examination of the crystal contacts in 3G5U and 4M1M P-gp shows that the NBDs of each P-gp protein are nested in a cavity created by the three surrounding P-gp molecules (Fig 6). In the 4KSB crystal lattice, each NBD contacts the NBD of the neighbouring molecule, while the apex of the TMD of a third protein sits between the NBDs. This suggests that the crystal lattice itself may stabilise the relatively large NBD separations observed in the P-gp crystal structures.

Finally, a hinging motion of the NBDs at the NBD:TMD interface is observed in the simulations initiated using 4M1M, and to a lesser extent, in those initiated using 4KSB (S8 Fig). It is unclear whether this hinging motion of the NBDs is an artefact of the simulation or a true representation of the motion in this region [61]. The 3G5U NBD:TMD interface appears to be more resilient of the three structural models examined. As noted above, there is a loss of helical structure in the ICLs initiated using 3G5U. This also occurred to a lesser extent in the 4KSB system. We assume that the increased disorder in the ICLs allows for the re-adjustment of the TMD:NBD interface in 3G5U, stabilising the interaction.

## Discussion

In this study, three crystal structures of mouse P-gp (PDBid: 3G5U, 4M1M, 4KSB), solved at the same resolution (3.8 Å) were simulated under identical unbiased conditions. In all simulations, the NBDs rapidly made direct contact with each other, favouring direct contact between the NBDs over the large separation found in the crystal structures. This is in line with a number of other simulation studies initiated using both 3G5U ([29, 30, 32, 39]) and 4KSB P-gp [33, 34, 62]. This said, it is also possible that ~ 60 residue linker region that spans the ~ 50 Å

gap between the C-terminus of NBD1 and the N-terminus of TM7 plays an important role is determining the relative positions of the NBDs and the final NBD1:NBD2 interface. It has been suggested that the linker region may affect the extent of motion of the NBDs [31, 32], but as the structure of this region has not been resolved in any P-gp crystal structure model, regardless of resolution, this conclusion is speculative. As there is little information regarding either the conformation of the linker region or how it lies relative to the rest of the protein, it was not included in these simulations. The scope of this study was to compare the performance of three different starting points without introducing further bias. Note that P-gp remains functional in experiments where the linker has been removed [63].

In this work we have attempted to determine if 200 ns is sufficient to obtain converged structural properties and to determine which of the three selected P-gp structural models, if any, is an appropriate starting point for future simulation studies. Pairwise RMSDs and PC analysis were used to illustrate the extent to which trajectories diverged both as a result of using different starting structures, and also between replica simulations starting from the same structure. However, some common features were observed in all simulations. In particular, the NBDs moved into close contact and adopted an occluded conformation in preference to the open conformation observed in the crystal. The TMD-NBD interface was found to be unstable in the simulations initiated from both 4M1M and 4KSB simulations. This resulted in a hinge motion of the NBDs, but the origin of this behaviour is not clear. These changes may result from the loss of crystal lattice contacts, but they could also reflect shortcomings of the crystallographic model and simulation conditions, such as the missing linker, the choice of the force field and simulation parameters.

The three models had discrepancies between the residue assignments within the TMDs and the simulations showed difference in the degree of secondary structure stability in the TMD between the models. Progressive deformation of the helices (unfolding, kinking, and bending) occurred during the simulations. The TMD helices were most stable in the 4M1M simulations, which showed the highest amount of helical content, while the loss of helicity was most notable in the 3G5U model. The unfolding of the transmembrane helices has been reported in multiple previous studies based on the 3G5U structure. These studies have involved different simulation protocols and a variety of force fields [29, 30, 32, 38, 46]. Kinking in the Gly/Pro regions of the portal helices (TM4/6, TM10/12) was observed in the 4M1M system, and the relatively high content of Gly/Pro residues in the TM helices was suggested to lead to enhanced flexibility in P-gp [46]. This observation is in agreement with the kinks in portal helices found in one of the most recent P-gp crystal structures, which extend from Pro219 to Tyr243 [18]. Weaker electron density in the TM4 region, indicating local disorder, has been observed in several crystallographic models of mouse P-gp [17, 18], and in the *C. merolae* P-gp structure [14]. A triple mutation (G277V/A278V/A279V) in the *C. merolae* P-gp improved the resolution from 2.75 Å in the original dataset to 2.6 Å in the mutant, but it was associated with reduced drug transport and ATPase activities. This result suggests that reduction of the flexibility in this region affects protein function.

The majority of the previous simulation studies of P-gp have involved limited sampling and simulation times in the order of 100 ns or less. Most have used the 3G5U model as the starting point. This study found significant divergence in the structural properties between the nine replicas initiated from three different models on a 200 ns time scale. Some structural changes were rapid, such as the NBD association, while the partial unfolding in the TMDs was a continuous process throughout the entire simulation. It is possible that the extent of these deformations would not be evident in shorter simulations. Even at 100 ns, the structural perturbations observed in all three systems could potentially be interpreted as the intrinsic flexibility of the protein. Previous studies have suggested that the extent to which the structural integrity of a



protein was maintained during an MD simulation might be used as a potential indicator of the quality of a membrane protein structure [52, 53]. Among the structures examined, the TMD of the 4M1M model showed the highest degree of convergence and stability under the simulated conditions. However, the hinging motion of the NBDs meant the final conformations were further from the initial state. The other two structures (3G5U, 4KSB) showed a higher degree of unfolding in the TMDs, but the TMD-NBD contact was more stable. The intrinsic structural plasticity and divergent behavior of P-gp observed for all three structural models make it impossible to assert which of these three is the most appropriate structural model for future studies. Variations arising from different parameters used in the setup, including the choice of the force field, lipids, salt concentration, pH, temperature, and others, make this task even more complex. Clearly, in the case of P-gp, the results obtained from short simulations of this flexible membrane protein relying on low resolution data should be interpreted with caution.

## Methods

### System setup

The starting configurations of the three P-glycoprotein structures for the MD simulations were taken from the murine P-gp crystal structures corresponding to the PDB IDs: 3G5U [15], 4M1M [16], and 4KSB [17] resolved to 3.8 Å in the absence of any ligands. In the 3G5U and 4M1M structural models, which contain two molecules of P-gp in the asymmetric unit, molecule A was used to initiate the simulations. The protonation states of all the titratable residues were set at their default values at pH 7.0, except for His149, His583, His 1228, which were protonated as reported by [29]. The N- and C-termini of all three structures were acetylated and aminated, respectively. In all three structural models, the linker region of approximately 60 residues connecting the two halves of P-gp was not resolved, and thus, was not included in the simulations. P-gp was embedded in a pre-equilibrated cholesterol enriched POPC (2-oleoyl-1-palmitoyl-sn-glycero-3-phosphocholine) bilayer. The parameters for POPC were taken from [64] and the parameters for cholesterol were obtained from [29]. The membrane used to build the 3G5U and 4M1M systems contained 440 POPC and 40 cholesterol molecules, while the 4KSB P-gp structure was embedded in a larger membrane (660 POPC, 66 cholesterol) to accommodate the larger separation of NBDs. Cholesterol was distributed evenly throughout both leaflets of the bilayer prior to membrane equilibration. Cholesterol distribution around P-gp at the beginning and the end of simulations for each simulated system is shown in S9, S10 and S11 Figs. All simulation systems were solvated using the SPC water model [65]. An electrolyte concentration of 150 mM NaCl and 1.5 mM MgCl<sub>2</sub> was added to each system to mimic physiological salt concentrations. Additional chloride ions were added to neutralize each system. The total size for 3G5U/4M1M systems was ~ 235 000 atoms, while 4KSB contained ~ 340 000 atoms.

### Simulation parameters

All molecular dynamics simulations were performed using GROMACS version 3.3.3 [66] in conjunction with the GROMOS 54A7 force field for proteins [67]. All simulations were performed under periodic boundary conditions in a rectangular box. The dimensions of the box were chosen such that the minimum distance of the protein to the box wall was at least 1.0 nm. A twin-range method was used to evaluate the non-bonded interactions. Interactions within the short-range cut-off of 0.8 nm were updated every step. Interactions within the long-range cut-off of 1.4 nm were updated every 8 fs, together with the pair list. A reaction field correction was applied using a relative dielectric constant of  $\epsilon_r = 78.5$ , to minimize the effect of

truncating the electrostatic interactions beyond the 1.4 nm long-range cut-off [68]. The LINCS algorithm [69] was used to constrain the lengths of all the covalent bonds (*all-bonds*). The geometry of the water molecules was constrained using the SETTLE algorithm [70]. In order to extend the timescale that could be simulated, explicit hydrogen atoms in the protein were replaced with virtual sites, the positions of which were calculated each step based on the positions of the heavy atoms to which they were attached. This eliminates high frequency degrees of freedom associated with the bond angle vibrations involving hydrogens, allowing a time step of 4 fs to be used to integrate the equations of motion without affecting thermodynamic properties of the system significantly [71]. The simulations were carried out in the NPT ensemble at  $T = 300$  K, and  $P = 1$  bar. The temperature was coupled to an external bath using the Berendsen thermostat with a relaxation time constant of 0.1 ps [72]. The Berendsen scheme was also used for semi-isotropic pressure coupling with a relaxation time constant of 0.5 ps. Data was collected every 50 ps during the unbiased MD simulations.

Each system was energy minimized using a steepest descent algorithm. In the 3G5U and 4M1M systems, this was followed by 10 ns of equilibration simulations in which the protein was restrained using a harmonic potential on all heavy protein atoms. These restraints were gradually lowered over 5 consecutive 2 ns simulations, employing force constants of  $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ ,  $500 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ ,  $200 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ ,  $100 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ , and  $50 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ . In the widely splayed 4KSB system, the position restraints were progressively lowered over a longer period of 30 ns (5 consecutive 6 ns simulations). This longer equilibration time allowed the highly splayed protein conformation of 4KSB P-gp to gradually adjust to the removal of the crystal lattice. After equilibration, new velocities were assigned, and unrestrained MD simulations lasting 200 ns were performed in triplicate for each system.

## Analysis

Trajectory analysis was carried out using the analysis module of the GROMACS package and the MDTraj python toolkit [73]. The root mean squared deviation (RMSD) was calculated using the method of Maiorov and Crippen [74] after first performing a rotational and translational fit of each frame of the trajectory to a reference structure or domain. Pairwise RMSDs were calculated with MDTraj using a concatenated trajectory containing all 9 simulations for the backbone atoms of the entire protein and TMD coordinate subset. Principal component analysis was performed using the ProDy package based on  $C\alpha$  atoms for the entire system and the TMDs only [75]. Distances and root mean square fluctuations (RMSF) were computed using Gromacs analysis tools. Secondary structure analysis of transmembrane domain was performed using simplified DSSP algorithm implemented in the MDTraj analysis package. In the analysis, the transmembrane domain (TMD) spanned residues Asp46-Phe362 (TMD1) and Val708-Ile1008 (TMD2), while the nucleotide binding domains spanned residues Asn387-Thr626 (NBD1) and Asn1030-Ala1271 (NBD2). Images were produced using VMD [76] and Matlab2015b [77].

## Data availability

All the relevant input files for running MD simulations, as well as resulting trajectories and accompanying videos have been deposited on Figshare (<https://doi.org/10.6084/m9.figshare.4806544.v4>) [78]. Each trajectory has two accompanying videos showing P-glycoprotein from two different perspectives (*front* and *side*). More detailed description of each movie is provided in an additional document uploaded on Figshare. Model data obtained using principal component analysis are also provided and can be further explored using `loadModel()` function in the Prody package. The extension for these files is *.npz*.

## Supporting information

**S1 Fig. P-glycoprotein back view.** P-glycoprotein with TMD2 (TM7-12) in front depicted using the 4M1M model. The conserved motifs in the NBDs required for ATP binding and hydrolysis, Walker A and Signature motif (LSSGQ) are highlighted in orange and yellow, respectively. The distance between the Walker A1 (located on NBD1) and the Signature motif on NBD2 is labelled as **d1**, and the distance between the Walker A2 (NBD2) and Signature motif 1 (NBD1) is labelled as **d2**. The arrows are pointing towards the Walker A motifs.

(TIFF)

**S2 Fig. RMSD time series.** The RMSD time series measured for the backbone atoms of (A) the entire protein, (B) transmembrane domain and (C, D) the two nucleotide binding domains for each replica, started from the 3G5U (orange), 4KSB (green) and 4M1M (blue) models. All the protein snapshots were aligned to the relevant domain of the reference structure before calculating the RMSD.

(TIFF)

**S3 Fig. Average RMSD.** The mean RMSD values and standard deviations were computed for each independent replica, started from 3G5U (orange), 4KSB (green) and 4M1M (blue). The mean RMSD values are given for the entire protein and each domain separately, namely the two nucleotide binding domains (NBD1, NBD2) and the full transmembrane domain (TMD). The TMD\* contains only the TM helices, while the connecting intracellular helices (ICL1-4) and extracellular loops have been removed from these calculations, resulting in lower RMSD values for all three systems.

(TIFF)

**S4 Fig. Principal component variance.** Square fluctuations (variances) corresponding to the first 5 principle components (PC1-5) calculated using a concatenated trajectory containing all 9 simulations started from three different crystal structure models (3G5U, 4KSB, 4M1M). The upper panel shows the principal components calculated for the entire protein, while the lower panel shows the principal components obtained from the analysis of the TMDs only (residues Asp46-Phe362 and Val708-Ile1008). The highest fluctuations in the protein correspond to the NBD movement, while in the TMDs the highest fluctuations are found in the intracellular helices ICL1 (connecting TM2-TM3) and ICL4 (connecting TM10-TM11), which form an interface with NBD1. The extracellular loop connecting TM1 and TM2 is also very dynamic.

(TIFF)

**S5 Fig. Fraction of variance.** Fraction of variance corresponding to the first 10 principal components (PC1-10) calculated using each simulation independently (coloured), and a concatenated trajectory containing all 9 simulations started from the three different crystal structure models (black) for (A) the entire protein and (B) only the TMDs. The concatenated trajectory yields similar values for the first three PCs for both the total protein and the TMDs only: 34%, 15%, and 10%, respectively. There is greater variability between the PCs resulting from the analysis of each independent simulation. Subspace overlap between the PCs obtained from each replica and the concatenated trajectory is given in S1–S3 Tables.

(TIFF)

**S6 Fig. RMSF.** Root mean squared fluctuations (RMSF) of  $C\alpha$  atoms in P-gp calculated using data points from three trajectories generated for each system: 3G5U (orange), 4KSB (green) and 4M1M (blue). The upper panel shows the RMSF of  $C\alpha$  atoms of the entire protein, while the lower panel shows RMSF for the TMD only. The highest fluctuations correspond to NBDs

and the intracellular loops forming the interface between TMD and NBDs. The portal helices TM4/6 and TM10/12 have higher RMSF compared to the other helices.

(TIFF)

**S7 Fig. Secondary structure analysis.** Secondary structure analysis of TMD1 (TM1-6) and TMD2 (TM7-12) calculated using a simplified version of the DSSP algorithm implemented in the MDTraj package. In the simplified version, only helical (blue), coil (cyan) and strand (yellow) elements are assigned. White vertical lines separate results obtained from each replica performed for the 3G5U, 4KSB and 4M1M systems, while the length of each TM helix is indicated on the right axis.

(TIFF)

**S8 Fig. Domain distances.** (A-C) Distances between the Walker A motif (GxxGxGKS) on one NBD and the signature motif (LSGGQ) located on the opposing NBD (shown as d1 and d2 on S1 Fig). (D-F) Distances between NBDs and the intracellular helices at the NBD-TMD interface during triplicate simulations of the 3G5U, 4KSB and 4M1M system.

(TIFF)

**S9 Fig. Cholesterol distribution in the 3G5U system.** Cholesterol distribution around P-glycoprotein in the simulations based on the 3G5U model at (A) the beginning of the simulations and at 200 ns for each replica: (B) 3G5U #1, (C) 3G5U #2, and (D) 3G5U #3. The cholesterol molecules are shown in dark violet (upper leaflet) and light violet (lower leaflet) space filling representation. Only the TMD of the protein is shown for clarity.

(TIFF)

**S10 Fig. Cholesterol distribution in the 4KSB system.** Cholesterol distribution around P-glycoprotein in the simulations based on the 4KSB model at (A) the beginning of the simulations and at 200 ns for each replica: (B) 4KSB #1, (C) 4KSB #2, and (D) 4KSB #3. The cholesterol molecules are shown in dark violet (upper leaflet) and light violet (lower leaflet) space filling representation. Only the TMD of the protein is shown for clarity. Please note that the simulations of the 4KSB model involved a larger bilayer, the ratio of POPC:cholesterol is the same as in the 3G5U and 4M1M simulations.

(TIFF)

**S11 Fig. Cholesterol distribution in the 4M1M system.** Cholesterol distribution around P-glycoprotein in the simulations based on the 4M1M model at (A) the beginning of the simulations and at 200 ns for each replica: (B) 4M1M #1, (C) 4M1M #2, and (D) 4M1M #3. The cholesterol molecules are shown in dark violet (upper leaflet) and light violet (lower leaflet) space filling representation. Only the TMD of the protein is shown for clarity.

(TIFF)

**S1 Table. Subspace overlap: 3G5U.** Pairwise comparison of the 10 principal components obtained for each simulation based on the 3G5U model (3G5U#1-#3 PCs) and the concatenated (3G5U-4KSB-4M1M) trajectory containing all 9 simulations (cPCs). Subspace overlap was calculated as the root mean squared inner product (RMSIP) of every pair of the principal components, as defined in [79] and implemented in the Prody package [75].

(CSV)

**S2 Table. Subspace overlap: 4KSB.** Pairwise comparison of the 10 principal components obtained for each simulation based on the 4KSB model (4KSB#1-#3 PCs) and the concatenated (3G5U-4KSB-4M1M) trajectory containing all 9 simulations (cPCs). Subspace overlap was

calculated as the root mean squared inner product (RMSIP) of every pair of the principal components, as defined in [79] and implemented in the Prody package [75].  
(CSV)

**S3 Table. Subspace overlap: 4M1M.** Pairwise comparison of the 10 principal components obtained for each simulation based on the 4M1M model (4M1M#1-#3 PCs) and the concatenated (3G5U-4KSB-4M1M) trajectory containing all 9 simulations (cPCs). Subspace overlap was calculated as the root mean squared inner product (RMSIP) of every pair of the principal components, as defined in [79] and implemented in the Prody package [75].  
(CSV)

**S4 Table. Covariance overlap.** The covariance overlap between each pair of the 9 replicas and the concatenated trajectory calculated for the entire protein and the TMD coordinate subset only. The covariance overlap was calculated as defined in [80] and implemented in the Prody package [75]. This measure assesses the similarity of the spaces sampled by two different trajectories and it ranges from 0, corresponding to no similarity between the fluctuations, to 1, where the fluctuations are identical. The measured covariance overlap between different replicas range between 0.1-0.2.  
(CSV)

**S1 Video. 3G5U #1 trajectory—Front view.** The protein movement in the 200 ns trajectory based on the 3G5U model (replica #1). P-glycoprotein is shown in the *front* view depicting the  $\Lambda$ -conformation with TMD1 in the centre, NBD1 on the right side and NBD2 on the left. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 3G5U #2 and 3G5U #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].  
(MP4)

**S2 Video. 3G5U #1 trajectory—Side view.** The protein movement in the 200 ns trajectory based on the 3G5U model (replica #1). P-glycoprotein is shown in the *side* view with the TMD1 and NBD1 in front and with the closest lipids (purple) interacting with TM4/TM6 (left) and TM10/TM12 (right) portals. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 3G5U #2 and 3G5U #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].  
(MP4)

**S3 Video. 4KSB #1 trajectory—Front view.** The protein movement in the 200 ns trajectory based on the 4KSB model (replica #1). P-glycoprotein is shown in the *front* view depicting the  $\Lambda$ -conformation with TMD1 in the centre, NBD1 on the right side and NBD2 on the left. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 4KSB#2 and 4KSB #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].  
(MP4)

**S4 Video. 4KSB #1 trajectory—Side view.** The protein movement in the 200 ns trajectory based on the 4KSB model (replica #1). P-glycoprotein is shown in the *side* view with the TMD1 and NBD1 in front and with the closest lipids (purple) interacting with TM4/TM6 (left) and TM10/TM12 (right) portals. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 4KSB #2 and 4KSB #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].  
(MP4)



**S5 Video. 4M1M #1 trajectory—Front view.** The protein movement in the 200 ns trajectory based on the 4M1M model (replica #1). P-glycoprotein is shown in the *front* view depicting the  $\Lambda$ -conformation with TMD1 in the centre, NBD1 on the right side and NBD2 on the left. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 4M1M#2 and 4M1M #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].

(MP4)

**S6 Video. 4M1M #1 trajectory—Side view.** The protein movement in the 200 ns trajectory based on the 4M1M model (replica #1). P-glycoprotein is shown in the *side* view with the TMD1 and NBD1 in front and with the closest lipids (purple) interacting with TM4/TM6 (left) and TM10/TM12 (right) portals. Frames shown in the video were taken every 100 ps. Corresponding videos for the replicas 4M1M #2 and 4M1M #3 are provided on Figshare: <https://doi.org/10.6084/m9.figshare.4806544.v4> [78].

(MP4)

## Acknowledgments

This work was supported by grants from the Australian Research Council (DE120101550, DP130102153), the National Health and Medical Research Council (APP1049685), and with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

## Author Contributions

**Conceptualization:** Karmen Condic-Jurkic, Alan E. Mark, Megan L. O'Mara.

**Data curation:** Nandhitha Subramanian.

**Formal analysis:** Nandhitha Subramanian.

**Funding acquisition:** Alan E. Mark, Megan L. O'Mara.

**Investigation:** Karmen Condic-Jurkic, Nandhitha Subramanian.

**Methodology:** Karmen Condic-Jurkic.

**Project administration:** Megan L. O'Mara.

**Resources:** Megan L. O'Mara.

**Supervision:** Alan E. Mark, Megan L. O'Mara.

**Writing – original draft:** Karmen Condic-Jurkic, Megan L. O'Mara.

**Writing – review & editing:** Karmen Condic-Jurkic, Alan E. Mark, Megan L. O'Mara.

## References

1. Holland IB, Blight MA. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *Journal of Molecular Biology*. 1999; 293(2):381–99. <https://doi.org/10.1006/jmbi.1999.2993> PMID: 10529352
2. Davidson AL. Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters. *Journal of Bacteriology*. 2002; 184(5):1225–1233. <https://doi.org/10.1128/JB.184.5.1225-1233.2002> PMID: 11844750
3. Higgins CF. ABC transporters: From microorganisms to man. *Annual Review of Cell Biology*. 1992; 8(1):67–113. <https://doi.org/10.1146/annurev.cb.08.110192.000435> PMID: 1282354

4. Gottesman MM, Fojo T, Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature Reviews Cancer*. 2002; 2(1):48–58. <https://doi.org/10.1038/nrc706> PMID: 11902585
5. Sharom FJ. ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics*. 2008; 9(1):105–127. <https://doi.org/10.2217/14622416.9.1.105> PMID: 18154452
6. Callaghan R, Luk F, Bebawy M. Inhibition of the Multidrug Resistance P-Glycoprotein: Time for a Change of Strategy? *Drug Metabolism and Disposition*. 2014; 42(4):623–631. <https://doi.org/10.1124/dmd.113.056176> PMID: 24492893
7. Gutmann DAP, Ward A, Urbatsch IL, Chang G, van Veen HW. Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. *Trends in Biochemical Sciences*. 2010; 35(1):36–42. <https://doi.org/10.1016/j.tibs.2009.07.009> PMID: 19819701
8. Chufan EE, Sim HM, Ambudkar SV. Molecular basis of the polyspecificity of P-glycoprotein (ABCB1). Recent biochemical and structural studies. *Advances in Cancer Research*. 2015; 125:71–96. <https://doi.org/10.1016/bs.acr.2014.10.003> PMID: 25640267
9. Dawson RJP, Locher KP. Structure of a bacterial multidrug ABC transporter. *Nature*. 2006; 443(7108):180–185. <https://doi.org/10.1038/nature05155> PMID: 16943773
10. Dawson RJP, Locher KP. Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP. *FEBS Letters*. 2007; 581(5):935–938. <https://doi.org/10.1016/j.febslet.2007.01.073> PMID: 17303126
11. Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences*. 2007; 104(48):19005–19010. <https://doi.org/10.1073/pnas.0709388104>
12. Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. *Nature Structural & Molecular Biology*. 2012; 19(4):395–402. <https://doi.org/10.1038/nsmb.2267>
13. Hohl M, Hurlimann LM, Böhm S, Schoppe J, Grütter MG, Bordignon E, et al. Structural basis for allosteric cross-talk between the asymmetric nucleotide binding sites of a heterodimeric ABC exporter. *Proceedings of the National Academy of Sciences*. 2014; 111(30):11025–11030. <https://doi.org/10.1073/pnas.1400485111>
14. Kodan A, Yamaguchi T, Nakatsu T, Sakiyama K, Hipolito CJ, Fujioka A, et al. Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog. *Proceedings of the National Academy of Sciences*. 2014; 111(11):4049–4054. <https://doi.org/10.1073/pnas.1321562111>
15. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*. 2009; 323(5922):1718 LP–1722. <https://doi.org/10.1126/science.1168750>
16. Li J, Jaimes KF, Aller SG. Refined structures of mouse P-glycoprotein. *Protein Science*. 2014; 23(1):34–46. <https://doi.org/10.1002/pro.2387> PMID: 24155053
17. Ward AB, Szewczyk P, Grimard V, Lee CW, Martinez L, Doshi R, et al. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *Proceedings of the National Academy of Sciences*. 2013; 110(33):13386–13391. <https://doi.org/10.1073/pnas.1309275110>
18. Szewczyk P, Tao H, McGrath AP, Villaluz M, Rees SD, Lee SC, et al. Snapshots of ligand entry, malleable binding and induced helical movement in P-glycoprotein. *Acta Crystallographica Section D: Biological Crystallography*. 2015; 71(3):732–741. <https://doi.org/10.1107/S1399004715000978>
19. Jin MS, Oldham ML, Zhang Q, Chen J. Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature*. 2012; 490(7421):566–569. <https://doi.org/10.1038/nature11448> PMID: 23000902
20. Ambudkar SV, Kim IW, Sauna ZE. The power of the pump: Mechanisms of action of P-glycoprotein (ABCB1). *European Journal of Pharmaceutical Sciences*. 2006; 27(5):392–400. <https://doi.org/10.1016/j.ejps.2005.10.010> PMID: 16352426
21. Bill RM, Henderson PJF, Iwata S, Kunji ERS, Michel H, Neutze R, et al. Overcoming barriers to membrane protein structure determination. *Nature Biotechnology*. 2011; 29(4):335–340. <https://doi.org/10.1038/nbt.1833> PMID: 21478852
22. Privé GG. Detergents for the stabilization and crystallization of membrane proteins. *Methods*. 2007; 41(4):388–397. <https://doi.org/10.1016/j.ymeth.2007.01.007> PMID: 17367711
23. Columbus L, Lipfert J, Jambunathan K, Fox DA, Sim AYL, Doniach S, et al. Mixing and matching detergents for membrane protein NMR structure determination. *Biophysical Journal*. 2009; 96(3):195a. <https://doi.org/10.1016/j.bpj.2008.12.1044>

24. Tate CG. Practical considerations of membrane protein instability during purification and crystallisation. *Methods in Molecular Biology* (Clifton, NJ). 2010; 601:187–203. [https://doi.org/10.1007/978-1-60761-344-2\\_12](https://doi.org/10.1007/978-1-60761-344-2_12)
25. Eyal E, Gerzon S, Potapov V, Edelman M, Sobolev V. The limit of accuracy of protein modeling: Influence of crystal packing on protein structure. *Journal of Molecular Biology*. 2005; 351(2):431–442. <https://doi.org/10.1016/j.jmb.2005.05.066> PMID: 16005885
26. Huang DB, Ainsworth CF, Stevens FJ, Schiffer M. Three quaternary structures for a single protein. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93(14):7017–21. <https://doi.org/10.1073/pnas.93.14.7017> PMID: 8692936
27. van Gunsteren WF, Dolenc J, Mark AE. Molecular simulation as an aid to experimentalists. *Current Opinion in Structural Biology*. 2008; 18(2):149–153. <https://doi.org/10.1016/j.sbi.2007.12.007> PMID: 18280138
28. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*. 2012; 41(1):429–452. <https://doi.org/10.1146/annurev-biophys-042910-155245> PMID: 22577825
29. O'Mara ML, Mark AE. The effect of environment on the structure of a membrane protein: P-glycoprotein under physiological conditions. *Journal of Chemical Theory and Computation*. 2012; 8(10):3964–3976. <https://doi.org/10.1021/ct300254y> PMID: 26593033
30. Ferreira RJ, Ferreira MJU, Dos Santos DJVA. Insights on P-glycoproteins efflux mechanism obtained by molecular dynamics simulations. *Journal of Chemical Theory and Computation*. 2012; 8(6): 1853–1864. <https://doi.org/10.1021/ct300083m> PMID: 26593820
31. Ferreira RJ, Ferreira MJU, Dos Santos DJVA. Assessing the stabilization of P-glycoprotein's nucleotide-binding domains by the linker, using molecular dynamics. *Molecular Informatics*. 2013; 32(5–6): 529–540. <https://doi.org/10.1002/minf.201200175> PMID: 27481670
32. Ma J, Biggin PC. Substrate versus inhibitor dynamics of P-glycoprotein. *Proteins: Structure, Function and Bioinformatics*. 2013; 81(9):1653–1668. <https://doi.org/10.1002/prot.24324>
33. Pan L, Aller SG. Equilibrated atomic models of outward-facing P-glycoprotein and effect of ATP binding on structural dynamics. *Scientific Reports*. 2015; 5(1):7880. <https://doi.org/10.1038/srep07880> PMID: 25600711
34. Domiciceva L, Biggin PC. Homology modelling of human P-glycoprotein. *Biochemical Society Transactions*. 2015; 43(5):952–958. <https://doi.org/10.1042/BST20150125> PMID: 26517909
35. Domiciceva L, Paramo T, Biggin PC. In silico characterization of P-glycoprotein substrate entrance pathways. *Biophysical Journal*. 2016; 110(3):141a. <https://doi.org/10.1016/j.bpj.2015.11.800>
36. Jara GE, Vera DMA, Pierini AB. Binding of modulators to mouse and human multidrug resistance P-glycoprotein. A computational study. *Journal of Molecular Graphics and Modelling*. 2013; 46:10–21. <https://doi.org/10.1016/j.jmgm.2013.09.001> PMID: 24095875
37. Prajapati R, Singh U, Patil A, Khomane KS, Bagul P, Bansal AK, et al. In silico model for P-glycoprotein substrate prediction: Insights from molecular dynamics and in vitro studies. *Journal of Computer-Aided Molecular Design*. 2013; 27(4):347–363. <https://doi.org/10.1007/s10822-013-9650-x> PMID: 23612916
38. O'Mara ML, Mark AE. Structural characterization of two metastable ATP-bound states of P-glycoprotein. *PLoS ONE*. 2014; 9(3):e91916. <https://doi.org/10.1371/journal.pone.0091916> PMID: 24632881
39. Subramanian N, Condic-Jurkic K, Mark AE, O'Mara ML. Identification of possible binding sites for morphine and nicardipine on the multidrug transporter P-glycoprotein using umbrella sampling techniques. *Journal of Chemical Information and Modeling*. 2015; 55(6):1202–1217. <https://doi.org/10.1021/ci5007382> PMID: 25938863
40. Ferreira RJ, Ferreira MJU, Dos Santos DJVA. Do drugs have access to the P-glycoprotein drug-binding pocket through gates? *Journal of Chemical Theory and Computation*. 2015; 11(10):4525–4529. <https://doi.org/10.1021/acs.jctc.5b00652> PMID: 26574244
41. Chang SY, Liu FF, Dong XY, Sun Y. Molecular insight into conformational transmission of human P-glycoprotein. *The Journal of Chemical Physics*. 2013; 139(22):225102. <https://doi.org/10.1063/1.4832740> PMID: 24329094
42. Prajapati R, Sangamwar AT. Translocation mechanism of P-glycoprotein and conformational changes occurring at drug-binding site: Insights from multi-targeted molecular dynamics. *Biochimica et Biophysica Acta (BBA)—Biomembranes*. 2014; 1838(11):2882–2898. <https://doi.org/10.1016/j.bbamem.2014.07.018>
43. Zou P, Mchaourab HS. Alternating access of the putative substrate-binding chamber in the ABC transporter MsbA. *Journal of Molecular Biology*. 2009; 393(3):574–585. <https://doi.org/10.1016/j.jmb.2009.08.051> PMID: 19715704

44. Verhalen B, Ernst S, Börsch M, Wilkens S. Dynamic ligand-induced conformational rearrangements in P-glycoprotein as probed by fluorescence resonance energy transfer spectroscopy. *Journal of Biological Chemistry*. 2012; 287(2):1112–1127. <https://doi.org/10.1074/jbc.M111.301192> PMID: 22086917
45. Marcoux J, Wang SC, Politis A, Reading E, Ma J, Biggin PC, et al. Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. *Proceedings of the National Academy of Sciences*. 2013; 110(24):9704–9709. <https://doi.org/10.1073/pnas.1303888110>
46. Wen PC, Verhalen B, Wilkens S, Mchaourab HS, Tajkhorshid E. On the origin of large flexibility of P-glycoprotein in the inward-facing state. *Journal of Biological Chemistry*. 2013; 288(26):19211–19220. <https://doi.org/10.1074/jbc.M113.450114> PMID: 23658020
47. Zarrabi N, Ernst S, Verhalen B, Wilkens S, Börsch M. Analyzing conformational dynamics of single P-glycoprotein transporters by Förster resonance energy transfer using hidden Markov models. *Methods*. 2014; 66(2):168–179. <https://doi.org/10.1016/j.ymeth.2013.07.026> PMID: 23891547
48. Ledwith KV, Gibbs ME, Barnes RW, Roberts AG. Cooperativity between verapamil and ATP bound to the efflux transporter P-glycoprotein. *Biochemical Pharmacology*. 2016; 118:96–108. <https://doi.org/10.1016/j.bcp.2016.08.013> PMID: 27531061
49. Verhalen B, Dastvan R, Thangapandian S, Peskova Y, Koteiche HA, Nakamoto RK, et al. Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature*. 2017; 543(7647):738–741. <https://doi.org/10.1038/nature21414> PMID: 28289287
50. Esser L, Zhou F, Pluchino KM, Shiloach J, Ma J, Tang Wk, et al. Structures of the multidrug transporter P-glycoprotein reveal asymmetric ATP binding and the mechanism of polyspecificity. *Journal of Biological Chemistry*. 2017; 292(2):446–461. <https://doi.org/10.1074/jbc.M116.755884> PMID: 27864369
51. Mackay JP, Landsberg MJ, Whitten AE, Bond CS. Whaddaya Know: A guide to uncertainty and subjectivity in structural biology. *Trends in Biochemical Sciences*. 2017; 42(2):155–167. <https://doi.org/10.1016/j.tibs.2016.11.002> PMID: 28089412
52. Law RJ, Capener C, Baaden M, Bond PJ, Campbell J, Patargias G, et al. Membrane protein structure quality in molecular dynamics simulation. *Journal of Molecular Graphics and Modelling*. 2005; 24(2):157–165. <https://doi.org/10.1016/j.jmgm.2005.05.006> PMID: 16102990
53. Ivetac A, Sansom MSP. Molecular dynamics simulations and membrane protein structure quality. *European Biophysics Journal*. 2008; 37(4):403–9. <https://doi.org/10.1007/s00249-007-0225-4> PMID: 17960373
54. Pan AC, Weinreich TM, Piana S, Shaw DE. Demonstrating an order-of-magnitude sampling enhancement in molecular dynamics simulations of complex protein systems. *Journal of Chemical Theory and Computation*. 2016; 12(3):1360–1367. <https://doi.org/10.1021/acs.jctc.5b00913> PMID: 26866996
55. Dahl ACE, Chavent M, Sansom MSP. Bendix: Intuitive helix geometry analysis and abstraction. *Bioinformatics*. 2012; 28(16):2193–2194. <https://doi.org/10.1093/bioinformatics/bts357> PMID: 22730430
56. Zhang J, Li D, Sun T, Liang L, Wang Q. Interaction of P-glycoprotein with anti-tumor drugs: the site, gate and pathway. *Soft Matter*. 2015; 11(33):6633–6641. <https://doi.org/10.1039/C5SM01028D> PMID: 26205623
57. Sjarheyeva A, Liu R, Sharom FJ. Characterization of an asymmetric occluded state of P-glycoprotein with two bound nucleotides: Implications for catalysis. *Journal of Biological Chemistry*. 2010; 285(10):7575–7586. <https://doi.org/10.1074/jbc.M109.047290> PMID: 20061384
58. Urbatsch IL, Gimi K, Wilke-Mounts S, Lerner-Marmarosh N, Rousseau ME, Gros P, et al. Cysteines 431 and 1074 are responsible for inhibitory disulfide cross-linking between the two nucleotide-binding sites in human P-glycoprotein. *Journal of Biological Chemistry*. 2001; 276(29):26980–26987. <https://doi.org/10.1074/jbc.M010829200> PMID: 11356825
59. Lee JY, Urbatsch IL, Senior AE, Wilkens S. Projection structure of P-glycoprotein by electron microscopy. Evidence for a closed conformation of the nucleotide binding domains. *Journal of Biological Chemistry*. 2002; 277(42):40125–40131. <https://doi.org/10.1074/jbc.M206871200> PMID: 12163504
60. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature*. 2015; 524(7566):433–438. <https://doi.org/10.1038/nature14953> PMID: 26266984
61. Loo TW, Clarke DM. The transmission interfaces contribute asymmetrically to the assembly and activity of human P-glycoprotein. *Journal of Biological Chemistry*. 2015; 290(27):16954–16963. <https://doi.org/10.1074/jbc.M115.652602> PMID: 25987565
62. McCormick JW, Vogel PD, Wise JG. Multiple drug transport pathways through human P-glycoprotein. *Biochemistry*. 2015; 54(28):4374–4390. <https://doi.org/10.1021/acs.biochem.5b00018> PMID: 26125482

63. Sato T, Kodan A, Kimura Y, Ueda K, Nakatsu T, Kato H. Functional role of the linker region in purified human P-glycoprotein. *FEBS Journal*. 2009; 276(13):3504–3516. <https://doi.org/10.1111/j.1742-4658.2009.07072.x> PMID: 19490125
64. Poger D, van Gunsteren WF, Mark AE. A new force field for simulating phosphatidylcholine bilayers. *Journal of Computational Chemistry*. 2010; 31(6):1117–1125. <https://doi.org/10.1002/jcc.21396> PMID: 19827145
65. Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM. A consistent empirical potential for water–protein interactions. *Biopolymers*. 1984; 23(8):1513–1518. <https://doi.org/10.1002/bip.360230807>
66. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*. 2001; 7(8):306–317. <https://doi.org/10.1007/s008940100045>
67. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*. 2011; 40(7): 843–856. <https://doi.org/10.1007/s00249-011-0700-9> PMID: 21533652
68. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics*. 1995; 102(13):5451–5459. <https://doi.org/10.1063/1.469273>
69. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*. 1997; 18(12):1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H)
70. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992; 13(8):952–962. <https://doi.org/10.1002/jcc.540130805>
71. Feenstra KA, Hess B, Berendsen HJC. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*. 1999; 20(8):786–798. [https://doi.org/10.1002/\(SICI\)1096-987X\(199906\)20:8%3C786::AID-JCC5%3E3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(199906)20:8%3C786::AID-JCC5%3E3.0.CO;2-B)
72. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*. 1984; 81(8):3684–3690. <https://doi.org/10.1063/1.448118>
73. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*. 2015; 109(8): 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015> PMID: 26488642
74. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*. 1994; 235(2):625–634. <https://doi.org/10.1006/jmbi.1994.1017> PMID: 8289285
75. Bakan A, Meireles LM, Bahar I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011; 27(11):1575–1577. <https://doi.org/10.1093/bioinformatics/btr168> PMID: 21471012
76. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14(1):33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
77. MATLAB. version 8.6 (R2015b). Natick, Massachusetts: The MathWorks Inc.; 2015.
78. Condic-Jurkic K. MD simulations of P-glycoprotein started from three different crystal structures: 3G5U, 4M1M and 4KSB. Figshare. 2017;
79. Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*. 1999; 36(4):419–424. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990901\)36:4%3C419::AID-PROT5%3E3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0134(19990901)36:4%3C419::AID-PROT5%3E3.0.CO;2-U)
80. Hess B. Convergence of sampling in protein simulations. *Phys Rev E*. 2002; 65:031910. <https://doi.org/10.1103/PhysRevE.65.031910>