# Multiscale molecular dynamics simulations of lipid interactions with P-glycoprotein in a complex membrane

Laura Domicevica[1], Heidi Koldsø[1,‡] and Philip C. Biggin[1*]

[1]Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom.

[‡]Present address: D. E. Shaw Research, 120 West 45th Street, 39th Floor, New York, New York 10036, United States of America.

Running Title: "P-glycoprotein in a complex membrane"

*To whom correspondence should be addressed.

Email: philip.biggin@bioch.ox.ac.uk

Tel. +44 1865 613305

## ABSTRACT

P-glycoprotein (P-gp) can transport a wide range of very different hydrophobic organic molecules across the membrane. Its ability to extrude molecules from the cell creates delivery problems for drugs that target proteins in the central nervous system (CNS) and also causes drug-resistance in many forms of cancer. Whether a drug will be susceptible to export by P-gp is difficult to predict and currently this is usually assessed with empirical and/or animal models. Thus, there is a need to better understand how P-gp works at the molecular level in order to fulfil the 3Rs: Refinement, reduction and replacement of animals in research. As structural information increasingly becomes available, our understanding at the molecular level improves. Proteins like P-gp are however very dynamic entities and thus one of the most appropriate ways to study them is with molecular dynamics simulations, especially as this can capture the influence of the surrounding environment. Recent parameterization developments have meant that it is now possible to simulate lipid bilayers that more closely resemble *in vivo* membranes in terms of their composition. In this report we construct a complex lipid bilayer that mimics the composition of brain epithelial cells and examine the interactions of it with P-gp. We find that the negatively charged phosphatidylserine lipids in the inner leaflet of the membrane tend to form an annulus around P-gp. We also observed the interaction of cholesterol with three distinct areas of the P-gp. Potential of mean force (PMF) calculations suggest that a crevice between transmembrane helices 10 and 12 has particularly favourable interaction energy for cholesterol.

### *Keywords*

Molecular dynamics, P-glycoprotein, ABCB1, bilayer, efflux pump, coarse-grained, membrane, lipids.

### Abbreviations

ABC, ATP-binding cassette; CNS, central nervous system; P-gp, P-glycoprotein; POPC, 1-palmitoyl-2-oleoylphosphatidylcholine; PS, phosphatidylserine; SD, standard deviation;

## 1. Introduction

Members of the ATP-Binding Cassette (ABC) superfamily are pumps that deliver their substrates across the plasma membrane by coupling ATP binding and hydrolysis to transport. In humans, 48 ABC proteins have been identified [1], but the one that has received the most attention by far is the multidrug transporter P-glycoprotein (P-gp) or ABCB1. P-gp can transport a wide variety of hydrophobic organic molecules across the membrane including vinca alkaloids, anthracyclines and taxanes [2]. While its ability to extrude cytotoxic drugs from the cell is the main cause for P-gp-mediated drug-resistance in cancer, it also forms an integral part of the blood brain barrier by extruding compounds from the central nervous system (CNS). P-gp can therefore, also reduce the bioavailability of potential treatments for CNS-based targets. For these reasons, the susceptibility of compounds to P-gp transport is often assessed during the development of new drugs and the methods to do this range from empirical models through to animal testing. There is thus a strong incentive to understand the processes involved in P-gp behaviour in order to promote the 3Rs agenda: Refinement, reduction and replacement of animals in research.

The plasma membrane of human cells is a complex assembly of proteins and lipids that protect the cell from the outer environment. It is also the focal point for a variety of signalling and transport processes. Biological membranes consist of many different types of lipid that are distributed asymmetrically across the two leaflets of the bilayer [3]. Typical molecular dynamics (MD) simulation studies of membrane proteins use very simplistic bilayers that usually consist of a single lipid species [4]. Consequently, the information these studies can provide on protein-lipid interactions is somewhat limited and in some case at least, may not reflect accurately the situation *in vivo*. In the case of P-gp, its ATPase activity can be influenced by the phospholipid and cholesterol content of the lipid bilayer [5]. Furthermore, the surrounding lipid environment, especially the addition of cholesterol [6], can also affect the structural stability of P-gp.

The increasing interest in protein-lipid interactions has spawned the development of lipid parameters for various force-fields [7-9]. Concomitantly, various efforts have been made to adopt so-called multi-scale approaches [7, 10-12], which take advantage of the increased sampling made possible through the use of coarse graining combined with the ability to convert back to an atomistic representation when detail is required [13, 14].

In this study P-gp was inserted in a complex asymmetric bilayer containing the main lipid headgroup types present in the human brain epithelial cell membranes [15]. After characterising the bilayer in the absence of protein through numerous multi-scale MD simulations, we investigated P-gp interactions with the lipids present in the bilayer and the effect of these interactions on the stability of the protein. The interactions with sterols were studied in greater detail, as there is much uncertainty about the nature and importance of these particular interactions [16, 17]. The simulations revealed several potential cholesterol binding sites on the surface of P-gp. The

strength of these interactions was investigated further by calculating the potential of mean force (PMF) via umbrella sampling (US) simulations. These calculations suggest that a crevice between transmembrane helices 10 and 12 may form a cholesterol-binding pocket.

## 2. Methods

### *2.1. Bilayer setup*

An initial bilayer of 500 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) lipid molecules was created by a self-assembly process in a 25 ns long coarse-grained simulation. Coarse-grained (CG) topologies under the MARTINI v2.2 coarse-grain forcefield [18] use a four-to-one mapping that typically represent 4 atomistic particles as one "bead'. The CG bilayer was then converted to a complex asymmetric bilayer by "editing" the POPC bilayer [7, 9]. This process converts the coordinates of a random POPC molecule to coordinates of another phospholipid with the same or shorter chain length as POPC, followed by re-labelling and particle type alteration. Cholesterol was superimposed on the headgroup and 3 of the hydrophobic tail beads of a POPC molecule. The final composition is shown in **Table 1**. Since the conversion provides different lipid coordinates after every run, this procedure was done 3 times to make three independent repeats. The box was then solvated with MARTINI water beads (each bead represents 4 water molecules). Since negative charge was introduced in the system by addition of POPS lipids, $Na^+$ ions were used to neutralise the system. Additional $Na^+$ and $Cl^-$ ions were added to the system to reach the physiological conditions of 150 mM NaCl. After conversion, 10 ns equilibration run for each repeat was done with protein backbone beads restrained with a force constant of 1000 kJ/mol. Coarse-grained production runs were 10 μs long.

**Table 1. The composition of the complex* asymmetric bilayer.**

| Lipid | This study | | | Siakotos *et al.* *(%)* | Tewes *et al.* *(%)* |
|---|---|---|---|---|---|
| | Outer leaflet (%) | Inner leaflet (%) | Total bilayer (%) | | |
| POPC | 45.0 (112) | 25.0 (62) | 35.0 | 28.7 | 33.2 |
| POPE | 15.0 (37) | 35.0 (89) | 25.0 | 22.8 | 25.2 |
| POPS | 0.0 (0) | 20.0 (51) | 10.0 | 6.8 | 10.7 |
| Sphingomyelin | 20.0 (49) | 0.0 (0) | 10.0 | 33.4 | 17.0 |
| Cholesterol | 20.0 (50) | 20.0 (50) | 20.0 | 0.0 | 20.0 |
| Phosphatidylinositol | 0.0 | 0.0 | 0.0† | 5.5 | 4.8 |
| Phosphatidylinositol + phosphatidic acid | 0.0 | 0.0 | 0.0† | 0.0 | 4.8 |

Absolute numbers of each species are given in brackets. *The complexity of the bilayer reflects the composition of blood epithelial cell membrane [15], which cites two papers ([19, 20]). †These lipids were omitted from the simulations here as no parameters available. The asymmetry was determined by trends observable in mammalian cells [3].

The coarse-grained topologies were then converted to atomistic resolution using the Backwards tool from the MARTINI toolkit [13]. The resulting atomistic system was energy minimised using the steepest descents method and allowed to equilibrate for 1 ns with the Berendsen barostat [21]. The topology files for lipids in the MARTINI and AMBER ff99SB-ILDN force-fields were acquired from the lipidbook repository [22].

### *2.2. P-glycoprotein-lipid system setup*

All simulations were run using the MARTINI coarse grained force field for lipids [18] and proteins [23]. The protein was converted to a coarse grained resolution using the martinize.py script which creates input files for proteins compatible with the Gromacs package and MARTINI forcefield [24]. The protein structure used in the coarse-grained simulations is an equilibrated conformation of P-gp that was obtained from a 100 ns simulation of an apo P-glycoprotein homology model in POPC bilayer as reported previously [25]. The coarse-grained protein was inserted in to a lipid bilayer consisting of 500 POPC molecules using the self-assembly method [14]. After 25 ns of self-assembly simulation, the POPC bilayer was formed completely around the protein. It was then converted to a complex asymmetric bilayer [7, 9]. Since the conversion provides different lipid coordinates after every run, this procedure was done 10 times to make 10 independent repeats. After conversion, 100 ns equilibration run for each repeat was done with protein backbone beads position-restrained. The production runs were 10 μs long.

Three of the resulting topologies were converted to atomistic resolution using the “Backwards” back-mapping tool [13] for lipid molecules whilst the cg2at approach [14] was used to convert the protein back to an atomistic representation. The Alchembed method was used to reduce potential steric clashes [26]. The system was then re-solvated and ions were added to neutralize the system and achieve a 150 mM concentration of NaCl. After steepest descent energy minimisation, a 1 ns long equilibration simulation was run with heavy protein atoms restrained. After the equilibration, production runs were 100 ns long.

### *2.3. Coarse-grained molecular dynamics*

All simulations (including all-atom) were run with GROMACS 4.6.x simulation package [27]. The Berendsen thermostat and barostat [21] were used for temperature and pressure coupling. Simulations were run at temperature of 310K and pressure of 1 bar. The LINCS algorithm [28] was used to constrain bond lengths with a timestep of 20 fs. The long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method [29]. During the equilibration process the protein backbone beads were constrained. The 100 ns equilibration period was continued with 10 μs production runs of MD simulations.

### *2.4. All-atom molecular dynamics*

The production runs of all-atom MD simulations used the V-rescale thermostat

[30] and the Parrinello-Rahman barostat [31] to keep the system at constant temperature and pressure respectively. The temperature was kept at 310K and the pressure was 1 bar. The AMBER ff99SB-ILDN forcefield [32] was used to describe protein and the Slipids forcefield [33] was used for the lipid molecules. The timestep was 2 fs. The LINCS algorithm [28] was used to constrain bond lengths and the long-range electrostatic interactions were calculated with the PME method [29].

*2.5. Umbrella sampling simulations and PMF calculations*

A frame with cholesterol bound to the site of interest was extracted from a 10 μs coarse-grained production run. The protein-lipid complex was inserted in to a POPC bilayer using a 50 ns long self-assembly simulation. We performed these calculations in a POPC bilayer rather than a more complex system to facilitate a convergence of the PMFs. To do the same calculations in a complex bilayer was beyond our current resources. P-gp and cholesterol were restrained during the self-assembly run using restraints of 1000 kJ/mol/nm$^2$. Self-assembly was followed by 10 ns equilibration run during which protein was restrained with 400 kJ/mol/nm$^2$ restraints in the XY plane to prevent rotation and translation while cholesterol and POPC lipids were unrestrained. Configurations along the collective variable (distance between the cholesterol headgroup and the center of mass of residues within 7 Å of cholesterol when bound) were created using steered molecular dynamics (SMD) simulations. These configurations ranged from a lipid-bound to a lipid-unbound state where cholesterol is present in the bulk lipid bilayer and exhibits a plateau in terms of the free energy (as expected). Cholesterol was pulled at a rate of 1 Å/ns over a distance of 30 Å with a force constant of 1000 kJ/mol/nm$^2$. The protein was restrained from positional and rotational movements by placing positional restraints on Ser237, Ala354, Ala883 and Ala1006 backbone beads on the X and Y axis (400 kJ/mol/nm$^2$). Additionally, a 100 kJ/mol/nm$^2$ restraint was placed on the cholesterol headgroup bead (ROH) in the direction perpendicular to the collective variable and the bilayer normal [34].

Umbrella sampling simulations were run with conformations obtained from steered molecular dynamics (SMD) simulations. The umbrella sampling windows were spaced asymmetrically with 0.5 Å spacing at low values of the collective variable (0 - 10 Å) and at a 1 Å spacing with larger collective variable values (distances over 10 Å). Each window was run for 500 ns. The protein-lipid separation was maintained by applying a harmonic umbrella potential to the collective variable used in the SMD simulations. The same positional restraints used in the SMD simulations were also applied in the umbrella sampling simulations.

The g_wham tool – the GROMACS implementation of the Weighted Histogram Analysis Method (WHAM) was used to combine and unbias the umbrella potentials. Bootstrapping was used to estimate the error on free energy profiles. Convergence was analysed by comparing free energy profiles from successive 100 ns increases, as well as non-overlapping 100 ns sections of simulation time. The first 100 ns of each window were discarded

from the final free energy profile as an equilibration period.

### *2.6. Lipid analysis and data visualization*

The diffusion of lipids was measured using the "Diffusion analysis in MD simulations" Python script [35]. The mean square displacement (MSD) was calculated separately for each lipid species present in the complex bilayer. To improve the speed of the calculations only every 100$^{th}$ frame was analysed. The window sizes for MSD calculation are measured in frames, not time values, as the simulation output frequencies and time steps may vary. The window sizes were set up in such a way that the separation between values was small at the start and larger as the values increased. The highest window size was half of the simulation frames available for analysis. Linear diffusion coefficients were calculated from the obtained data (MSD). Additionally, the MSD was also fitted to an anomalous diffusion model (Eq. 1) to investigate the nature of diffusion of the lipids [36]. If the scaling exponent (α) of the non-linear fit is 1, the lipids diffuse via a random walk, however, if the alpha value is lower than 1, the system is undergoing sub-diffusion.

$$MSD = ND_{\alpha} t^{\alpha} \qquad \text{Eq 1.}$$

The distribution of lipids and radial distribution were determined by GROMACS tools g_density and g_rdf respectively. Lipid densities were analysed and displayed with VMD VolMap plugin. The cholesterol binding sites were determined using MDAnalysis [37] and in-house Python scripts. Graphs were plotted with Xmgrace and Matplotlib and Visual Molecular Dynamics (VMD) [38] was used to view trajectories, analyse lipid densities and create cartoon representations.

## 3. Results

### *3.1. Characterisation of the complex bilayer*

We began by constructing a complex lipid bilayer system (see Methods) designed to mimic as closely as possible the composition of membranes in the blood-brain barrier [15] whilst at the same time considering the availability of parameters at both the coarse-grained and atomistic levels. As a result, the complex bilayer used here consists of several phospholipid species, including anionic lipids, as well as sphingomyelin and cholesterol (see **Table 1**). The identity and total percentage of lipid headgroups closely replicated the experimentally determined composition of human brain epithelial cell membrane [15]. The lipid distribution between leaflets was asymmetric with phosphatidylcholine (PC) and sphingomyelin (SM) appearing mostly in the outer leaflet and phosphatidylethalomine (PE) and the negatively charged phosphatidylserine (PS) headgroups being the dominant species of the inner leaflet consistent with experimental data [15]. Hydrocarbon tails of the phospholipids (18:1, 16:0) and sphingomyelin (15:1, 16:0) were of similar length and composed of one saturated and one unsaturated fatty acid.

The complex bilayer was investigated first in the absence of P-gp to check compare its behaviour under both a coarse-grained or atomistic regimen was consistent with previously reported systems of similar complexity [8, 39]. Coarse-grained simulations of the complex asymmetric bilayer in a box of water were run in triplicate for 10 μs, while three repeats of the atomistic simulations were run for 100 ns. The difference in resolution allows for comparison between methods and with available experimental and simulation data.

Anomalous diffusion (sub-diffusion) was observed for all lipid species in both coarse grained and atomistic simulations (**Fig. 1**). Experimental work on the diffusion of dye molecules in lipid bilayers also shows anomalous diffusion [40]. In coarse-grained simulations the diffusion constant of all lipids was 2873 ± 360 Å$^2$/μs (standard deviation, SD) with cholesterol diffusing slightly faster than the phospholipids (3085 ± 316 Å$^2$/μs). These diffusion constants are within close agreement with coarse-grained simulations of a large patch of idealized mammalian plasma membrane [8]. The lipids in atomistic simulations showed a lower diffusion constant (1826 ± 1060 Å$^2$/μs). Similar trends and diffusion constants have been observed in atomistic simulations of various phospholipid bilayers with 20% cholesterol [8].

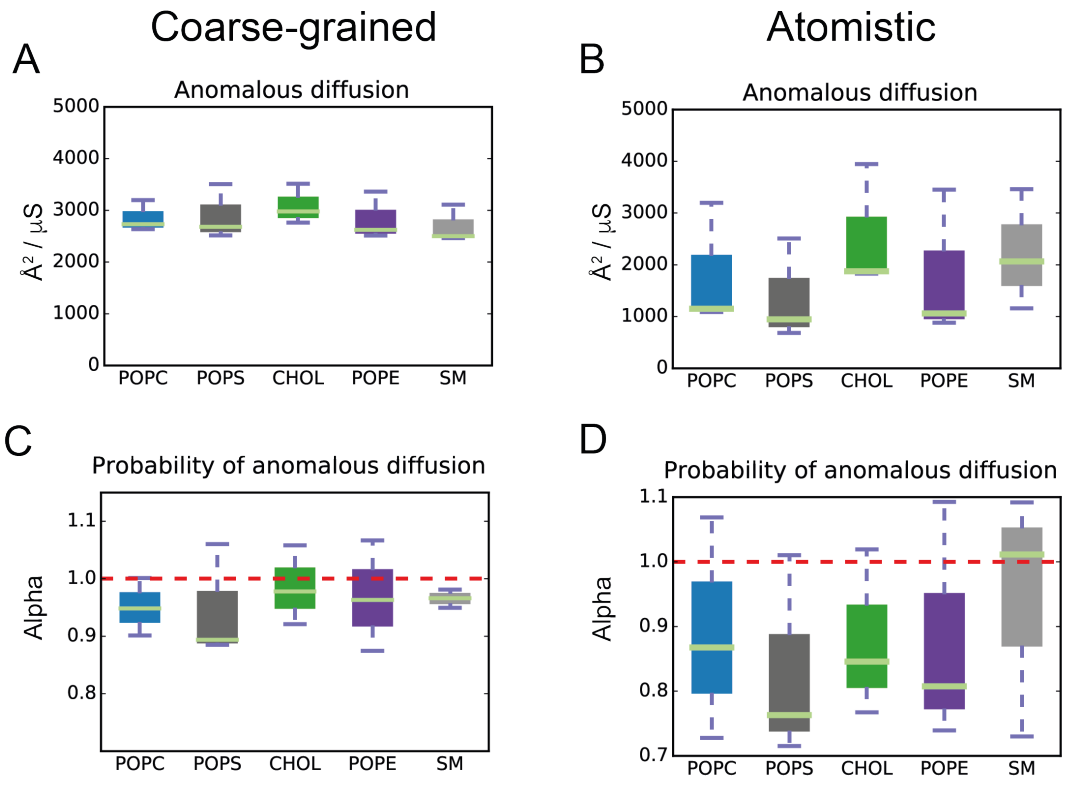



***Figure 1***. *The lipid diffusion constants for coarse grained (A) and atomistic (B) MD simulations of the asymmetric lipid bilayer patch shown as boxplots. Boxes show the interquartile range (IQR), error bars show the full range excluding outliers (points) defined as being more than ± 1.5 IQR outside the box. The results for MSD fitting to anomalous diffusion model for CG (C) and*

*atomistic (D) simulations suggest that sub-diffusion was observed for all lipid species, as the α value from fitting is lower than 1 [36].*

The design of the complex asymmetric bilayer included an even distribution of cholesterol across the leaflets with each leaflet containing 20% cholesterol. However, an asymmetric distribution of cholesterol molecules with more cholesterol found in the outer leaflet was observed after 10 μs of simulation time within the coarse grained simulations of the bilayer (**Fig. 2A**). This trend is even more pronounced in the atomistic simulations, and reflects an increase in cholesterol flip-flopping (**Fig. 2B**) as observed by visual inspection of the trajectories. Both asymmetric distribution and fast flip-flop rates have been observed in coarse-grained simulations of bilayers containing cholesterol previously [8].

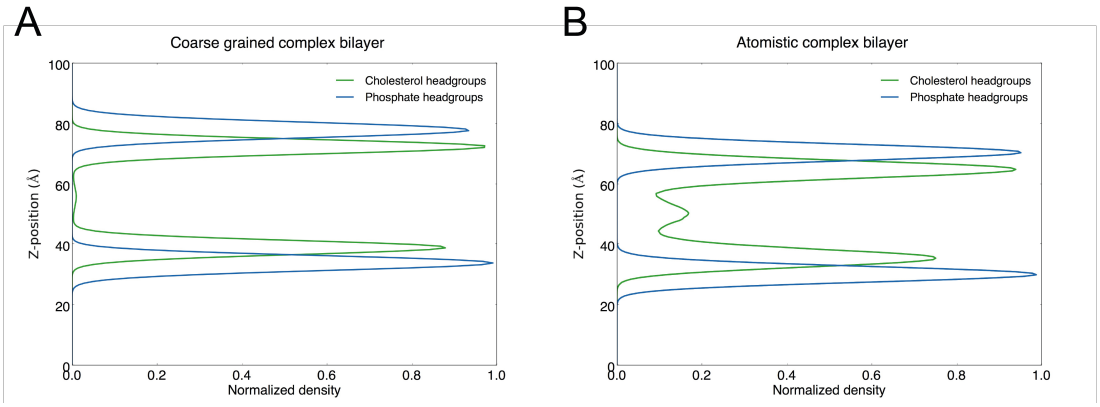



***Figure 2.*** *Distribution of phospholipids (blue line) and cholesterol (green line) headgroups along the z-dimension (membrane normal) of the simulation box in coarse grained (A) and atomistic (B) MD simulations showing an asymmetric distribution of cholesterol.*

### *3.2. P-glycoprotein strongly affects the local bilayer environment*

Satisifed that the complex bilayer was consistent with previous reports on similar or more complex systems [8, 39] we proceeded to examine the behaviour and influence of P-gp within it (**Fig. 3**). The coarse-grained simulations of this system were used to investigate the effect P-gp has on the local environment of the lipid bilayer. The local ordering of lipids in the vicinity of P-gp was analysed by calculating the spherical radial distribution functions (RDFs) of the lipid headgroups with respect to the protein. This analysis revealed a distinct pattern of lipid ordering around P-gp. For all lipid species, the highest probability distance was observed around the first interaction shell at the distance of 5 Å (**Fig. 4**). However, other peaks with decreasing probability can be observed at 10 and 15 Å, suggesting that the lipids arrange around the P-gp in ordered ring-like patterns. The distinct ordering of the local lipid bilayer has been observed previously for much smaller systems such as a transmembrane α-helix monomer [7] and is therefore unsurprising

for a larger, more complex protein. Altered fluidity and modified packing organisation of fluorescent lipid probes has previously been observed after incorporating P-gp in a PC:PE bilayer within experimental studies [41].

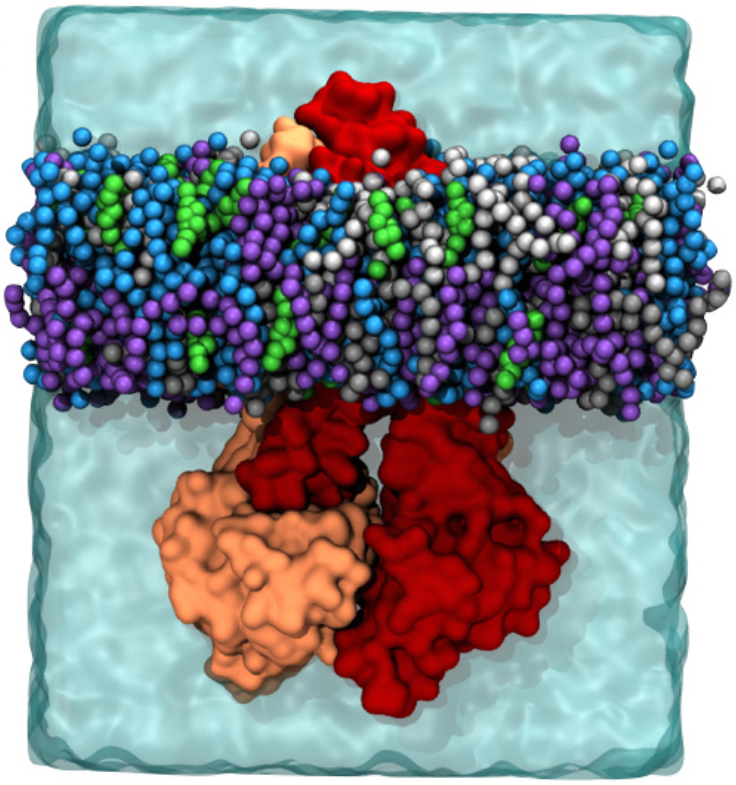


***Figure 3**. P-glycoprotein in a box of water and inserted in a complex asymmetric bilayer. TMD1/NBD1 of P-gp are shown in red and TMD2/NBD2 – in orange. The coarse grained lipids are depicted in van der Vaals representation, with POPC in blue, POPE in purple, POPS in dark gray, sphingomyelin in light gray and cholesterol in green.*

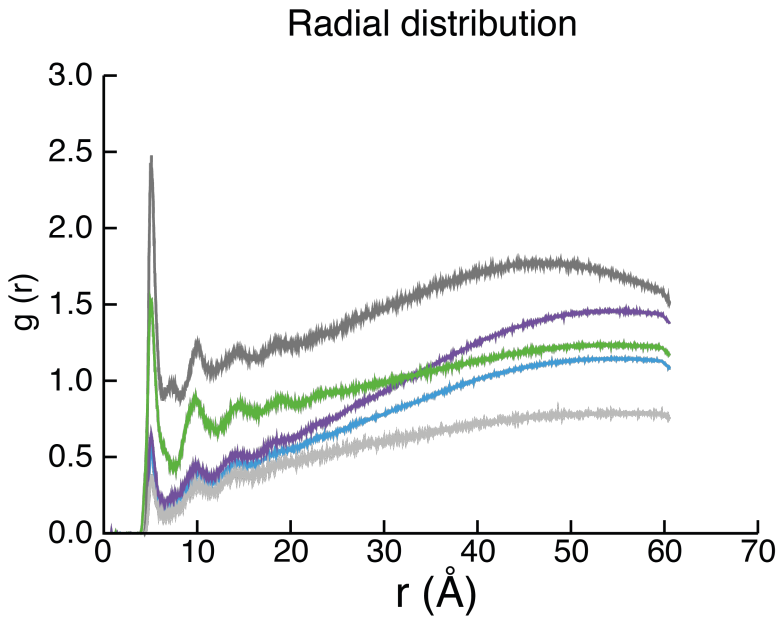



***Figure 4.*** *The spherical radial distribution functions for the various lipids in the CG simulations of P-gp in a complex lipid bilayer. POPC:blue, POPE:purple, POPS:dark grey, sphingomyelin: light grey line and cholesterol:green.*

We then performed a more specific analysis to examine whether the lipids exhibited any preferences in terms of location with respect to P-gp. Whilst the first annular shell of the outer-leaflet of the bilayer did not show any preferences for particular lipids species (**Fig. 5A**), it could be seen in the inner-leaflet that POPS forms a distinct annular ring around P-gp.

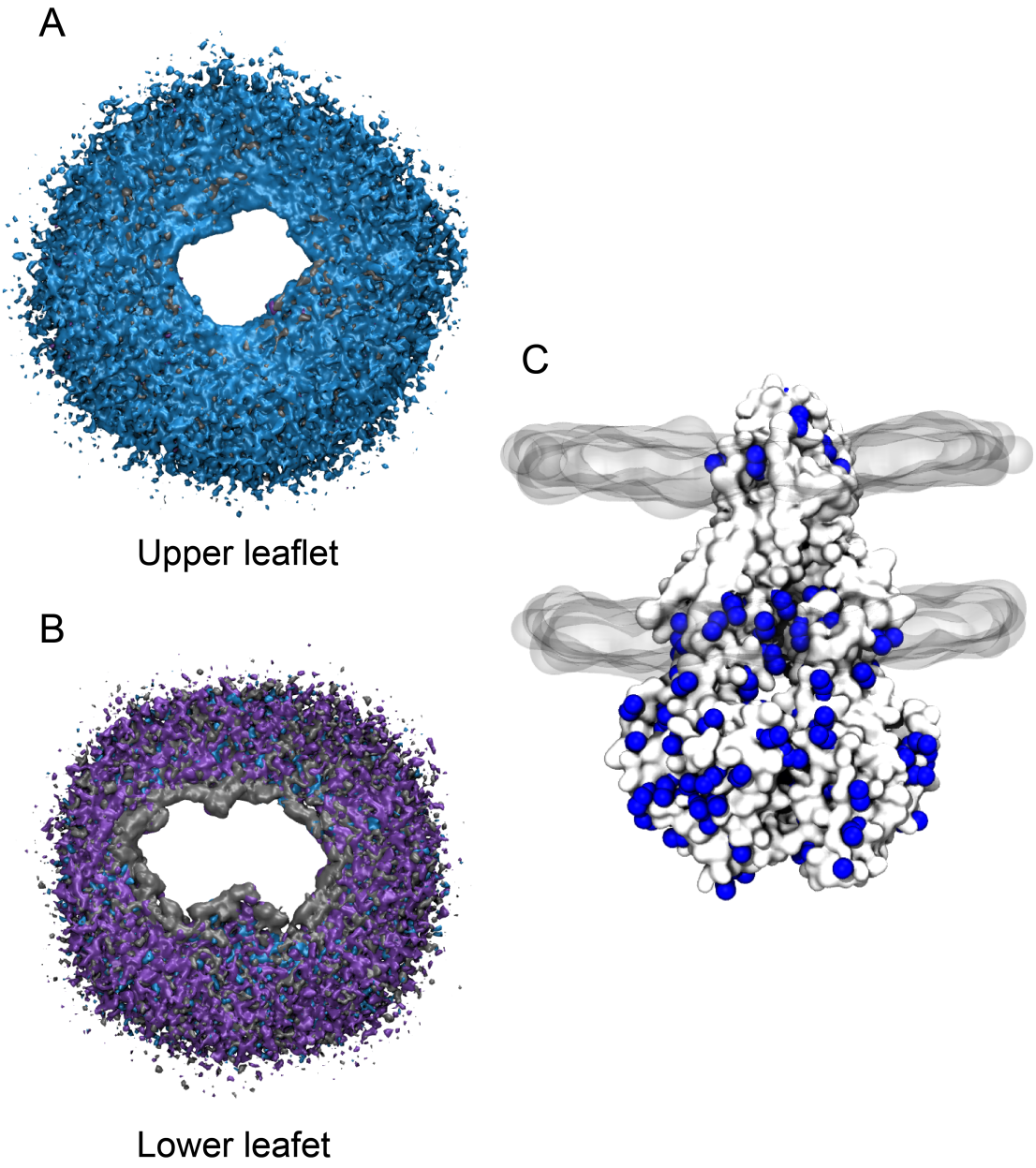



***Figure 5.*** *Phospholipid headgroup spatial occupancy density maps for the upper (A) and lower (B) leaflet of the complex asymmetric bilayer with POPC (blue), POPE (purple), POPS (dark grey) and sphingomyelin (light grey) headgroup densities shown. (C) Representation of positive residues (lysines and arginines) in blue on the P-gp surface. Phospholipid headgroup regions are depicted as grey surfaces.*

The visualisation of density maps of different lipid headgroup spatial occupancy revealed that POPS forms an annular ring in the inner leaflet of the bilayer around the positively charged residues of P-gp (**Fig. 5B,C**). The first interaction shell of the outer leaflet did not show any preference for particular lipid species (**Fig. 5A**). Similar affinity for anionic lipids has been observed for other ABC transporters [42]. This feature is proposed to be important for the function of the protein as delipidation causes a decrease in ATP activity [42]. Mass spectrometry of purified P-gp suggested more preferential binding of

negatively charged lipids [43] and the lipid species that have been associated with purified P-gp were mostly PS and PE with small quantities of PC [44].

We also observed an effect on the rate of cholesterol flip-flop by P-gp. In the coarse-grained simulations of the complex bilayer alone, we observed 368 ± 36 (standard deviation) cholesterol flip-flop events, whilst in the presence of P-gp this number was 525 ± 14.

Diffusion constants of the lipids within the bilayer with P-gp present were lower (**Fig. 6**) in comparison to the lipid only systems (**Fig. 1**). Additionally, when compared to other phospholipids in the coarse grained system with P-gp present, POPS has a lower diffusion constant, consistent with the formation of an annular ring. Lipids are still expected to diffuse anomalously, albeit the α constant is increased, suggesting that the insertion of P-gp has an influence on lipid bilayer fluidity and packing. Loe and Sharom investigated the changes in lipid bilayer properties in P-gp expressing cell lines and concluded that molecular packing of the lower leaflet increases with higher levels of P-gp expression [45]. These results are in line with our observations of tighter packing and the ring-like lipid ordering which would also be expected to result in lower diffusion constants.

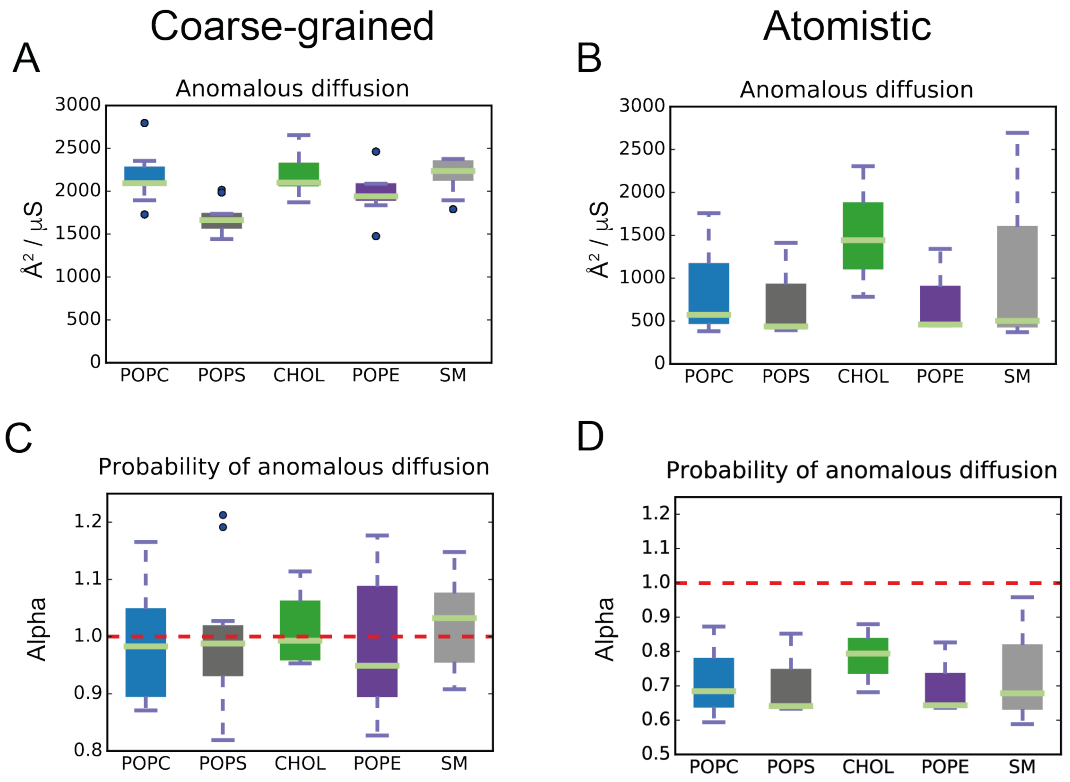



***Figure 6.*** *The lipid diffusion constants for coarse grained (A) and atomistic (B) MD simulations of P-gp inserted in to an asymmetric lipid bilayer patch shown as boxplots. Boxes show the IQR, error bars show the full range excluding outliers (points) defined as being more than ± 1.5 IQR outside the box. The results for MSD fitting to an anomalous diffusion model for coarse grained (B) and atomistic (D) simulations suggest that sub-diffusion was observed for all lipid species, as the α value from fitting is lower than 1 [36] as*

*observed for the pure lipid system in Fig. 1.*

Analysis of the root mean square deviation (RMSD) and root mean squared fluctuations (RMSF) of the Cα atoms (**Fig. 7**) suggests that P-gp is, on these relatively short timescales at least, conformationally stable in this complex lipid environment.

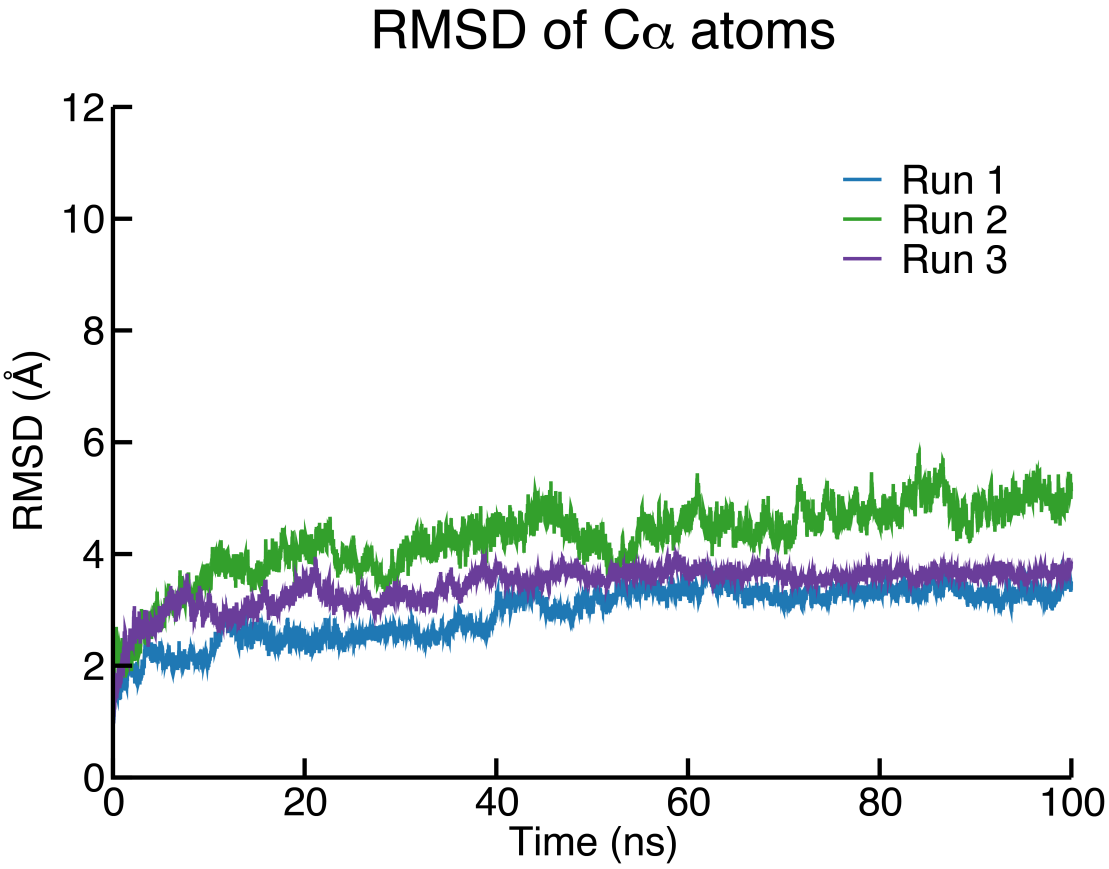



***Figure 7.*** *RMSD of Cα atoms of P-gp. Repeat runs are indicated as blue, green and purple lines. The RMSDs suggest that the presence of the complex asymmetric bilayer does not induce significant conformational changes in the 100 ns timescale of these atomistic simulations.*

### *3.3. Cholesterol binds in the clefts of the P-gp surface*

From the analysis of averaged lipid density maps cholesterol was observed to be present in the clefts of the P-gp surface, often for long periods of simulation time. The same clefts were occupied in different repeats of the simulations, strongly suggesting a preference that was most prominently observed for the gaps between transmembrane helices (TMH) 4/6 and 10/12. This lipid binding arrangement, known as “non-annular surface lipids”, is not unusual for membrane protein that exhibit clefts, cavities or pores on their surface [46]. For lipid interaction to be considered stable two conditions have to be fulfilled – the lipid headgroup needs to form hydrophilic interactions with

the protein, while the tails can interact with a nonpolar surface of the protein [46].

The simulations were analysed to define the cholesterol binding sites by identifying the residues that formed the site. These were defined as residues that remain within close distance (< 6 Å) of cholesterol for at least half of the simulation length. This analysis revealed the cleft between TMH 10/12 (**Fig. 8A**) to be the most stable potential binding site as it was visited in 65% of all simulation time and more than three quarters of the residues associated with this binding site remained within 6 Å of the cholesterol molecule for the length of the interaction. Other sites of interest were situated between TMH 7/8 (**Fig. 8B**) and 9/12 (**Fig. 8C**). All three sites fulfilled the criteria for stable non-annular surface lipid interactions with P-gp, as their headgroups were interacting with basic residues (Lys or Arg) and the tails generally interacted with the nonpolar surface of transmembrane domains (**Fig. 8**).

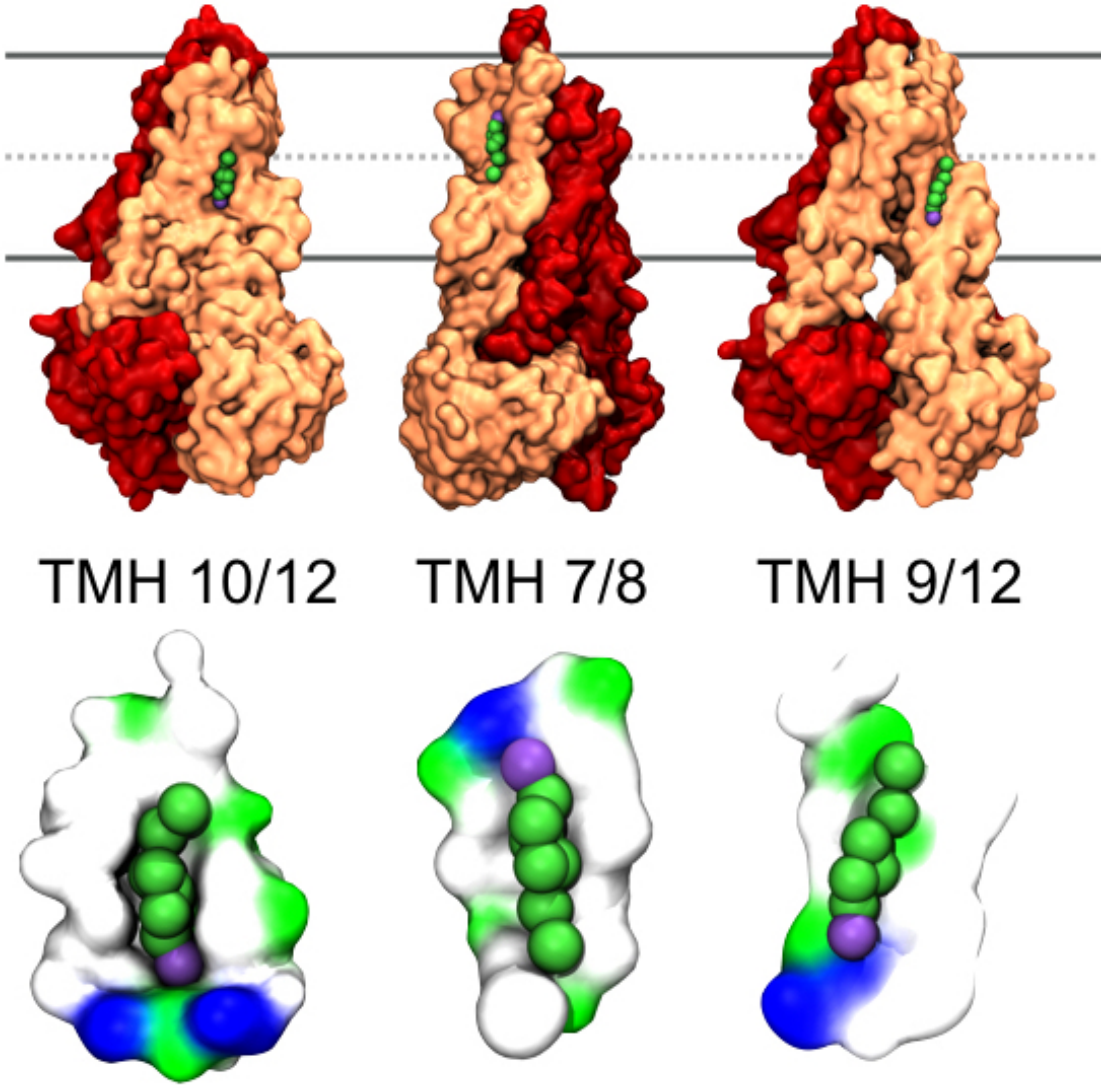



***Figure 8.*** *The top panel shows the location of cholesterol (green spheres)*

*binding sites on the P-gp surface with TMD1/NBD1 colored in red and TMD2/NBD2 colored in orange. The cholesterol headgroup bead is shown in purple. The cholesterol binding sites are shown in more detail on the lower panel, where P-gp residues within 10 Å of a cholesterol molecule are shown in surface representation with positive residues in blue and polar residues in green. The cholesterol headgroup bead (purple) is interacting with positive residues in all binding sites.*

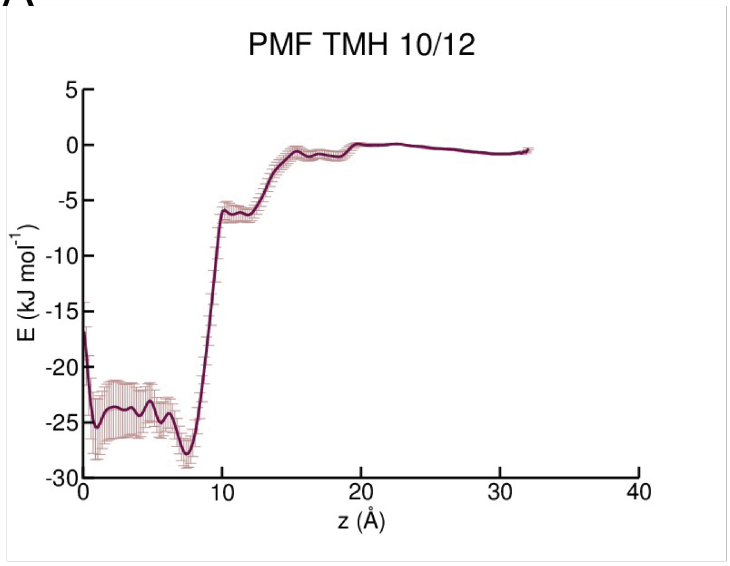
### *3.4. Cholesterol binds tightly in the groove between TMH 10/12*

The strength of potential cholesterol binding sites on P-gp surface could not be estimated from unbiased MD simulations, therefore, umbrella sampling simulations were used to calculate the potential of the mean force of cholesterol binding for each of the proposed sites (TMH 10/12, TMH 7/8 and TMH 9/12).

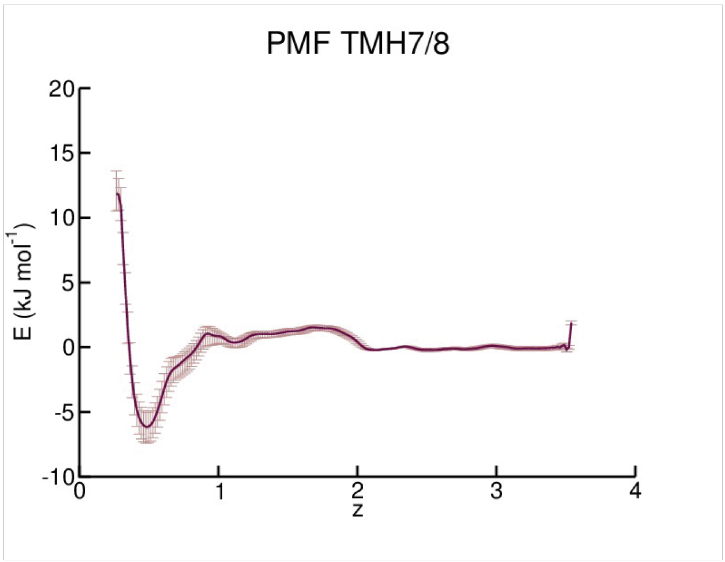
The free energy profile obtained from these calculations suggests that cholesterol binds most tightly to the TMH 10/12 with a broad energy minimum close to the initial position (**Fig. 9A, Fig. 10A**). Visual inspection of the umbrella sampling windows provides a potential explanation of the free energy profile. The groove formed by TMHs 10 and 12 **(Fig. 10A)** is relatively deep and the movement through the groove was reflected in the broad energy minimum. During this stage cholesterol was still interacting with the transmembrane helices. The lowest energy state (-27.9 ± 1.2 kJ/mol) was reached when cholesterol had almost left the groove rather than at its deepest point (-25.6 ± 2.8 kJ/mol), however, the difference is not great.

The other potential binding sites explored in this study showed similar free energy values with -6.4 ± 1.2 kJ/mol for the TMH7/8 site (**Fig. 9B**) and -4.4 ± 1.0 kJ/mol at the site between TMHs 9/12 (**Fig. 9C**). Interestingly, these values are similar to the free energy of a potential secondary binding site next to TMH 10 (**Fig. 10B**). The lower free energy value suggests weaker binding, which could be explained by the depth of the grooves in protein surface that are used by cholesterol to bind. The binding pockets are made of almost all hydrophobic amino acids with a polar group required to interaction with the OH group of the cholesterol (a lysine in the main and secondary binding sites in **Fig. 10**).

A


PMF TMH 10/12
E (kJ mol-1)
z (Å)

B


PMF TMH7/8
E (kJ mol-1)
z

C

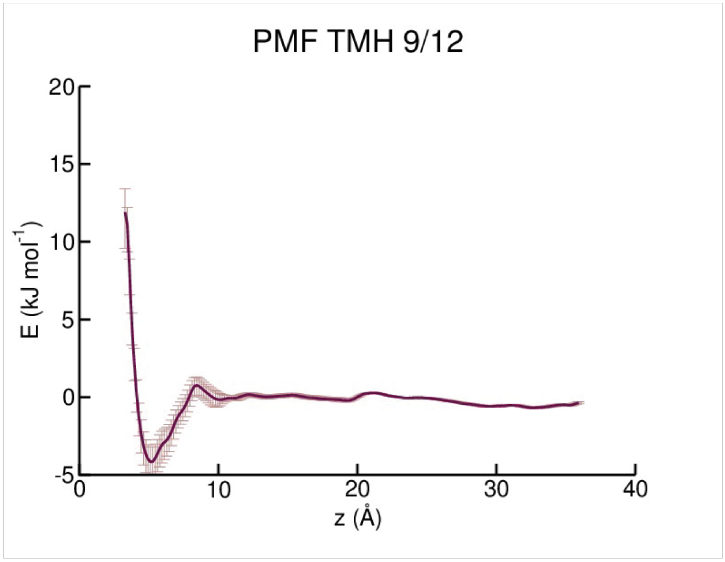

PMF TMH 9/12
E (kJ mol-1)
z (Å)

***Figure 9.*** *The PMF profiles of cholesterol binding sites at (A) TMH 10/12, (B) TMH 7/8 and (C) TMH 9/12. The error bars are derived from bootstrapping calculations (see Methods).*

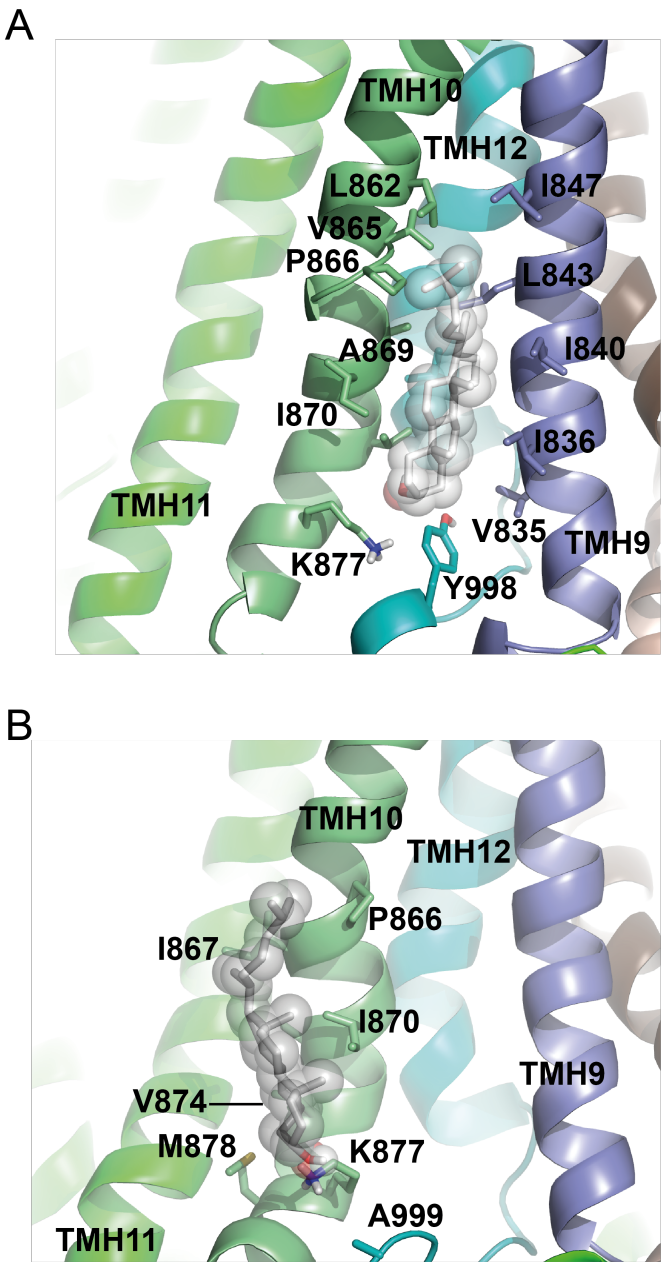



***Figure 10.*** *Cartoons of the main cholesterol binding site (A) TMH 10/12 and the secondary site (B) obtained from the PMF shown in Fig. 9A. K877 at the bottom of TMH10 interacts with the OH group of cholesterol and likely provides an orientation anchor in an otherwise very hydrophobic cleft.*

## 4. Discussion

The lipid bilayer environment has a considerable effect on structure and function of P-gp [5, 6, 47]. The extent of this dependency and the importance

of studying membrane proteins, particularly, P-gp in their native environment has been shown by two recent studies [48, 49]. The first study voiced concerns over the physiological relevance of crystal structures acquired from environment that consists of detergents [49], while the latter revealed a requirement for lipids to activate a switch at the transmission interface between TMD2 and NBD1 [48]. Other studies have shown the importance of individual components of the lipid bilayer, for example, cholesterol, which, although unlikely to be a transported substrate of P-gp, can intimately interact with the transporter at specific sites [17].

In this study, we used a multi-scale MD simulation approach to analyse a bilayer patch designed to mimic the epithelial cells found in the human blood-brain barrier. Analysis of the diffusion and distribution of lipids revealed that this bilayer behaves similarly to other complex bilayers in both coarse grained and atomistic resolutions reported in the literature [8, 39], including the movement of cholesterol.

P-gp interactions with the complex bilayer were also investigated using a multi-scale approach. Coarse-grained simulations provided information on the effects P-gp has on the surrounding lipid bilayer, while atomistic MD simulations were used to determine the stability of the protein in a complex bilayer.

Insertion of P-gp in the complex bilayer caused the lipid ordering in ring-like structures and this effect was most profound for POPS. The negatively charged lipid species formed an annular ring around P-gp in the lower leaflet, most likely attracted by the large amount of positive residues in the headgroup region of the lower leaflet. More specific type of binding was observed for cholesterol molecules, which appeared to bind strongly to the crevices on P-gp surface, in particular the cavities formed by TMH 4/6 and TMH 10/12 which form the putative substrate entrance gates [50, 51].

The analysis of P-gp dynamics in atomistic simulations of the protein in complex bilayer showed that the bilayer composition did not have a large effect on the overall stability of the structure. Nevertheless, the occurrence of some more subtle changes that would be revealed by longer timescales is possible.

Three of the potential cholesterol binding sites were investigated by umbrella sampling simulations to obtain the potential of mean force. The binding of cholesterol was strongest at a binding site formed by TMH 10/12, potentially because the cavity was deeper than the other sites in the umbrella sampling study. The shallower sites had similar free energy of cholesterol binding, which could suggest less specific or non-specific binding on the surface of P-gp.

A clear definition of potential cholesterol binding sites was achieved in the coarse grained simulations of P-gp in complex asymmetric bilayer. These results provide a hypothesis of cholesterol binding to P-gp in a detail never explored before and suggest that cholesterol uses a binding site that is not

formed by a CRAC/CARC motif. Whether this is present in other ABC transporters would be difficult to predict at this time, because even very small changes in sequence could render the site incompetent.

Cholesterol binding sites have been attributed to sequence interaction motifs such as CARC/CRAC motifs, which are also present in P-gp [17]. Experimental evidence of P-gp co-eluting with [$^3$H]cholesterol suggested that P-gp might interact with cholesterol directly [52]. Lipids that bind in the grooves of the protein surface or in the spaces between domains can potentially stabilize the protein structure [53].

## 5. Conclusion

Here we have shown that lipid components in a complex lipid bilayer system designed to mimic brain epithelial cells tend to exhibit anomalous diffusion, an observation consistent with previous observations on mixed lipids systems. We also find that P-glycoprotein is conformationally stable on a relatively short time-scale in this complex lipid bilayer and that phosphatidylserine within the inner leaflet can form an annulus around the protein. We were also able to see cholesterol spontaneously interact with three potential pockets on the surface of P-gp and upon further investigation with potential of mean force calculations were able to show that one of these, located between transmembrane helices 10 and 12 is the most probably one in terms of binding. This prediction awaits experimental confirmation. Reduced snapshot trajectories are available for download from http://doi.org/10.5281/zenodo.847637.

## Acknowledgements

LD is funded by a NC3Rs studentship and was also a Cephalosporin Scholar at Lady Margaret Hall, Oxford. We thank the Advanced Research Computing (ARC) facility, the EPSRC UK National Service for Computational Chemistry Software (NSCCS) at Imperial College London (grant no. EP/J003921/1) and the ARCHER UK National Supercomputing Services for computer time granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://www.hecbiosim.ac.uk), supported by EPSRC (grant no. EP/L000253/1). H.K. acknowledges a research fellowship from the Alfred Benzon Foundation. Work in PCB's laboratory is supported by the MRC (MR/M000435/1) and the BBSRC (BB/M006395/1).

## References

[1] Dean, M., Rzhetsky, A., Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 2001, 11, 1156-66.

[2] Sarkadi, B., Homolya, L., Szakács, G., Váradi, A. Human multidrug resistance ABCB and ABCG transporters: Participation in a chemoimmunity defense system. Physiol. Rev. 2006, 86, 1179.

[3] van Meer, G., de Kroon, A.I.P.M. Lipid map of the mammalian cell. J. Cell. Sci. 2010, 124, 5.
[4] Biggin, P.C., Bond, P.J. Molecular dynamics simulations of membrane proteins. Methods. Mol. Biol. 2015, 1215, 91-108.
[5] Sharom, F.J. Complex interplay between the P-glycoprotein multidrug efflux pump and the membrane: Its role in modulating protein function. Front. Oncol. 2014, 4, 41.
[6] Pollock, N.L., McDevitt, C.A., Collins, R., Niesten, P.H.M., Prince, S., Kerr, I.D., et al. Improving the stability and function of purified ABCB1 and ABCA4: The influence of membrane lipids. Biochim. Biophys. Acta. 2014, 1838, 134-47.
[7] Koldsø, H., Sansom, M.S.P. Local lipid reorganization by a transmembrane protein domain. J. Phys. Chem. Letts. 2012, 3, 3498-502.
[8] Ingólfsson, H.I., Melo, M.N., van Eerden, F.J., Arnarez, C., Lopez, C.A., Wassenaar, T.A., et al. Lipid organization of the plasma membrane. J. Am. Chem. Soc. 2014, 136, 14554-9.
[9] Koldsø, H., Sansom, M.S.P. Organization and dynamics of receptor proteins in a plasma membrane. J. Am. Chem. Soc. 2015, 137, 14694-704.
[10] Hedger, G., Shorthouse, D., Koldsø, H., Sansom, M.S.P. Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor. J. Phys. Chem. B. 2016, 120, 8154-63.
[11] Buyan, A., Kalli, A.C., Sansom, M.S.P. Multiscale simulations suggest a mechanism for the association of the dok7 PH domain with pip-containing membranes. PLOS Comp. Biol. 2016, 12, e1005028.
[12] Hedger, G., Sansom, M.S.P., Koldsø, H. The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. Sci. Reps. 2015, 5, 9198.
[13] Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J., Tieleman, D.P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. J. Chem. Theor. Comput. 2014, 10, 676-90.
[14] Stansfeld, P.J., Sansom, M.S.P. From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. J. Chem. Theory Comput. 2011, 7, 1157-66.
[15] Campbell, S.D., Regina, K.J., Kharasch, E.D. Significance of lipid composition in a blood-brain barrier–mimetic PAMPA assay. J. Biomol. Screen. 2013, 19, 437-44.
[16] Eckford, P.D.W., Sharom, F.J. Interaction of the P-glycoprotein multidrug efflux pump with cholesterol: Effects on ATPase activity, drug binding and transport. Biochemistry. 2008, 47, 13686-98.
[17] Clay, A.T., Lu, P., Sharom, F.J. Interaction of the p-glycoprotein multidrug transporter with sterols. Biochemistry. 2015, 54, 6586-97.
[18] Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., Vries, A.H.D. The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations. Journal of Physical Chemistry B. 2007, 111, 7812-24.

[19] Siakotos, A.N., Rouser, G. Isolation of highly purified human and bovine brain endothelial cells and nuclei and their phospholipid composition. Lipids. 1969, 4, 234–9.
[20] Tewes, B.J., Galla, H.J. Lipid polarity in brain capillary endothelial cells. Endothelium. 2001, 8, 207-20.
[21] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 1984, 81, 3684-90.
[22] Domański, J., Stansfeld, P.J., Sansom, M.S.P., Beckstein, O. Lipidbook: A public repository for force-field parameters used in membrane simulations. J. Memb. Biol. 2010, 236, 255-8.
[23] Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.-J. The MARTINI coarse-grained force field: Extension to proteins. J. Chem. Theory Comput. 2008, 4, 819-34.
[24] de Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L.V., et al. Improved parameters for the martini coarse-grained protein force field. J. Chem. Theory Comput. 2013, 9, 687-97.
[25] Domicevica, L., Biggin, P.C. Homology modelling of human p-glycoprotein. Biochem. Soc. TRans. 2015, 43, 952-8.
[26] Jefferys, E., Sands, Z.A., Shi, J., Sansom, M.S.P., Fowler, P.W. Alchembed: A computational method for incorporating multiple proteins into complex lipid geometries. J. Chem. Theor. Comput. 2015, 11, 2743-54.
[27] Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simuation. J. Chem. Theory Comput. 2008, 4, 435-47.
[28] Hess, B., Bekker, J., Berendsen, H.J.C., Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. J. Comp. Chem. 1997, 18, 1463-72.
[29] Darden, T., York, D., Pedersen, L. Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems. J. Chem. Phys. 1993, 98, 10089-92.
[30] Bussi, G., Donadio, D., Parrinello, M. Canonical sampling through velocity rescaling. J. Chem. Phys. 2007, 126.
[31] Parrinello, M., Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 1981, 52, 7182-90.
[32] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J., Dror, R., et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Struc. Func. Genet. 2010, 78, 1950-8.
[33] Jämbeck, J.P.M., Lyubartsev, A.P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. J. Phys. Chem. B. 2012, 116, 3164-79.
[34] Hedger, G., Rouse, S.L., Domański, J., Chavent, M., Koldsø, H., Sansom, M.S.P. Lipid-loving ANTs: Molecular simulations of cardiolipin interactions and the organization of the adenine nucleotide translocase in model mitochondrial membranes. Biochemistry. 2016, 55, 6238-49.
[35] Reddy, T., Shorthouse, D., Parton, Daniel L., Jefferys, E., Fowler, Philip W., Chavent, M., et al. Nothing to sneeze at: A dynamic and integrative computational model of an influenza A virion. Structure. 2015, 23, 584-97.

[36] Kneller, G.R., Baczynski, K., Pasenkiewicz-Gierula , M. Communication: Consistent picture of lateral subdiffusion in lipid bilayers: Molecular dynamics simulation and exact results. J. Chem. Phys. 2011, 135, 141105.
[37] Michaud-Agrawal, N., Denning, E.J., Woolf, T., Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 2011, 32, 2319-27.
[38] Humphrey, W., Dalke, A., Schulten, K. VMD - Visual molecular dynamics. J. Mol. Graph. 1996, 14, 33-8.
[39] Jeon, J.-H., Monne, H.M.-S., Javanainen, M., Metzler, R. Anomalous diffusion of phospholipids and cholesterols in a lipid bilayer and its origins. Phys. Rev. Letts. 2012, 109, 188103.
[40] Schwille, P., Haupts, U., Maiti, S., Webb, W.W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. Biophys. J. 1999, 77, 2251-65.
[41] Rothnie, A., Theron, D., Soceneantu, L., Martin, C., Traikia, M., Berridge, G., et al. The importance of cholesterol in maintenance of P-glycoprotein activity and its membrane perturbing influence. Eur. Biophys. J. 2001, 30, 430-42.
[42] Bechara, C., Nöll, A., Morgner, N., Degiacomi, M.T., Tampé, R., Robinson, C.V. A subset of annular lipids is linked to the flippase activity of an ABC transporter. Nat Chem. 2015, 7, 255-62.
[43] Marcoux, J., Wang, S.C., Politis, A., Reading, E., Ma, J., Biggin, P.C., et al. Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. Proc. Natl. Acad. Sci. USA. 2013, 110, 9704-9.
[44] Sharom, F.J., Yu, X., Chu, J.W., Doige, C.A. Characterization of the ATPase activity of P-glycoprotein from multidrug-resistant Chinese hamster ovary cells. Biochem. J. 1995, 308, 381-90.
[45] Loe, D.W., Sharom, F.J. Interaction of multidrug-resistant Chinese hamster ovary cells with amphiphiles. Br. J. Cancer. 1993, 68, 342-51.
[46] Palsdottir, H., Hunte, C. Lipids in membrane protein structures. Biochim. Biophys. Acta. 2004, 1666, 2-18.
[47] Urbatsch, I.L., Sankaran, B., Bhagat, S., Senior, A.E. Both P-glycoprotein nucleotide-binding sites are catalytically active. J. Biol. Chem. 1995, 270, 26956-61.
[48] Loo, T.W., Clarke, D.M. P-glycoprotein ATPase activity requires lipids to activate a switch at the first transmission interface. Biochem. Biophys. Res. Comms. 2016, 472, 379-83.
[49] Zoghbi, M.E., Cooper, R.S., Altenberg, G.A. The lipid bilayer modulates the structure and function of an atp-binding cassette exporter. J. Biol. Chem. 2016, 291, 4453-61.
[50] Kodan, A., Yamaguchi, T., Nakatsu, T., Sakiyama, K., Hipolito, C.J., Fujioka, A., et al. Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog. Proc. Natl. Acad. Sci. USA. 2014, 111, 4049-54.
[51] Jin, M.S., Oldham, M.L., Zhang, Q., Chen, J. Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. Nature. 2012, 490, 566-9.

[52] Kimura, Y., Kioka, N., Kato, H., Matsuo, M., Ueda, K. Modulation of drug-stimulated ATPase activity of human MDR1/P-glycoprotein by cholesterol. Biochem. J. 2007, 401, 597-605.
[53] Baier, C.J., Fantini, J., Barrantes, F.J. Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. Sci. Rep. 2011, 1, 69.