Molecular Dynamics Studies of P-Glycoprotein



Laura Domicevica

Lady Margaret Hall

University of Oxford

Submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2016

Abstract

Molecular Dynamics Studies of P-Glycoprotein

Laura Domicevica
Lady Margaret Hall
University of Oxford
Doctor of Philosophy
Trinity Term 2016

Due to its involvement in multidrug resistance in cancer and other disorders, P-glycoprotein has been a subject of many studies, however, the evidence from structural and functional investigations is often conflicting. Homology models of human P-glycoprotein were prepared and analysed to choose the most suitable model for further experiments. Molecular dynamics simulations of the homology models revealed closure of potential substrate entrance gates. The results of the steered molecular dynamics and other simulations suggest that amitriptyline enters the central cavity through the cytoplasmic part of the entrance gates formed by transmembrane helices 4 and 6. P-glycoprotein was inserted in a complex asymmetric bilayer designed to mimic the composition of blood-brain barrier and the system was subjected to coarse-grained simulations and more interesting cases were converted to atomistic resolution. Three major insights were obtained. Firstly, P-glycoprotein affects the local bilayer environment by causing local ordering of lipids and decrease in lipid diffusion constants. Secondly, several potential cholesterol binding sites were observed in the simulations. Lastly, the presence of complex bilayer might affect the flexibility of substrate entrance gate-forming helices. Taken together, these results highlight the importance of structural studies that can generate new and testable hypotheses.

Acknowledgements

First I would like to thank my supervisor, Professor Phil Biggin, for giving me the chance to undertake this project and study something that was entirely new to me. He has supported me throughout the degree and provided the opportunities to implement and expand my ideas.

I would also like to thank members of the lab for supporting me through the highs and the lows of DPhil student life, as well as providing me with much needed advice, especially Maira, Maria, Riley, Matteo, Tyler, Vishal, Nick, Eric, Jo, Phil F., George and Anna. A huge thanks to Heidi, who inspired the lipid project and is herself an inspiration for any aspiring woman scientist, and to Tere, who helped me to understand programming better and was always happy to hear out my ideas (also thanks for all the charity cakes). Tom and Fiona - thank you for your friendship and all the exciting conversations. A very special thanks goes to Amanda and Julia, who have been very supporting friends and have shown me that the world is diverse and beautiful.

Many thanks to the friends from LMH, including Sadia, Gerda, Dominic, Shiv, Paton, Hannah and Ana. A special thank you to my college mother Qin, who has remained one of my best friends throughout the years in Oxford.

Finally, I would like to thank my family, especially my mother, father and grandmothers, for their unwavering support and always believing in me. To Rihards - thank you for your love and support and allowing me to chase my dreams.

Laura Domicevica
September 2016, University of Oxford

List of Publications Arising From This Thesis

Domicevica, L. and Biggin, P.C. "Homology modelling of human P-glycoprotein." *Biochemical Society Transactions*, 2015.

Domicevica, L., Paramo, T. and Biggin, P.C. "Exploring Drug Access Routes to P-glycoprotein." (Manuscript in Preparation for PLOS Comp Biol.)

Contents

List of Figures x

List of Tables xii

List of Abbreviations, Acronyms and Symbols xiii

1 Introduction 1
- 1.1 Multidrug resistance . . . 1
  - 1.1.1 ABC transporters in multidrug resistance . . . 3
- 1.2 P-glycoprotein . . . 4
  - 1.2.1 P-glycoprotein in the blood-brain barrier . . . 5
  - 1.2.2 P-glycoprotein as a lipid flippase . . . 5
  - 1.2.3 P-glycoprotein structure . . . 6
    - 1.2.3.1 Crystal structures of proteins related to human P-gp 10
  - 1.2.4 P-glycoprotein drug binding site . . . 11
  - 1.2.5 P-gp substrates . . . 12
  - 1.2.6 P-gp modulators and inhibitors . . . 14
- 1.3 P-glycoprotein transport cycle . . . 16
  - 1.3.1 Substrate transport . . . 20
- 1.4 Multidrug efflux pump and the lipid bilayer . . . 22
  - 1.4.1 P-glycoprotein interactions with sterols . . . 23
- 1.5 Molecular modelling studies of P-glycoprotein . . . 24
  - 1.5.1 Homology models of human P-glycoprotein . . . 24
  - 1.5.2 Studies that involve dynamics . . . 25
  - 1.5.3 Molecular dynamics studies of P-gp substrate entrance . . . 27
- 1.6 Protein-ligand interactions in MD . . . 28
- 1.7 Thesis aims . . . 30
  - 1.7.1 P-gp substrate entrance pathways . . . 30
  - 1.7.2 P-glycoprotein interactions with lipids . . . 31
  - 1.7.3 Dynamic protein-ligand interactions . . . 32

2 Simulation Theory and Methods 33
- 2.1 Introduction . . . 33
- 2.2 Molecular dynamics simulation . . . 34
  - 2.2.1 Equations of motion . . . 34

2.3 Force fields . . . 36
2.3.1 Bonded interactions . . . 37
2.3.2 Nonbonded interactions . . . 38
2.3.3 All-atom forcefields . . . 39
2.3.4 Coarse-grained forcefields . . . 40
2.3.5 Conversion from coarse-grained to atomistic representation . . . 40
2.4 Molecular dynamics simulation technique . . . 41
2.4.1 Constraints . . . 41
2.4.2 Periodic boundary conditions . . . 42
2.4.3 Treatment of long range interactions . . . 42
2.4.4 Constant temperature and pressure simulations . . . 43
2.4.5 Energy minimisation . . . 45
2.4.6 General simulation setup . . . 46
2.5 Free energy calculations . . . 47
2.5.1 Steered molecular dynamics . . . 48
2.5.2 Potential of mean force . . . 48
2.5.3 Umbrella sampling . . . 50
2.5.4 The weighted histogram analysis method (WHAM) . . . 50
2.6 Homology modeling . . . 51

3 Homology Modelling and Atomistic Simulations of P-glycoprotein 53
3.1 Homology modelling of human P-glycoprotein . . . 53
3.1.1 Methods . . . 54
3.1.2 Available Templates . . . 55
3.1.3 Homology model from the *C. elegans* crystal structure . . . 57
3.1.4 Homology model from *C. merolae* crystal structure . . . 58
3.1.5 Homology model from refined mouse crystal structure . . . 58
3.1.6 Template selection . . . 59
3.2 Atomistic simulations of human P-glycoprotein homology models . . . 61
3.2.1 Methods . . . 61
3.2.2 Atomistic simulations of the human homology model from *C. elegans* crystal structure . . . 62
3.2.3 Atomistic simulations of human homology model from the *C. merolae* crystal structure . . . 63
3.2.4 Atomistic simulations of human homology model from refined mouse crystal structure . . . 64
3.2.5 Atomistic simulations of *C. elegans*, refined mouse ABCB1 and *C. merolae* crystal structures . . . 65
3.2.6 Justification of the mouse template . . . 67
3.3 Atomistic simulations of human P-glycoprotein homology model with ATP and Mg^{2+} ions . . . 68
3.3.1 Methods . . . 68
3.3.2 Comparison between apo and ATP and Mg^{2+} bound P-glycoprotein atomistic simulations . . . 69

3.3.3 Behaviour of the potential substrate entrance gates in MD simulations . . . 70
3.3.4 Conclusions . . . 74

4 Investigation of P-glycoprotein substrate entrance pathways 75
4.1 Investigation of access to the central cavity of P-gp . . . 76
4.1.1 Methods . . . 77
4.1.2 Results . . . 77
4.2 The binding site of amitriptyline . . . 81
4.2.1 Free diffusion of amitriptyline in the central cavity . . . 81
4.2.2 Methods . . . 83
4.2.3 Consensus docking . . . 83
4.2.4 Summary of amitriptyline binding sites . . . 86
4.3 The unbinding of amitriptyline . . . 86
4.3.1 Methods . . . 86
4.3.2 Unbinding pathways of amitriptyline . . . 87
4.4 Amitriptyline pathway to its binding site is not linear . . . 89
4.4.1 Further experiments . . . 90

5 P-glycoprotein in a complex lipid bilayer 93
5.1 Blood-brain membrane lipid bilayer . . . 94
5.1.1 Methods . . . 97
5.1.2 Description of complex asymmetric bilayer by multiscale MD simulations . . . 98
5.2 Coarse grained MD simulations of P-gp in complex bilayer . . . 100
5.2.1 Methods . . . 101
5.2.2 P-glycoprotein strongly affects the local bilayer environment . . . 103
5.2.3 Lipid access to the central cavity . . . 107
5.3 Potential of mean force simulations of potential cholesterol binding sites 110
5.3.1 Methods . . . 111
5.3.2 Cholesterol binds tightly in the groove between TMH 10/12 . . . 112
5.4 Atomistic simulations of P-gp in complex bilayer . . . 113
5.4.1 Methods . . . 115
5.4.2 Results . . . 115
5.4.3 Lipid bilayer effect on P-gp . . . 118
5.5 Conclusions . . . 120
5.5.1 Further experiments . . . 123

6 LINTools - dynamic representation of protein-ligand interactions 124
6.1 Introduction . . . 124
6.2 LINTools workflow . . . 127
6.2.1 Data input and analysis . . . 128
6.2.1.1 Obtaining coordinates of the ligand . . . 128
6.2.1.2 Determine protein residues of interest . . . 129
6.2.1.3 Determine hydrogen bonding . . . 130

6.2.1.4 RMS fluctuations of the ligand . . . 130
6.2.2 Plotting the residues of interest . . . 132
6.2.3 2D representation of protein-ligand interactions . . . 132
6.2.3.1 Drawing of 2D ligand . . . 132
6.2.3.2 Determining residue placement in 2D . . . 135
6.2.4 Compilation of final image . . . 137
6.3 Dependencies . . . 137
6.4 Open-source and licensing . . . 139
6.5 Good software development practices . . . 139
6.6 P-glycoprotein and amitriptyline binding . . . 141
6.6.1 Free diffusion of amitriptyline in the central cavity of P-gp . . . 142
6.6.2 MD simulations of docked amitriptyline . . . 142
6.6.3 Comparison of amitriptyline binding sites from LINTools analysis 144
6.7 Drug binding in bromodomains . . . 146
6.7.1 Alprazolam . . . 147
6.7.2 3,5-Dimethylisoxazole derivative . . . 147
6.7.3 RVX-208 and RVX-OH . . . 149
6.7.4 Dual bromodomain-kinase inhibitors . . . 150
6.7.5 Summary of bromodomain interactions . . . 152
6.8 Conclusions . . . 154
6.8.1 Reasons to use LINTools . . . 155
6.8.2 Future directions for LINTools development . . . 156

7 Conclusions and future directions **158**
7.1 Summary and conclusions . . . 158
7.2 Future directions . . . 162

A Sequence alignments for homology modelling **164**

B NBD dynamics in the atomistic simulations of P-gp homology models and their templates **171**

C Convergence of PMF profiles **174**

Bibliography **178**

List of Figures

1.1 P-glycoprotein structure 7
1.2 ATP binding site 9
1.3 Inward-open conformations of P-gp 11
1.4 The drug binding site of P-gp 13
1.5 The models of ATPase cycle 18
1.6 The transport cycle 19
1.7 The substrate transport models 21

3.1 Homology models 60
3.2 The dynamics of homology model from *C. elegans* ABCB1 crystal structure 62
3.3 The dynamics of homology model from *C. merolae* crystal structure . 63
3.4 The dynamics of homology model from refined mouse ABCB1 crystal structure 64
3.5 The dynamics of ABCB1 crystals 66
3.6 The dynamics of ATP-Mg^{2+} bound P-glycoprotein 69
3.7 Entrance gates 71
3.8 Distances between TMHs forming substrate entrance gates in nucleotide-free P-gp simulations 72
3.9 Distances between TMHs forming substrate entrance gates in ATP and Mg^{2+} bound P-gp simulations 73

4.1 Entrance pathways for amitriptyline 79
4.2 Amitriptyline interactions with protein during SMD simulations . . . 80
4.3 Amitriptyline free diffusion 82
4.4 Consensus docking of amitriptyline 85
4.5 Unbinding pathways from PELE simulations 88

5.1 Lipids present in the bilayer 96
5.2 Lipid diffusion in the bilayer 99
5.3 Lipid distribution 100
5.4 P-glycoprotein in complex bilayer 102
5.5 Radial distribution functions of lipids 104
5.6 Lipid ordering around P-gp 105
5.7 Lipid diffusion in coarse grained simulations 106

5.8 Potential cholesterol binding sites 108
5.9 PMF of cholesterol binding 114
5.10 Lipid diffusion in atomistic simulations 116
5.11 The dynamics of P-gp in complex bilayer 117
5.12 The distances between intramembrane and cytoplasmic portions of substrate entrance gates 119

6.1 RMS fluctuations of the ligand 131
6.2 Conversion from three dimensional coordinates to 2D diagram 134
6.3 Method of residue placement 136
6.4 An example of a final image created by LINTools 138
6.5 Example of residue information file 138
6.6 LINTools logo 140
6.7 Free diffusion of amitriptyline in P-gp central cavity 143
6.8 Amitriptyline docked in the central cavity of P-gp 145
6.9 Alprazolam interactions with BRD4(1) 148
6.10 3,5-dimethylisoxazole derivative interactions with BRD4(1) 149
6.11 RVX-208 interactions with BRD4(1) 151
6.12 RVX-OH interactions with BRD4(1) 151
6.13 BI2356 interactions with BRD4(1) 153
6.14 TG101348 interactions with BRD4(1) 153

A.1 The sequence alignment between *C. elegans* and human P-glycoprotein 166
A.2 The sequence alignment between *C. merolae* and human P-glycoprotein 168
A.3 The sequence alignment between mouse and human P-glycoprotein . 170
B.1 The distance between NBDs of the homology model from the *C. elegans* crystal structure 172
B.2 The distance between NBDs of the homology model from the refined mouse crystal structure 172
B.3 The distance between NBDs of the homology model from the *C. merolae* crystal structure 173
B.4 The distance between NBDs of the homology model from the refined mouse crystal structure 173
C.1 Histograms of umbrella sampling windows 175
C.2 Evolution of PMF profiles with time 176

List of Tables

1.1 Overview of P-glycoprotein homologue crystal structures 8

3.1 Overview of human homology models of P-glycoprotein 56

4.1 Work required to enter the central cavity 78

5.1 The composition of the complex asymmetric bilayer 96

6.1 LINTools options and flags 127
6.2 The styles of residue plots 133

List of Abbreviations, Acronyms and Symbols

ADME	Absorbtion, distribution, metabolism and excretion
ADP	Adenosine diphosphate
ATP	adenosine triphosphate
BBB	Blood-brain barrier
CARC	Inverted cholesterol recognition/interaction amino acid consensus (CRAC) motif
CG	Coarse grained
CG	Coarse grained
CHOL	cholesterol
CNS	Central nervous system
COM	center of mass
CRAC	cholesterol recognition/interaction amino acid consensus
ICH	Intra-cellular coupling helix
IQR	Interquartile range
LJ	Lennard-Jones
MD	Molecular Dynamics
MDR	Multidrug resistance
NBD	Nucleotide binding domain
P-gp	P-glycoprotein
PBC	Periodic boundary conditions
PC	phosphatidylcholine
PE	phosphatidylethanolamine

PME	Particle Mesh Ewald
PMF	Potential of mean force
POPC	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine
POPS	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-L-serine
PPCS	Sphingomyelin
PS	phosphatidylserine
QM	Quantum Mechanical
RAMD	Random accelerated molecular dynamics
RMSD	Root mean square deviation
SAR	Structure-activity relationship
SEM	Standard error of the mean
SMD	Steered molecular dynamics
SVG	Scalable Vector Graphic
TMD	Transmembrane domain
WHAM	Weighted histogram analysis method

1

Introduction

1.1 Multidrug resistance

Multidrug resistance (MDR) can be defined as insensitivity to several structurally and functionally different drugs even if sensitivity was previously present. Antimicrobial resistance is considered a serious threat to the public health worldwide since almost all agents capable of infection (e.g. bacteria, fungi, parasites and viruses) have shown high levels of MDR[1]. This has lead to increased mortality and morbidity from diseases such as tuberculosis, pneumonia, HIV, malaria and influenza, indicating that MDR

is a global concern[1].

Amongst the several possible reasons for cancer treatment failure (e.g. variations in ADME , tumour location and vasculature) resistance to the cytotoxic agents used in therapy is highly prominent. Due to nearly universal resistance to any single compound, therapy consisting of combination of multiple drugs is frequently used[2]. However, simultaneous resistance to different drugs with various targets greatly reduces the efficacy of this type of therapy. Multidrug resistance can be an intrinsic property of cancer cells, observed once the cancer is discovered, or acquired through chemotherapy and exhibits itself during the course of treatment or upon recurrence.

Various changes in the cancer cells can cause MDR by affecting the capability of cytotoxic drugs. These include alterations in drug metabolism and cell cycle, enhanced DNA repair and blocked apoptosis, allowing the cancer cells to evade the treatment. While reduced uptake of water-soluble drugs is caused by removal of the transporters required for import, the bioavailability of lipophilic drugs is lowered by energy-dependent pumps.

The molecular mechanisms that cause multidrug resistance are often characterised by the developmental properties of cancer cells and tumours. The vascular network that surrounds tumours is often structurally and functionally abnormal due to the imbalance of angiogenic regulators. The resulting blood flow leads to acidity and can induce hypoxia in some regions of the tumour[3], which in turn can lead to multidrug resistance through various mechanisms described above[4].

Increased efflux of hydrophobic drugs by "multidrug pumps" is the MDR mechanism that has been most widely studied due to its frequent occurrence. The transporters involved in this process belong to five different families - ATP-binding cassette (ABC) transporters, multidrug and toxic compound extrusion (MATE) transporters, major facilitator superfamily (MFS) transporters, resistance nodulation division (RND)

transporters, and small multidrug resistance (SMR) transporters[5]. Instead of having narrow substrate specificity, they are able to transport a very diverse set of seemingly unrelated chemicals making them challenging targets for drug discovery.

A lot of research has been focused on ABC transporter mediated MDR (see[6;7]) due to the important roles these proteins play in cellular processes that require transport of substrates across membranes, as well as the absorption, distribution and excretion of drugs.

1.1.1 ABC transporters in multidrug resistance

The ABC protein family consists of conserved proteins that contain the ATP-binding cassette (ABC) domain that gives the family its name and can be classified as importers and exporters. Both of these classes require ATP hydrolysis for substrate transport. The exporters, including P-glycoprotein, consist of transmembrane domains (TMDs) and nucleotide binding domains (NBD) that bind and hydrolyse ATP[8]. The transmembrane domain sequence and topology is not conserved across the transporters, while the NBDs are generally common to all transporters and have several highly conserved sequence motifs. In humans, 48 genes encode proteins from this superfamily that has been divided in seven subfamilies labelled A-G[9].

In a physiological context, ABC transporters can form a complex network involved in drug detoxification and protection of tissues from potentially harmful xenobiotics. This role can be deduced from the expression pattern of ABC transporters, namely their prominence on the barriers to sensitive sites such as the brain and testis, as well as the epithelial cells of liver, kidneys and interstines[10]. However, overexpression of some ABC transporters in cancer cells has been linked with multidrug resistance. In particular, P-glycoprotein (P-gp/ABCB1), multidrug resistance associated protein-1 (MRP1/ABCC1), and breast cancer resistant protein (ABCG2/BCRP) are amongst

the most extensively studied ABC proteins due to their involvement in multidrug resistance. At least 12 ABC transporters from four subfamilies are involved in multidrug resistance in cancer cell lines[6].

1.2 P-glycoprotein

The importance of active drug transport was realised when MDR Ehrlich ascites cells were observed to lower their intracellular daunorubicin concentration through active extrusion of the drug[11]. In 1976, a 170 kDa protein was detected as the main factor that changed the cellular permeability to drugs and it is now known as P-glycoprotein[12]. Over the next decade the transporter and its role in human cancer multidrug resistance were further characterised. It was shown that P-gp could export a structurally diverse set of chemical compounds that were generally lipophilic and proposed that its main physiological function is to protect sensitive sites such as the brain and kidneys from poisons[13].

Hundreds of P-gp substrates have been identified over the years, which include many clinically important drugs and naturally occurring compounds. The basal ATPase activity of P-gp is often stimulated by substrates and ATP hydrolysis is used to power the drug transport against the concentration gradient[14]. The transporter is expressed in epithelia of various tissues that could potentially come into contact with xenobiotics, such as the colon, small intestine, liver, pancreas, kidney, uterus and placenta[15]. Additional sites of expression include the barriers to highly sensitive tissues such as brain and testis[16].

1.2.1 P-glycoprotein in the blood-brain barrier

P-glycoprotein can transport a wide variety of hydrophobic organic molecules ranging in size from 300 to 2000 Da including vinca alkaloids, anthracyclines and taxanes[17]. This ability to extrude cytotoxic drugs from a cell is the main cause of P-gp mediated drug resistance in cancer. However, it can also potentially transport various unrelated compounds across the blood-brain barrier (BBB) thereby reducing bioavailability in the brain and causing treatment failures for various central nervous system (CNS) disorders. This is likely since drug delivery to the brain can be increased if coadministered together with an effective P-gp inhibitor, potentially increasing the efficacy of treatment for CNS diseases, AIDS or brain tumours[18].

The transport of antidepressant imipramime across BBB was improved by inhibition of P-gp by verapamil[19]. This study highlights P-gp potential involvment in treatment-resistant depression[20] which could be caused by limited intracerebral concentrations of antidepressants. Not all CNS acting drugs are subject to P-gp efflux, potentially explaining the clinical success of clozapine, a non-substrate of P-gp[21]. Clozapine is remarked as one of the most effective antipsychotic drugs[22] with high efficacy in treatment-resistant schizophrenia[23].

1.2.2 P-glycoprotein as a lipid flippase

Interestingly, xenobiotics are only some of the substrates P-gp can transport. It has been shown to extrude amyloid beta peptide from the CNS, resulting in its involvement in normal ageing and Alzheimer's disease. Decreased P-gp expression on the BBB is observed in aging brains for both normal and diseased specimens[24]. P-gp is encoded by three genes in rodent and two genes in human, each gene classified as a separate isoform of the protein[25]. P-gp Class III isoforms (human MDR3, mouse

mdr2) are known to be phospholipid flippases that are phosphatidylcholine (PC) specific[26] while the Class I isoforms (human MDR1, mouse mdr1a and mdr1b) are primarily drug transporters. Due to the high sequence similarity between the isoforms (78%), the drug-transporting isoform of P-gp was suspected to be able to function as a lipid flippase[27]. Sharom *et al.* showed that purified P-glycoprotein reconstituted in proteoliposomes can transport both short-chain and long-chain nitrobenzo-2-oxa-1,3-diazole labeled phospholipids[27]. Simple glycosphingolipids were also transported by the reconstituted P-gp[28]. Both activities required ATP hydrolysis and were inhibited by the phosphate analog, ortho-vanadate. Since flippase activity was inhibited by known P-gp substrates in a concentration-dependant manner, it was proposed that drugs and lipids are transported in a very similar manner[29].

However, the lipid flippase function is unlikely to be the main task of P-gp, since the rate of flipping is relatively low[27;28]. This could be explained by the activity of the plasma membrane aminophospholipid translocase, ATP8a1, a P4-ATPase that translocates phosphatidylethanolamine (PE) and phosphatidylserine (PS) to the inner leaflet to maintain the bilayer asymmetry[30]. If these flippases would act at the same rate, it would result in a futile and energy-wasteful cycle of phospholipid flip-flop. P-gp also could not rescue a knockout of MDR3[31], although the differing expression levels of both proteins could account for this effect.

1.2.3 P-glycoprotein structure

Similarly to other ABC exporters P-gp consists of two transmembrane domains and two nucleotide binding domains that bind ATP. Each transmembrane domain is formed by 6 transmembrane helices and together the domains form the substrate binding site[32]. TMD1 and NBD1 forms one half of the transporter and TMD2/NBD2 the other in a N-TMD1-NBD1-TMD2-NBD2-C topology. Since P-gp is a full size

ABC transporter, a flexible linker joins NBD1 and TMD2. The flexible linker region has been proposed to have a role in coupling ligand binding and ATP hydrolysis[33]. Another possibility is that it regulates the substrate specificity[34]. The intra-cellular coupling helices connect TMDs to the NBDs and coordinate ATP binding and hydrolysis with substrate transport[35]. The contact on NBD2 is established through a hydrophobic interface - "ball-and-socket" joint - between intra-cellular coupling helix 2 and the second nucleotide binding domain[36].

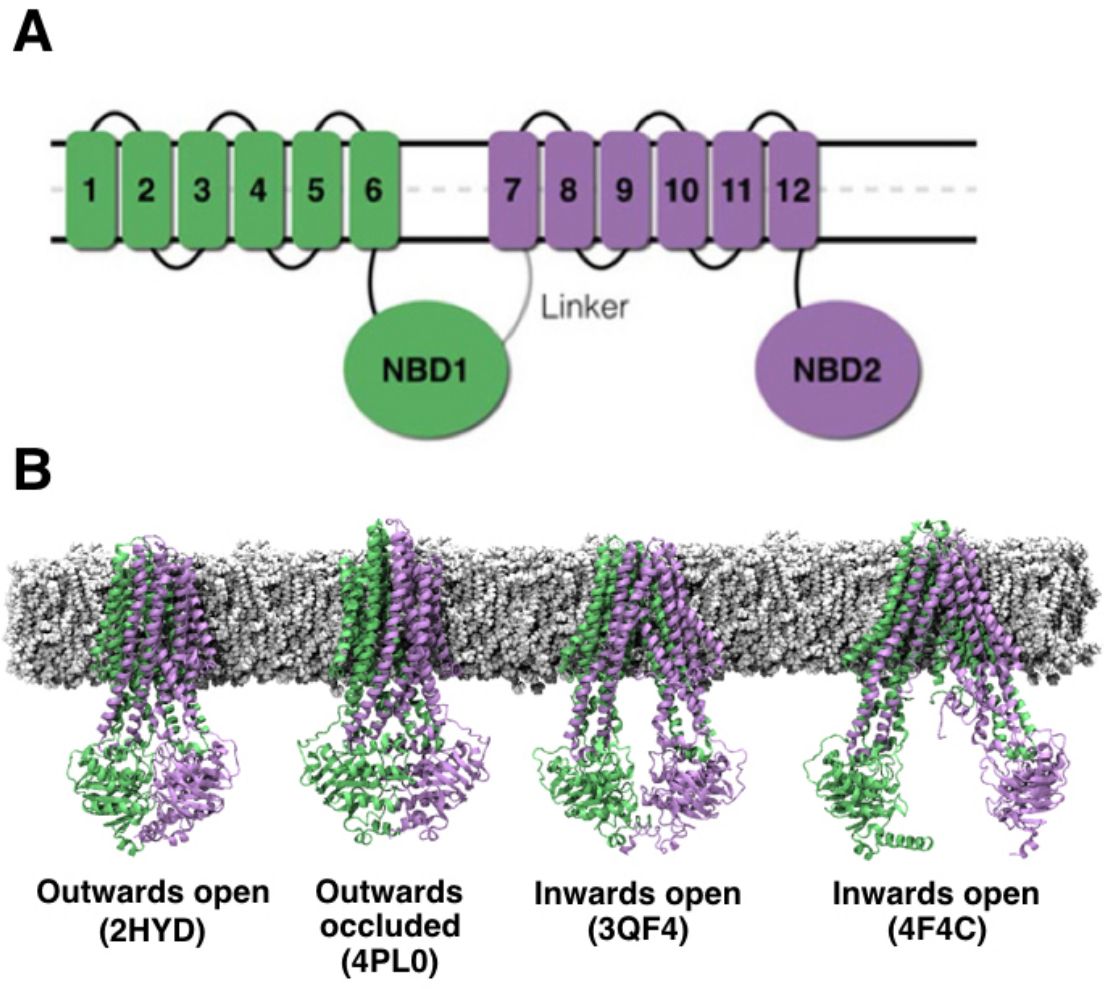


Figure 1.1: (A) The topology of P-glycoprotein with TMD1/NBD1 in green, TMD2/NBD2 in purple and the flexible linker as grey line and parallel lines depict the bilayer. (B) The range of conformational states for ABC proteins captured in crystal structures. The protein is shown in cartoon representation and TMD1/NBD1 is in green and TMD2/NBD2 is in purple. The bilayer lipids are depicted as grey spheres.

The ATP binding site is formed by conserved motifs on both NBDs with the ABC signature sequence (LSGGQ motif, also known as C-loop) on one subunit and Walker A and Walker B motifs on the other (see Figure 1.2). Therefore for ATP hydrolysis to take place NBDs have to be in close proximity to form the "sandwich dimer" (see Figure 1.2 B). The evolutionary conserved ABC signature motif is unique to the ABC protein family[37], while the Walker A (GXXGXGKS/T) and Walker B motifs are commonly found in nucleotide binding proteins[38].

The availability of crystal structures of P-gp and related proteins have given a higher resolution view on the structural features of the protein (Table 1.1). The publication of inwards and outwards open bacterial ABC transporters[35] allowed the application of the alternating access model[39] to this protein family (see Figure 1.1 B). This model predicts that the transporter constantly switches between inwards and outwards open conformations allowing the substrates to enter and exit in these subsequent stages, with the process is powered by ATP hydrolysis.

Table 1.1: Overview of P-glycoprotein homologue crystal structures

Protein	**PDB**	**Resolution (Å)**	**Distance between NBDs (Å)**	**Sequence identity**	**Conformation**	**Reference**
S. aureus Sav1866	2HYD	3.0	8.8	34%	Outwards open	[40]
E. coli MsbA	3B5W	5.3	66.6	37%	Inwards open	[41]
V. cholera MsbA	3B5X	5.5	35.8	34%	Inwards closed	[41]
S. typhimurium MsbA	3B60	3.7	7.8	37%	Outwards open	[41]
Mouse ABCB1	3G5U	3.8	22.4	87%	Inwards open	[32]
T. maritima TM287/288	3QF4	2.9	15.6	30%	Inwards open	[42]
C. elegans ABCB1	4F4C	3.4	40.4	46%	Inwards open	[43]
Mouse ABCB1	4KSB	3.8	38.0	87%	Inwards open	[44]
Human ABCB10	3ZDQ	2.85	24.3	37%	Inwards open	[45]
C. merolae CmABCB1	3WME	2.75	32.3	39%	Inwards open	[46]
Mouse ABCB1 (refined)	4M1M	3.8	22.0	87%	Inwards open	[47]
E. coli McjD	4PL0	2.7	8.4	34%	Outwards occluded	[48]
T. maritima TM287/288	4Q4H	2.53	14.9	30%	Inwards open	[49]

Crystal structures of P-gp homologues. The distance between NBDs is represented by the distance of the Nζ atom on the lysine of the Walker A motif in NBD1 and the Cα atom of the serine in the signature motif in NBD2.

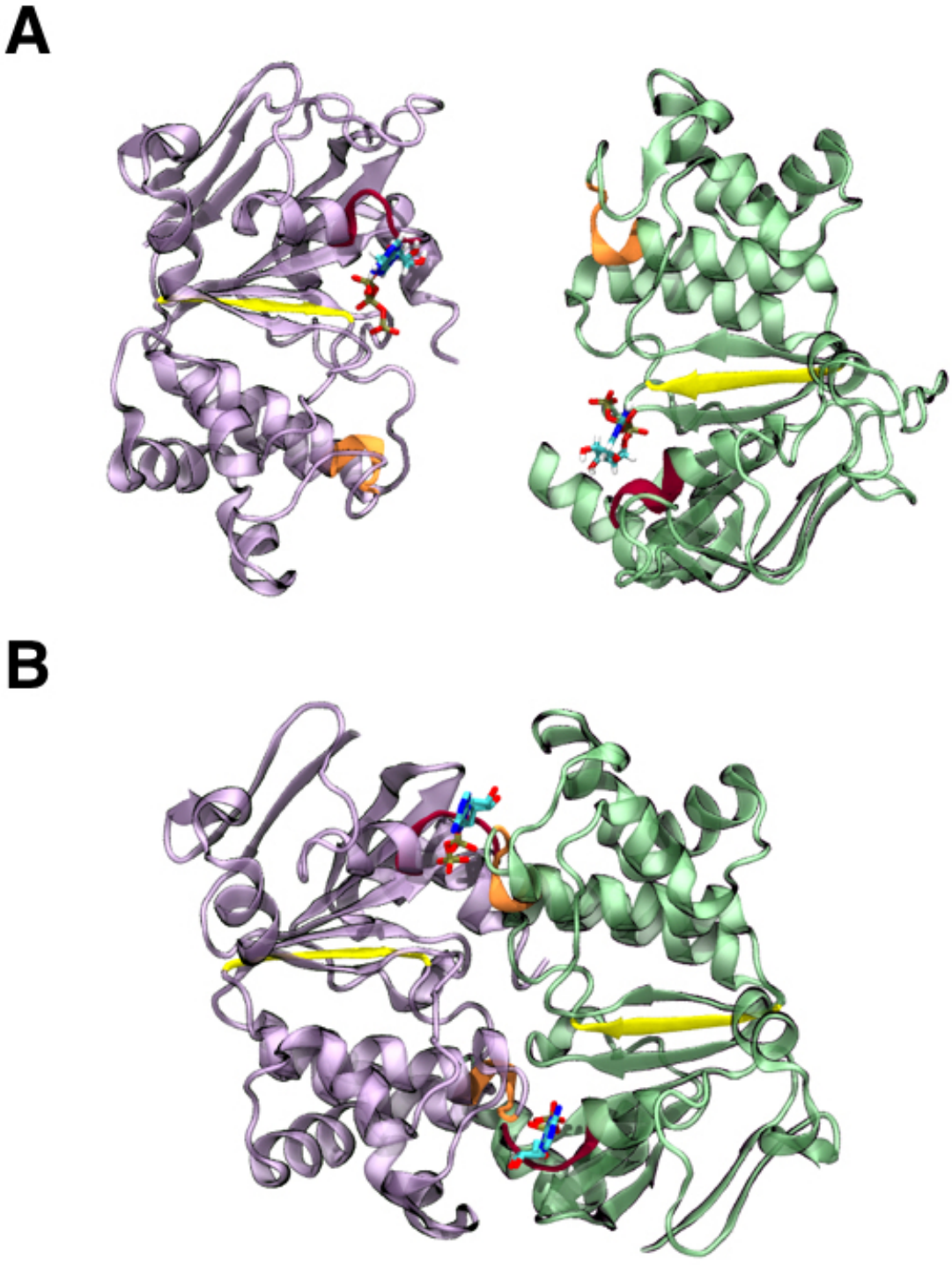


Figure 1.2: (A) The ATP binding site in inwards open mouse crystal structure (PDB 4M1M). The approximate nucleotide coordinates were obtained by overlapping the NBDs from *T. maritima* TM287/288 structure (PDB ID: 3QF4) with nucleotides present. (B) The ATP binding site in outwards open *S. aureus* Sav1866 crystal structure (PDB 2HYD). The protein is shown in cartoon representation with NBD1 in green and NBD2 in purple. The Walker A motif is in red, the Walker B motif is shown in yellow and the Signature motif is shown in orange. The ATP molecules are shown in licorice representation.

1.2.3.1 Crystal structures of proteins related to human P-gp

In 2009, the crystal models of mouse P-gp were published, including both the apo protein [PDB ID 3G5U] and P-gp in complex with two cyclic inhibitors in absence of nucleotide[32]. Due to the high sequence identity (87 %) to human P-gp it has been useful to interpret the existing biochemical data. However, with the publication of the *C. elegans* P-gp crystal structure, some doubts about the correctness of the mouse P-gp crystal model were raised[43]. Interpretation of arginine mutagenesis data suggested that the helix register on TMH5 is possibly wrong in the mouse structure[50]. This was corrected in the refined mouse P-gp crystal structure published in 2014[47].

Although the only available crystal structures of eukaryotic P-gp homologues are in different conformations of an inwards-open state (see Figure 1.3), bacterial ABC transporters have been captured in various conformational states corresponding to different parts of the transport cycle (see Table 1.1). A potential explanation for this disparity could be a difference in the energetic and mechanistic regulation of the conformational changes as shown by cryo-EM of bacterial MsbA and mammalian P-gp[51]. While both transporters showed to be dynamic and move from inwards-open to outwards-open conformations and vice versa, P-gp was predominantly found in inwards-open conformation even in presence of high concentrations of Mg and ATP[51].

The conditions in which these crystals have been obtained has been a subject of discussion, since none of the available crystal structures have been obtained in lipid bilayer with detergents used instead. Recently, it was shown that lipids are required for P-gp ATPase activity[52] and that the large separation between NBDs observed in crystal structures (see Table 1.1) could be caused by purification of proteins in detergents[53]. The "nucleotide-free" state has also been considered to be potentially physiologically irrelevant due to the high concentrations of ATP in the cell[54].

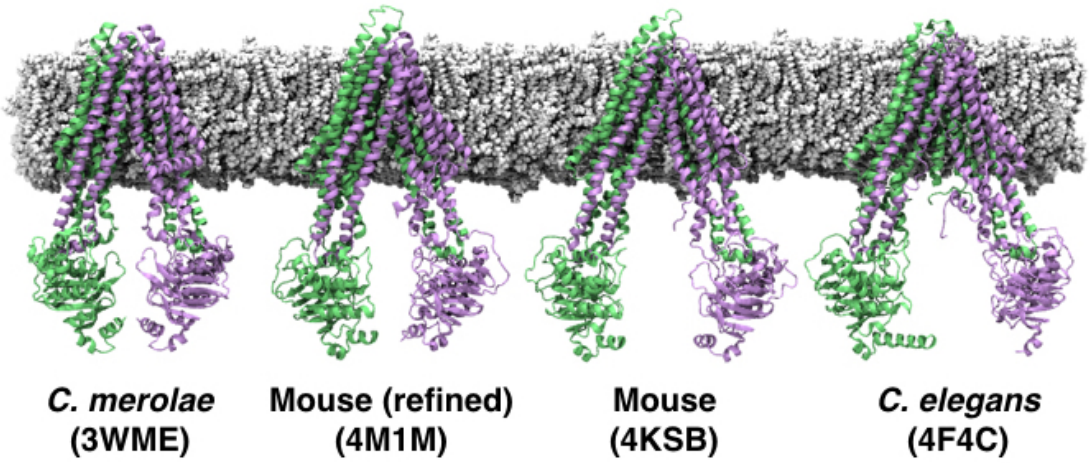


Figure 1.3: The eukaryotic structures of P-gp have been captured in range of inward-open conformations. The main difference between these conformations is the distance between NBDs ranging from 22.0 Å for refined mouse P-gp crystal structure (PDB 4M1M) and 40.4 Å for *C. elegans* P-gp crystal structure (see Table 1.1). The protein is depicted in cartoon representation and TMD1/NBD1 is shown in green, TMD2/NBD2 in purple. The lipid molecules are grey spheres.

1.2.4 P-glycoprotein drug binding site

The crystal structures of mouse P-gp revealed that the substrate binding site is a large central cavity formed by transmembrane helices[32]. The central cavity is lined with aromatic and hydrophobic residues whose involvement in substrate binding has been investigated by many mutagenesis studies[55–58]. These residues are proposed to bind the substrates via hydrophobic and van der Waals interactions. The drug binding site could be accessed through either the inner leaflet of the bilayer or straight from the cytoplasm.

However, there is no clear understanding of the binding modes of P-gp substrates. The binding site has been proposed to be either large and flexible, with the possibility for substrates to bind in secondary binding sites, if the primary sites are not available[59], or consist of several more specific sites[60]. In the initial mouse structure two molecules of the same cyclic peptide could fit in the drug binding site utilising two separate

sites (see Figure 1.4)[32]. This ability could explain some of the promiscuity of the transporter.

1.2.5 P-gp substrates

P-gp substrates vary in their size, structure and function, and can range from small molecules such as organic cations and amino acids to macromolecules such as proteins[61]. Extensive studies have tried to understand the characteristics of the interactions between P-gp and its small substrates, to identify potential substrates and develop more potent inhibitors[61]. Attempts have been made to find the common set of structural and functional features that allow substrates to interact with P-gp[62].

Prior to computational analysis of structure-activity relationship (SAR), simple rules for substrate-inhibitor relationships were tried to be devised. Initially, a requirement for a basic nitrogen atom and two planar aromatic rings was suggested[63], however, a wider screening of compounds revealed the requirement for a basic nitrogen atom not to be a strict rule[64;65]. Even though it is widely accepted that hydrophobic and amphiphatic molecules make up the bulk of P-gp substrates, some relatively hydrophilic compounds such as morphine, can also be weak substrates[62]. In a screening of a hundred chemically diverse compounds, it was demonstrated that a compound can interact with P-gp if it contains either: (a) two electron donor groups with a spatial separation of 2.5 $\pm$ 0.3 Å, or (b) two electron donor groups with a spatial separation of 4.6 $\pm$ 0.6 Å, or (c) three electron donor groups with a spatial separation of the outer two groups of 4.6 $\pm$ 0.6 Å[62]. Other features, such as lipophilicity, hydrogen-bonding ability, molecular weight, and surface area, may also affect substrate binding to P-gp[66;67]. It was found that the partitioning into the lipid membrane is a rate-limiting step in the interactions which might be unsurprising, considering the hydrophobic nature of most P-gp substrates[68].

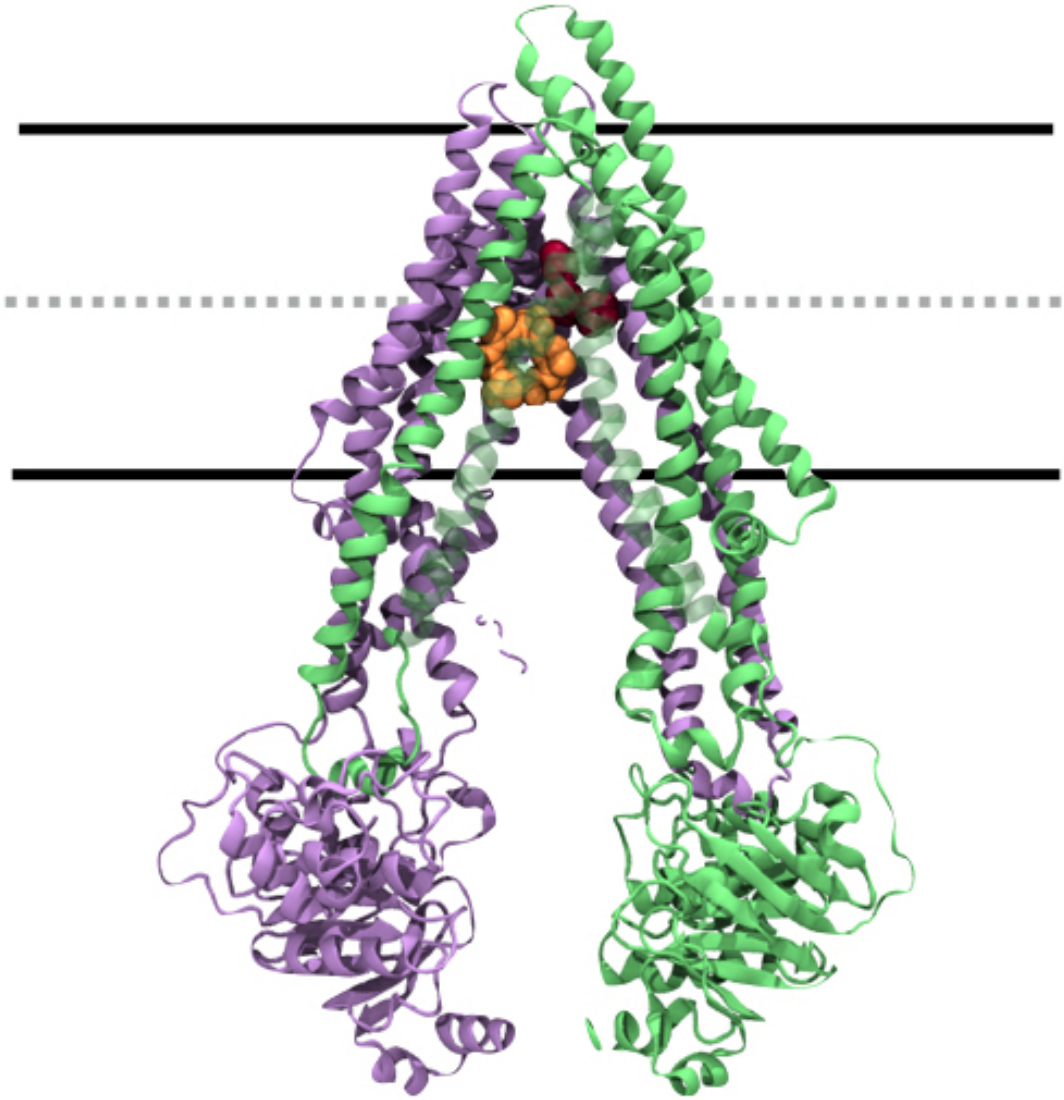


Figure 1.4: The mouse crystal structure (PDB 3G61) with two cyclic inhibitor QZ59-SSS molecules bound. The protein is shown in cartoon representation with TMD1/NBD1 in green and TMD2/NBD2 in purple. Some transmembrane helices are transparent for clarity. The inhibitor molecules are shown as red and orange spheres. The lines depict the approximate location of the lipid bilayer.

In silico models for the prediction of probability of a drug to interact with P-gp are valuable in the early phase of the drug discovery process. The models for predicting substrates tend to use binary classification (substrate/non-substrate), since the data obtained from transport assays often does not present precise IC_{50} values[69]. A large dataset of P-gp substrates/non-substartes was compiled by Li *et al.* (423 substrates, 399 non-substrates)[47]. Analysis of the distribution of eight basic physicochemical properties suggest that the main factors that differentiate P-gp substrates from non-substrates appear to be molecular weight and solubility. Additional analysis show that P-gp substrate-likeness is composed of H-bond acceptors arranged in distinct spatial patterns as well as flexibility of the molecule. When compared to 735 inhibitors, substrates appear to be less hydrophobic and tended to have more H-bond donors.

1.2.6 P-gp modulators and inhibitors

The reversal of the MDR phenotype through interference with P-gp substrate efflux can be caused by a structurally diverse group of compounds known as P-gp modulators (also called inhibitors or chemosensitizers)[70]. Modulators can affect the P-gp substrate transport process through several mechanisms. While some of these compounds (e.g. cyclosporin)[71] are themselves transported and act as competitive substrates, others disrupt transport through tight binding to the drug binding site for prolonged periods (e.g. LY335979)[72]. Another mechanism of inhibition involves interference with ATP hydrolysis through interactions with NBDs. Examples of such compounds include steroids and disulfiram[73]. Drugs that affect the lipid bilayer have also been shown to reverse MDR (e.g. benzyl alcohol, Nonidet P40 or Cremaphor EL)[74] even though they do not interact with P-gp directly.

It has often been advantageous to coadminister P-gp inhibitor together with an anticancer drug to enhance the latter's intracellular accumulation via reduction of P-gp

efflux in a process known as chemosensitization. Numerous compounds are known to possess P-gp inhibitory activity and can reduce the cellular resistance to conventional chemotherapy. The P-gp inhibitors are categorised as first, second and third generation in order of their reducing toxicity, an issue that remains one of the major clinical hurdles.

The first generation of P-gp inhibitors included existing therapeutic agents such as verapamil[75] and other calcium channel blockers, as well as chlorpromazine and prochlorperazine, indole alkaloids, the anti-malarial quinine and the anti-arrhythmic quinidine[76]. The most effective first generation inhibitor proved to be cyclosporin-A, a commonly used immunosuppressant. However, the first generation inhibitors inhibit P-gp in concentrations that is higher than their therapeutic window and therefore cause serious immunosuppresive and cardiovascular side effects[77]. In addition, some of the compounds possess an inhibitory action on CYP3A, potentially causing unfavourable drug-drug interactions.

The search for less toxic compounds resulted in the more potent and less toxic second generation P-gp inhibitors[78–80] - e.g. dexverapamil (R-enantiomer of verapamil), emopamil, and valspodar (non-immunosuppressive analog of cyclosporin-A). However, even though these drugs were very effective at lower concentrations, they still had an effect on CYP3A enzyme activity.

Structure-activity relationship and combinatorial chemistry approaches were used to discover the third generation P-gp inhibitors with inhibitory action in the nanomolar range (20-100 nM). While several of these compounds block only P-gp (e.g. zosuquidar (LY335979)[81], elacridar (GF120918)[82]), others are dual blockers of P-gp and MRP1 (timcodar (VX-853)and biricodar (VX-710))[83]. One of the most promising compounds - tariquidar (XR9576) - is a dual P-gp and BCRP inhibitor[84]. However, in clinical trials they failed to improve the patient outcome[85].

Some chemicals of natural origin and plant extracts have also been reported to act as P-gp modulators (for a review see[86]). Plant derived polyphenolic compounds, such as flavonoids and stilbenes and their synthetic derivatives have been shown to modulate the activity of several ABC transporters involved in MDR[87]. Other classes of naturally occurring compounds such as coumarins, terpenoids, alkaloids and saponins are also been reported as modulators of P-gp transport activity. These compounds do not possess the disadvantageous toxicity of other P-gp modulators, however, they still have to be tested in clinical trails.

The clinical trials have so far provided only disappointing results and in order to create more successful P-gp modulators, we will need to acquire more knowledge about the intricacies of the P-gp transport cycle.

1.3 P-glycoprotein transport cycle

The drug transport by P-gp is powered by ATP hydrolysis at the NBDs that form the composite ATP binding site. The crystal structures show two major conformations with drug binding site in the nucleotide-free structures open to cytoplasm (see Figure 1.1B), while the nucleotide binding leads to formation of the ATP binding sites and close proximity of NBDs. In this conformation, the TMDs have flipped and the drug binding site is open to the extracellular space. This switch leads to reduced affinity for the substrate at the drug binding site and subsequent release. The gates are never open to both cytoplasm and extracellular space, however, crystal structures of bacterial ABC transporters hint to an existence of occluded states with both gates closed (see Figure 1.1B). Therefore, it might be possible that the conformational pathway of P-gp transport cycle involve transition from inwards-open to outwards-open conformations with intermediate occluded states.

However, the molecular mechanisms that link ATP binding and hydrolysis with conformational changes and transport are still poorly understood. ATP can be hydrolysed at either binding site and it is unclear whether P-gp hydrolyses one or both ATP molecules. Similarly, the sequence of conformational and catalytic events that happen during the transport cycle is a subject of discussion. Experiments that have tried to understand the molecular mechanisms have resulted in two conflicting models - Switch and Constant Contact. The Constant Contact model suggested a more flexible NBD arrangement, where only one NBD undergoes ATP hydrolysis, while the other one remains closed with the ATP molecule present in the binding site (see Figure 1.5 A)[88]. As the NBDs never completely dissociate, it is the main difference from the Switch model. The Switch model proposed a closed NBD dimer with two ATP molecules bound that dissociates into open dimer after ATP hydrolysis (see Figure 1.5 B)[89].

A recent study showed that single Walker A mutants are capable of conformational changes, while a double mutant is trapped in the inwards-open conformation[91]. This suggests that only one hydrolysis event might be required for the transport cycle and the hydrolysis events are random, i.e. the hydrolysis does not happen in a strictly alternating order. The ATP hydrolysis is preceded by affinity switch at the substrate binding site and conformational change to outwards-open conformation caused by ATP binding[91]. Crosslinking studies have shown that NBDs can function even if clamped together in a closed conformation[92]. All these studies support the Constant Contact model and are at times in disagreement with the available structural data, in particular, the wide opening of NBDs observed in mouse and other crystal structures. A potential explanation for this discrepancy has been provided from studies in lipid and detergent environments[53].

In conclusion, the transport cycle of P-gp starts with binding of substrate to the

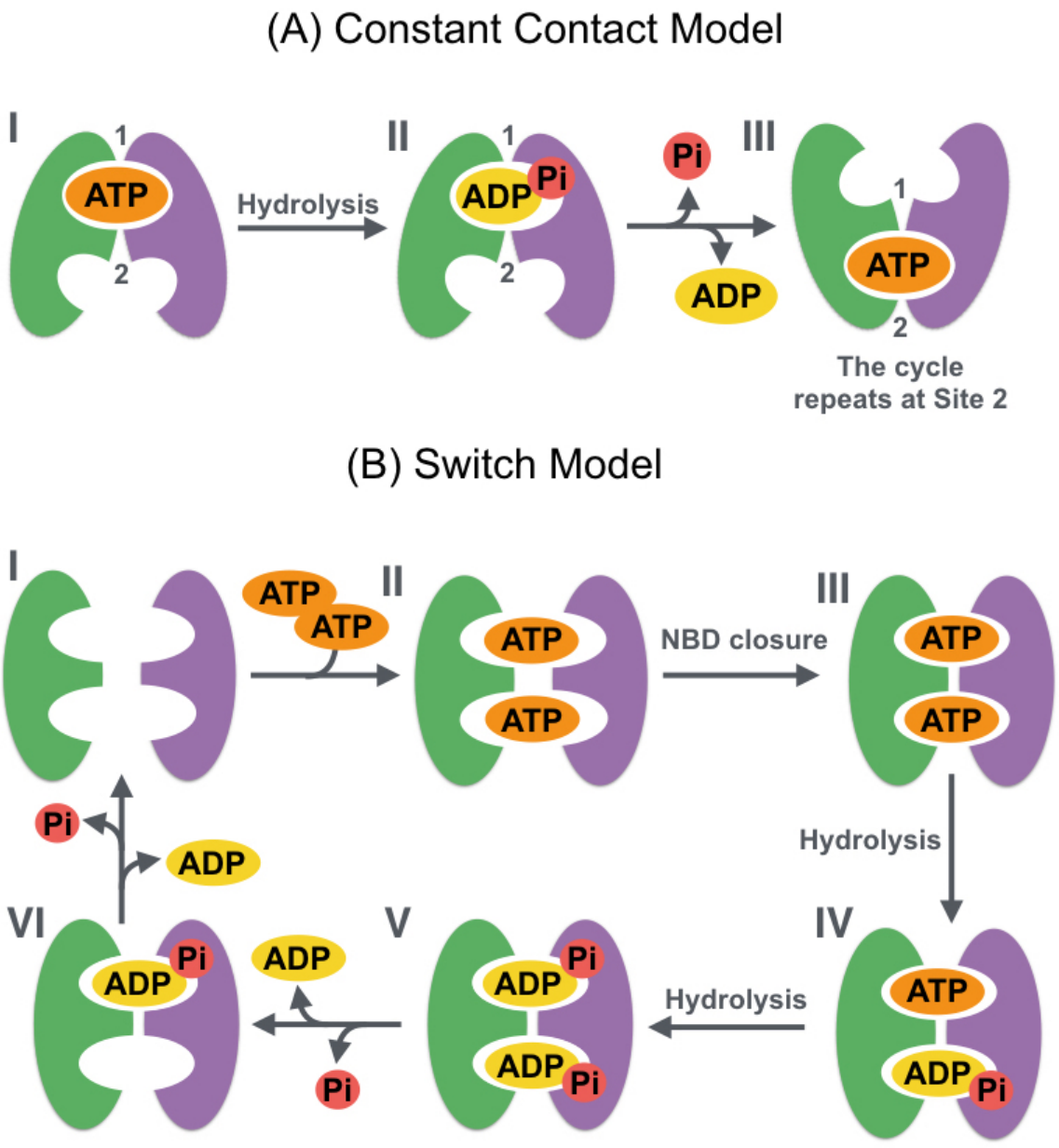


Figure 1.5: The models of ATPase cycle.(A) Constant Contact model and (B) Switch model. The main difference between the models is that in Constant Contact model the NBDs remain in contact throughout the cycle, while the Switch model requires dissociation (steps II and III in B). Adapted from[90].

binding site formed by TMDs, followed by ATP binding and conformational change to outwards-open state. The affinity for substrate has switched from high to low and it is released from binding site. Subsequent hydrolysis of one of the ATP molecules in a randomly selected site leads to conformational change back to inwards-open state and release of adenosine diphosphate (ADP) and inorganic phosphate (see Figure 1.6).

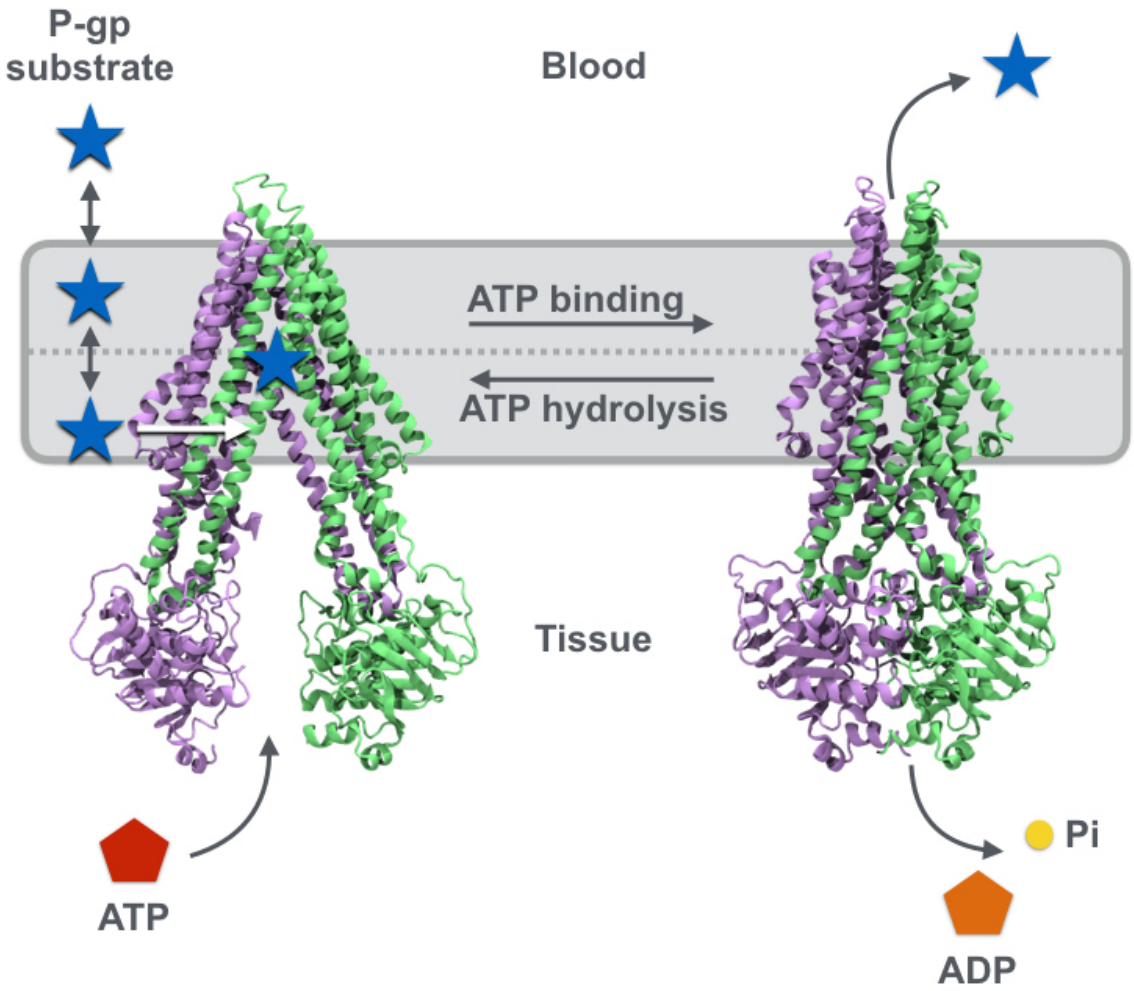


Figure 1.6: The P-glycoprotein transport cycle as discussed in the main text, also including the substrate entrance and binding steps. The P-gp is proposed to cycle between inwards-open (refined mouse crystal structure PDB 4M1M) and outwards-open (Sav1866 crystal structure PDB 2HYD) conformations. The protein is shown in cartoon representation with TMD1/NBD1 in green and TMD2/NBD2 in purple. The bilayer is depicted in grey.

1.3.1 Substrate transport

The crystal structure of mouse P-gp suggested that substrates could access the drug binding site from either lipid bilayer or cytoplasm. The gaps formed by TMH4/6 and TMH10/12 are positioned partly in the inner leaflet of lipid bilayer and, by extension of the transmembrane helices, in the cytoplasm as well. While several models for substrate transport have been proposed, no certainty has been achieved in this important question (see Figure 1.7). Two of the models, named as "hydrophobic vacuum cleaner" and "flippase" models[93], suggest that substrates diffuse into the inner leaflet of the bilayer and from there reach the binding site on P-gp. The difference in these models lies in the export step of the transport cycle - the drugs could be either extruded to the extracellular space as suggested by "hydrophobic vacuum cleaner" model or diffuse in the outer leaflet first. On the other hand, the "pore" model suggests that drug uptake straight from the cytoplasm.

The consensus for "hydrophobic vacuum cleaner" model is that the drugs partition into the lipid bilayer and are recognised by P-gp in the inner leaflet of the bilayer, then extruded directly to the extracellular space[93]. The action of the multidrug pump creates a concentration gradient across the lipid bilayer with the higher concentrations in the extracellular space and can potentially intercept the compounds before they have reached the cytosol, thereby protecting the cell[94]. The "hydrophobic vacuum cleaner" is the most widely accepted model of substrate transport due to its agreement with structural data from the mouse P-gp and other ABC transporter structures[32;43;46]. *C. elegans* and *C. merolae* structures show unwinding of TMH 10 at the level of inner leaflet of bilayer and it has been suggested that this feature could act as a hinge for the substrate entrance gate[43;46]. The model has also been supported by experimental evidence of P-gp substrates being extracted from the inner leaflet of the bilayer[95]. Similarly, the vast majority of P-glycoprotein substrates are lipophilic and MD and

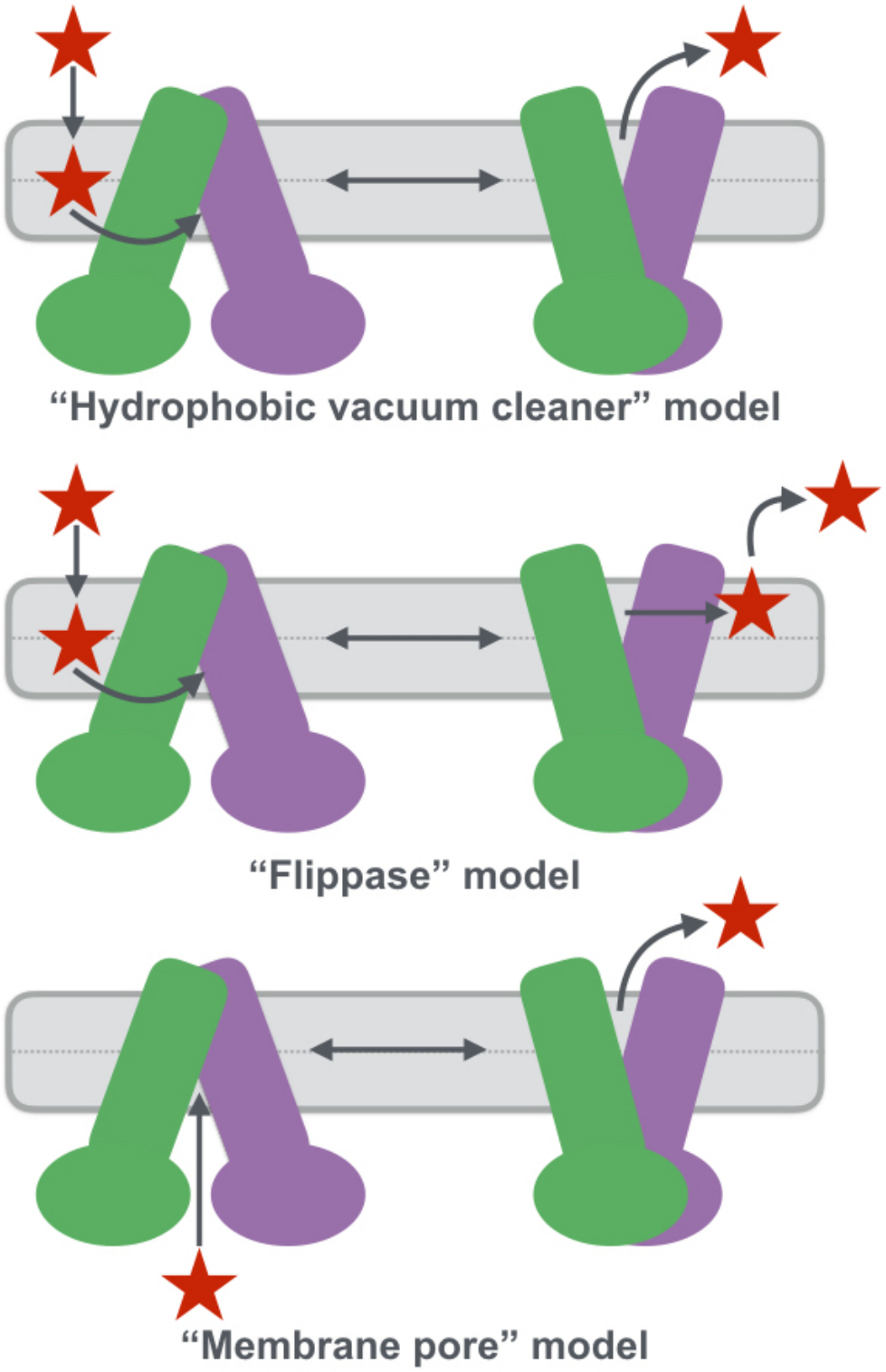


Figure 1.7: The models of substrate transport. The "hydrophobic vacuum cleaner" and "flippase" models suggest that the substrate enters by partitioning into the lipid bilayer and then diffusing towards P-gp and the substrates are obtained from the lipid bilayer. The difference between these models concerns the exit of a substrate. On the other hand, the "membrane pore" model suggests that substrates can bind directly from the cytoplasm. The TMD1/NBD1 is depicted in green, TMD2/NBD2 in purple and the bilayer is grey. The substrate molecule is shown as a red star.

NMR studies have shown that the substrates concentrate between lipid headgroups and the first few atoms of the lipid acyl chains[96].

However, it could be argued that the structural data can support both "hydrophobic vacuum cleaner" and "pore" models, since the transmembrane helices that most likely form the drug entrance gates (TMH4/6 and TMH10/12) extend to the cytoplasm and potentially provide direct access to the drug binding site from either membrane or cytoplasm.

1.4 Multidrug efflux pump and the lipid bilayer

It has been known for some time that the surrounding lipid bilayer has a large effect on P-gp-mediated MDR[97]. The hydrophobic substrates are affected through trans-bilayer movement and lipid partitioning, which is known to be the rate-limiting step in drug efflux[68]. Local lipid bilayer environment can influence P-gp stability and conformation, as well as its function at all stages of the transport cycle. As discussed previously, drugs that can affect properties of the lipid bilayer can reverse the MDR phenotype[74]. Conformational changes dependant on the surrounding environment have been observed, for example, differences in conformation dependant on whether P-gp is surrounded by lipids or detergents[53]. This is of particular importance for crystal structures since most crystals are obtained by stripping them from the lipid bilayer and exchanging lipids for detergents. Similarly, it has been shown that the lipid bilayer is required for ATPase activity of P-gp most likely by introducing structural changes to the protein[52].

Lipids that are closely associated to P-gp have been identified through metabolic labelling of the extracted protein and, surprisingly, these are enriched in PE and PS, but contain relatively little amounts of PC and SM lipids[98]. This is in contrast to the

composition of the mammalian plasma membrane, of which the latter lipid species are major components and suggests a selectivity in the phospholipid environment. It has to be noted that PS and PE are mostly found in the inner leaflet of the bilayer. A mass spectrometry study of P-gp interactions with lipids in detergent micelles showed that only one or two detergent molecules were retained in the gas phase, potentially in the substrate binding site[99]. Simultaneous binding of lipids, drugs and nucleotides was observed and it was proposed that lipids could interact with P-gp in either the drug binding site or at the protein-bilayer interface.

1.4.1 P-glycoprotein interactions with sterols

Cholesterol is a major component (up to 50 %) of the mammalian plasma membrane and is well known to alter its packing, order and fluidity. Therefore, it is perhaps not surprising that many studies report a link between cholesterol and P-gp function. Protein function can be regulated through two nonexclusive mechanisms - specific interaction with a bound lipid or nonspecific, lipid-induced changes to the membrane environment of the protein[100]. In the case of cholesterol, these effects can be hard to distinguish due to its effect on the membrane environment. However, several high resolution crystal structures of integral proteins have revealed cholesterol bound at specific sites on the surface of the protein (see[100]).

The amount of cholesterol present in the bilayer affects both the basal and drug-stimulated ATPase activity of P-glycoprotein in a dose-dependant fashion[101]. The partitioning of the drugs in the phospholipid bilayer is significantly reduced by the presence of cholesterol, thereby affecting P-gp transport function[102]. P-gp co-eluted with [^{3}H]cholesterol which also altered drug interactions with the transporter, suggesting a direct interaction between the sterol and P-glycoprotein[103]. Interestingly, this effect was size-dependant, since cholesterol improved binding affinity for smaller

(>500 Da) drug molecules, but not for larger ones (800-900 Da). These studies lead to the proposition of "cholesterol fill-in" model which suggests that cholesterol is present in the substrate binding site simultaneously with the smaller substrates and by taking up the extra space promotes binding of these molecules[104]. A recent study has looked at interactions of purified P-gp with fluorescent and nonfluorescent cholesterol derivatives and suggests that sterols may directly interact with the transporter potentially via specific binding regions[105].

Even though P-gp has shown some ability to flip fluorescent derivates of phospholipids, it is unlikely to function as a cholesterol flippase[102]. Since P-gp is highly active in phospholipid bilayers alone[106], the presence of cholesterol in a lipid bilayer may not be a strict requirement for P-gp to function normally. However, the P-gp modulating ability of cholesterol is clear from the evidence, even if the exact nature of those interactions remains yet unknown.

1.5 Molecular modelling studies of P-glycoprotein

1.5.1 Homology models of human P-glycoprotein

At the time of writing of this thesis (September 2016) there is no crystal structure of human P-gp available, however, most of the biochemical data has been obtained for the human protein. However, structures of related proteins can and have been used for homology modelling of the human protein. The homology models can be used to generate experimentally testable structure-based hypotheses. Although some studies have only looked at the static snapshots of the homology models to either gain structure-activity relationship (SAR) data or compare different conformational states, others have incorporated dynamics.

Existing SAR data is often rationalized using homology models and P-gp is no exception, especially since it is an important target for drug discovery[107;108]. For example, homology modelling in combination with docking was used to understand P-gp sensitivity to aminated thioxanthone derivates. These compounds with cytotoxic properties were shown to inhibit P-gp and led to an improved *in vitro* efficacy in a doxorubicin-resistant cell line[109]. Other similar studies have examined the binding of potential novel inhibitors such as 6-(methylsulfinyl)hexyl isothiocyanate[110] and guanidine alkaloid[111].They found in each case that these drugs bind in similar locations to known P-gp inhibitors.

The binding sites of the third generation inhibitors, tariquidar and elacridar, were investigated using homology models based on mouse P-gp and Sav1866 crystal structures. This study showed that their binding sites might overlap with the Hoechst 33342 and rhodamine 123 binding sites, potentially interacting with both. Although in most cases studies focus on the drug binding site in the transmembrane cavity, there have been studies that investigate compounds such as piperine, that might act via the NBD[112].

1.5.2 Studies that involve dynamics

However, studies of static structures do not reveal the dynamics of protein-ligand interactions or conformational changes - subjects that are of a great interest in regard to P-glycoprotein. This issue can be circumvented using molecular dynamics (MD) simulations. The binding of anticancer compounds docked in a homology model from *C. elegans* template was investigated using MD simulations and free energy calculations[113]. One of the compounds, NSC745689, did not show any stable interactions with the binding site, suggesting that it is not a substrate of P-gp and this was confirmed with *in vitro* experiments. The interactions of C_{60} fullerene, a potential drug

carrier, with P-gp were investigated by a combination of *in vitro* methods and MD simulations[114].

More ambitiously, attempts to predict conformational changes have also been made. The transport mechanism of P-gp was investigated using targeted MD on a homology model based on Sav1866 to obtain structural information about the intermediate states not available as crystal structures[115]. Docking of 21 substrates and inhibitors on the transitional structures mostly agreed with biochemical evidence for the drug-binding site. These structures were used in a later study to screen for selective compounds that would inhibit ATP hydrolysis by binding to NBDs and not targeting the substrate binding site in the transmembrane domains[116].

The communication interface between the NBDs and the TMDs was investigated by homology modelling and MD showing that conserved motifs on NBDs and the intracellular coupling helices form hydrogen bond networks[117]. Molecular docking was shown to be able to discriminate between P-gp substrates, modulators and inhibitors, however, care must be taken when interpreting the possible binding site position of a drug[118]. The transport mechanisms of mouse P-gp and Sav1866 were compared by investigating the inward and outward open models of P-gp with emphasis on the allosteric effect of ATP on the conformational dynamics[119].

Even with the appearance of more crystal structures, it is unlikely that they will provide a complete picture of P-gp transport cycle and, therefore, modelling can provide useful and experimentally testable hypotheses. Since there is no available human P-gp crystal structure, one must start with homology modelling. Similarly, more information about the conformational cycle and substrate recognition is likely to come from multidisciplinary studies that make use of modelling and MD simulations to refine our understanding of P-gp.

1.5.3 Molecular dynamics studies of P-gp substrate entrance

Due to the existence of several models of substrate entrance to the binding site, MD studies have attempted to solve this question using a variety of methods, starting with unbiased simulations[120], as well as more biased methods such as random accelerated molecular dynamics (RAMD)[121] and potential of mean force simulations[122;123].

Partial entrance of a lipid molecule from the bilayer, through the gap between TMHs 4/6, has been observed in a molecular dynamics simulation, suggesting this region to be a potential entrance portal that allows direct drug uptake from the bilayer[124]. Simulations of paclitaxel and doxorubicin egress also suggest the same region as the most likely drug entrance site, however, in their simulations, both drugs exited through the cytoplasmic part of the portal[121]. Umbrella sampling simulations of colchicine and tariquidar entrance through the other portal formed by TMH10/12, suggested no free energy penalty for crossing the gate and the movement from the bilayer core into the drug binding site is a downhill energy process[123].

Other studies have attempted to look at the access from cytoplasm and substrate approach to the binding site. Free diffusion of paclitaxel and doxorubicin through the central cavity was observed in unbiased MD simulations[120]. The substrates approached their putative binding sites via obstacle-overcoming process, with the residues in the central cavity showing flexibility to allow further passage of the compounds. Recently, O'Mara and colleagues computed potential of mean force pathways that considered access directly from the cytoplasm for different drugs but concluded that the pathways were not clearly defined[122].

These studies give certain insight in the substrate binding processes, highlighting the apparent flexibility of P-gp and its drug binding site. The molecular modelling approach allows us to investigate substrate binding pathways in atomistic detail, however, the timescales available to MD may still be too small to investigate this

process by unbiased MD simulations. Nevertheless, the advances in the field of biased MD methods could be of considerable help in this regard.

1.6 Protein-ligand interactions in MD

Due to the increased timescales and the possibility to study ever larger macromolecular systems molecular dynamics has become an essential technique in drug discovery. This is particularly true for studies of protein-ligand interactions that often try to predict the binding mode of a particular ligand for a specific biological target. The longer timescales available to MD and the exponential growth in the number of published target structures has allowed the study of protein-ligand interactions on a larger scale and to consider the importance of conformational changes in the drug binding process.

A very common use of molecular dynamics in this context, is to analyse binding site interactions after docking a compound (e.g.[125]). On the other hand, novel target conformations can be obtained from MD simulations and the ligand docked in the states of potential interest[124]. These approaches, although not necessarily complementary, give insight to the conformation-specific protein-ligand interactions, aiding drug discovery.

The protein-ligand complex is subjected to a MD simulation for an appropriate amount of time which may extend into the microsecond scale and several repeats may be required to ensure proper sampling. The resulting trajectory is then analysed to extract information such as distances and interactions between ligand atoms and residues of interest. Additionally, the stability of the complex can be analysed through root-mean-square deviations from a reference, as well as the root-mean-square fluctuations of atom positions. Specific interactions with individual ligand atoms, such as

salt bridges, hydrogen bonding and hydrophobic interactions can also be understood in greater detail from the MD simulation data.

An important use of MD in the context of drug design concerns the estimation of free energy differences ΔG, most importantly, the free energy of binding ΔG^{bind} - the difference between free energy in the ligand-bound state and ligand-free state. Some of the most rigorous methods for free energy calculation using MD are thermodynamic integration (TI)[126] and free energy perturbation (FEP)[127]. Although most accurate amongst free energy methods, these are also most computationally expensive. Alternative, more efficient approaches include linear interaction energy (LIE)[128] and the combination of molecular mechanics (MM) calculations and continuum solvation models such as Poisson-Boltzmann surface area (PBSA)[129] and generalized Born surface area (GBSA)[130].

Another case that exploits the dynamic approach of molecular dynamics, is the study of ligand binding and unbinding pathways using either long unbiased simulations[131] or a plethora of biased methods, such as steered MD[132], funnel metadynamics[133] or random accelerated MD[134] amongst many others (for a review see[135]). The development of open source tools such as PLUMED[136], which is compatible with most simulation engines, allows easier integration of these methods and has contributed to the popularity of this approach. These methods can be used to qualitatively discriminate binders from non-binders[137] or to gain mechanistic insight in the unbinding pathways[138].

As more and more MD studies of protein-ligand interaction appear, the importance of the technique becomes clearer, however, some limitations still exist. Optimal searching and sampling of the protein-ligand system conformational space is still difficult and can limit the accuracy of MD. For transmembrane proteins, their surrounding environment is of extreme importance and must be considered at great length. The

forcefield, a central factor in MD simulations, is yet another limitation, especially owing to negligence of electronic polarization effects due to fixed permanent partial atomic charges, as well as correct parametrization.

1.7 Thesis aims

1.7.1 P-gp substrate entrance pathways

There is a lot of controversy surrounding the models of P-gp substrate transport as the structural and experimental data appear to disagree. The lack of human P-gp structure in any conformation is also a problem for further structural studies and drug development. However, the availability of related structures allows us to make homology models of human protein.

In chapter 3, human homology models from different templates are compared by a score that predicts the absolute quality of the structure. The three best models from this comparison are further subjected to molecular dynamics simulations to assess the stability and conformation of the structures during the simulations. The homology model from a refined mouse structure template is deemed as most fitting for further experiments and its behaviour in apo and ATP-bound states is compared. The closure of intramembrane sections of putative substrate entrance gates suggests that more detailed investigation of the substrate entrance pathways is required.

In chapter 4, steered MD is used to investigate the substrate entrance pathways using the homology model described in chapter 3. Both intramembrane and cytoplasmic entrance routes are explored and the resulting work profiles strongly suggest that the cytoplasmic route through TMH4/6 is favoured by amitriptyline, an antidepressant and P-gp substrate. The SMD simulations were followed by unbiased simulations

where amitriptyline was allowed to explore the drug binding site and an event of free diffusion towards the binding site was observed. To gain more information about the amitriptyline binding site, consensus docking on conformations, extracted from simulations described in chapter 3, was performed. The sites suggested by consensus docking and free diffusion were slightly different, however, they contained common features such as similarity to the rhodamine binding site. Finally, egress of the compound was simulated with the Protein Energy Landscape Exploration method. Similarly to SMD, egress simulations showed preference to the cytoplasmic part of the TMH 4/6 portal.

1.7.2 P-glycoprotein interactions with lipids

Recent experimental work has shown the importance of the lipid bilayer in P-gp function and conformation. It is well known that P-gp is affected by its environment and it can function as a low-activity phospholipid flippase, however, the interactions have not been described in detail. There is a particular interest in the interactions between P-gp and cholesterol, as recent studies propose an intimate relationship between the molecules.

Chapter 5 describes multiscale simulations of P-gp in a complex asymmetric bilayer that is designed to fit the composition of the blood-brain barrier. The bilayer itself is first characterised by a set of multiscale simulations and its characteristics are compared with other similar MD studies. The multiscale simulations of P-gp in the complex bilayer provided data on lipid diffusion, P-gp conformation and lipid ordering. Several potential cholesterol binding sites on the P-gp surface were observed during the coarse-grained simulations and these were further investigated by umbrella sampling simulations that calculated potential of mean force. The most likely site of cholesterol binding is the groove formed by TMH 10/12 or the putative substrate

entrance portal. The same portal was used by several types of lipids to enter the central cavity. This fits with P-gp function as a lipid flippase, however, it does not support the conclusions from chapter 4.

1.7.3 Dynamic protein-ligand interactions

MD simulations of protein-ligand interactions are becoming an essential method in drug development, however, there is a lack of tools that can analyse and display the results of dynamic protein-ligand interactions. Due to the high efficacy and low computational cost of molecular docking it has been the method of choice to analyse ligand binding and interactions and, therefore, several methods for analysis and display of these interactions have been developed.

A novel open-source software that displays dynamic protein-ligand interactions in 2D is described in chapter 6. The LINTools (Ligand Interaction Network Tools) script is written in Python and exploits the availability of various open-source libraries that can import and analyse molecular dynamics data, draw 2D depictions of molecules and calculate geometrical properties. LINTools takes topology and trajectory files as input and creates 2D diagrams of protein-ligand interactions according the user specifications. At the moment, the script can analyse residence time of residues within cutoff distance of the ligand, RMSF of the ligand and hydrogen bonding.

2

Simulation Theory and Methods

2.1 Introduction

The motions and interactions within a biological system such as a protein embedded in a membrane can be investigated with computational techniques such as molecular dynamics (MD). In the context of this thesis, MD has been used to investigate P-glycoprotein interactions with a substrate and the surrounding lipid bilayer. MD simulation was chosen as the primary method to be used in this thesis since it provides atomistic resolution on the interactions between protein, lipids, solvent and

drugs. This chapter describes the theoretical basis behind some of the MD simulation techniques used in this thesis with the associated considerations and limitations. Additionally, brief outlines of the simulation protocols, parameters and analysis tools are also provided.

2.2 Molecular dynamics simulation

When simulating a biological system, the level of resolution is an important consideration. Ideally one would like to simulate a system to the quantum level to observe the movements of the electrons and therefore events such as bond forming and breaking. However, the computational cost of such calculations is very high and can reasonably only be used for small systems, for example, a ligand with a few surrounding amino acids[139].

Classical molecular dynamics uses the Born-Oppenheimer approximation which states that the electrons can almost instantaneously adjust to any changes in position of the nuclei due to their difference in mass. This allows one to represent all atoms as a set of Cartesian coordinates and therefore simulate larger systems without the considerable computational cost associated with quantum mechanics simulations.

2.2.1 Equations of motion

In a classical molecular dynamics simulation, Newton's equation of motion is used to evolve the system over time. The initial masses (m_i) and the positions (r_i) of every atom are known from the initial configuration of the system. The force acting on each atom needs to be calculated, as knowledge of the force allows us to determine the acceleration of each atom, yielding a trajectory with the information about the position, velocity and acceleration for each atom in time. Newton's equation of motion,

which relates the force F_i acting on atom i to the atoms' mass m_i and acceleration $a_i(t)$:

$$F_i = m_i \vec{a_i}(\vec{t}) \tag{2.1}$$

can be rearranged to obtain:

$$\frac{d^2 r_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{2.2}$$

providing the differential equations that must be solved in order to obtain the trajectory. The force F_{x_i} on atom i acts in x direction, where r_i is the position of atom i with mass m_i. The force F_i on atom i can be obtained from:

$$F_i = -\frac{dV}{dr_i} \tag{2.3}$$

where V is a potential energy function.

In all biomolecular systems the motions of all particles are coupled together, leading to a many-body problem. Since it cannot be solved analytically, finite difference methods have been implemented and the problem is solved numerically in discrete timesteps δt. The Verlet algorithm is the most popular finite difference method used for MD simulation. A variation of the Verlet algorithm, known as the "leap-frog" algorithm is used in GROMACS[140]. Instead of calculating velocities at full time intervals, it first computes the velocities v_i at time $t + \frac{1}{2}\delta t$ using the velocities from time $t - \frac{1}{2}\delta t$ and the acceleration at time t:

$$v_i(t + \frac{1}{2}\delta t) = v_i(t - \frac{1}{2}\delta t) + a_i(t)\delta t \tag{2.4}$$

The velocities are then used to calculate the position r_i at time $t + \delta t$:

$$r_i(t + \delta t) = r_i(t) + v_i(t + \frac{1}{2}\delta t) \tag{2.5}$$

Thereby the velocities at time t can be calculated from:

$$v_i(t) = \frac{1}{2}\left[v_i(t + \frac{1}{2}\delta t) + v_i(t - \frac{1}{2}\delta t)\right] \tag{2.6}$$

2.3 Force fields

In classical molecular dynamics, the forcefield is a set of parameters that are used to compute the interatomic forces between all atoms in a particular system. These parameters contribute to the potential energy of the system ($U(r^N)$), where r denotes the position of the particles and N is the number of particles (e.g. atoms or coarse-grained beads) and can be described in bonded (U^{bonded}) and nonbonded ($U^{nonbonded}$) energy terms.

$$U(r^N) = U^{bonded} + U^{nonbonded} \tag{2.7}$$

The potential energy functions that represent the bonded interactions describe the bond stretching, bending of bond angles, as well as the rotation of the bond (torsion). The nonbonded energy term is usually calculated for electrostatic and van der Waals interactions between all pairs of atoms in different molecules or separated by at least three bonds.

2.3.1 Bonded interactions

$$U^{bonded} = U^{bonds} + U^{angles} + U^{dihedrals} + U^{impropers} \tag{2.8}$$

Bonded interactions between particles are modelled by harmonic potentials to describe the potential energy of springs, since the use of harmonic potentials is computationally efficient and they are valid over the range of bond lengths and angles that are sampled in an MD simulation of proteins. Bond stretching between two atoms is often described by Hooke's law formula:

$$U^{bonds} = \sum_{bonds} \frac{k_l}{2}(l - l_0)^2 \tag{2.9}$$

where oscillation about reference bond length l_0 is assumed. Forces between the atoms are characterised by the force constant k_l, while l denotes the bond length at any point of the simulation. Similar harmonic potential describes bond angle movements:

$$U^{angles} = \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.10}$$

that are characterised by the reference bond angle θ_0, bond angle θ and force constant k_θ.

The torsion energy term is described by the sum of proper and improper dihedral potential energy. The dihedral angle between four bonded atoms is described by a cosine function. The dihedral function as described in AMBER forcefields[141]:

$$U^{dihedrals} = \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] \tag{2.11}$$

where ϕ is the torsion angle and γ is the phase factor that determines where the torsion angle passes through its minimum, while n determines how minimums the

cosine function will have as the angle moves through the 360°.

Improper torsion energy term limits out of plane bending for planar groups and can be used to achieve a particular geometry and is given as a harmonic potential:

$$U^{impropers} = \sum_{impropers} \frac{K_\omega}{2}(\omega - \omega_0)^2 \tag{2.12}$$

where K_ω is the force constant, ω is the torsion angle at any instant of the simulation and ω_0 is the equilibrium value of the angle.

2.3.2 Nonbonded interactions

The potential energy of nonbonded interactions is calculated from the short ranged Lennard-Jones (LJ) potential and long ranged Coulomb potential. These describe van der Waals and electrostatic interactions respectively.

$$U^{nonbonded} = U^{van\,der\,Waals} + U^{electrostatic} \tag{2.13}$$

These potentials are used to describe interactions between atoms that are separated by at least three bonds or are parts of different molecules.

The Lennard-Jones potential calculating the van der Waals interactions between two atoms i and j takes the form:

$$U^{van\,der\,Waals} = \sum_{pairs(i,j)} 4\epsilon_{i,j}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.14}$$

where $\epsilon_{i,j}$ describes the attraction between atoms i and j, σ_{ij} is the distance at which energy between two atoms is zero, while r_{ij} is the distance between i and j. The

inverse sixth power term describes the long-range attractive forces in the system, such as London dispersion forces. The short-range repulsion forces are described by the inverse twelfth power term.

Coulomb's Law describes the electrostatic term of nonbonded interactions:

$$U^{electrostatic} = \sum_{pairs(i,j)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.15}$$

where q_i and q_j represent the partial charges of particles i and j, r_{ij} is the distance between two particles and ϵ_0 describes vacuum permittivity.

The parameters for Lennard-Jones and Coulomb potentials are derived from experimental data or quantum mechanical (QM) calculations[142].

2.3.3 All-atom forcefields

MD simulations can be used to calculate and represent motions and interactions of individual atoms. This approach, although computationally expensive, has the advantage of system being in inherently high resolution. The combined bonded and nonbonded energy terms describe the total potential of the system and are defined as a forcefield.

The four most widely used classical all-atom forcefields for macromolecular MD simulations are OPLS[143], GROMOS[144], CHARMM[145] and AMBER[141]. The parameter fitting for each of these forcefields was done differently, however, the main approaches involved fitting to experimental or QM data, with the assumption that the parameters tested on small molecules can be used on larger systems.

In recent years, the AMBER forcefield development has concentrated on optimisation to replicate the NMR data, concentrating on corrections of the backbone and

sidechain torsion potentials. When compared to other major forcefields in their ability to replicate experimental data, the ff99SB-ILDN[146] performs reasonably well[147]. However, several aspects must be taken into account when selecting a forcefield for simulation of a specific system, for example, the availability of parameters of lipids, drugs and other macromolecules.

For the all-atom simulations in this thesis, the AMBER ff99SB-ILDN forcefield has been used to describe the protein molecule, while a specialized forcefield was used for lipids - the Slipids (Stockholm lipids) forcefield[148], which is compatible with the AMBER forcefields.

2.3.4 Coarse-grained forcefields

In addition to all-atomistic forcefields, coarse-grained (CG) forcefields are used in molecular dynamics simulations to achieve much longer timescales, as well as larger systems that can even reach the size of systems investigated in experiments. This can be achieved due to a reduction in resolution by combining approximately four non-hydrogen atomistic particles into one large coarse-grained particle. Since the number of particles in a system are reduced by a factor of 4, the calculations for each timestep are simplified, which can lead to an increase to the speed of calculation by up to a factor of three compared to an all-atom simulation. The simulations in this thesis were run with the MARTINI v2.1 forcefield[149;150] with lipid topologies[151].

2.3.5 Conversion from coarse-grained to atomistic representation

Coarse-grained simulations can be used to assess the macroscopic properties of a system, such as protein-lipid interactions over a longer timescale, however, the simplified

system does not allow us to investigate the microscopic properties of a system such as side chain movement and conformational changes. The interesting phenomena observed in coarse-grained simulations can be investigated in more detail in atomistic simulations, utilising the coarse-grained to atomistic conversion. Two approaches have been used in this thesis - the CG2AT program[152] was used for the conversion of a protein molecule, while lipids were converted to atomistic resolution using the Backward backmapping program[153]. CG2AT uses fragment based approach[152], which is generally considered to be more precise, however, it is limited to the molecules and forcefields that have been included in the package. On the other hand, Backward geometrically projects the particles and the resultant system is slowly relaxed by MD simulations[153]. The backmapping process only requires a file that lists the particles in the molecule in the different resolutions and therefore can be used for any forcefield and any molecule, provided such a comparison can be reasonably made.

2.4 Molecular dynamics simulation technique

2.4.1 Constraints

The length of the time step used in MD simulations is determined by the motion with the highest frequency, typically bond stretching. However, since the conformational changes in a biomolecular system usually result from the changes in torsion angles and nonbonded interactions, bond stretching is less interesting in this context. Therefore, the time step can be increased by constraining the bond lengths between atoms during the simulation. This process is achieved through the use of algorithms such as SHAKE[154], RATTLE[155] and LINCS[156], that reset the bond length to its reference value after each time step. The 2 fs time step in the atomistic simulations in this thesis was enabled by the LINCS algorithm.

2.4.2 Periodic boundary conditions

Even with the expansion of computational resources according to Moore's Law, there are still limitations to how many particles in a system it is feasible to simulate. Simulations of finite systems are going to be prone to artefacts due to interactions with the boundary of the box in which the system is contained. To ensure that in such systems all particles can experience forces as if in bulk solution, periodic boundary conditions (PBC) are imposed. In a system with PBC enabled, a particle that exits one side of the box will reenter from the opposite side with the same velocity, since the system is treated as a lattice with infinite copies of itself spreading out in all directions. Periodic boundary conditions can be used on different box shapes, with the most common being cubic. The shapes used in this thesis were cubic and rectangular, owing to the shape and setup of the system, which in most cases involved a bilayer.

2.4.3 Treatment of long range interactions

Since interactions can also take place across periodic boundaries, in an infinitely stretched system the potential energy for all interactions is an infinite sum. A simple solution to this problem is to disregard all interactions that are further than a certain distance threshold with the potential reduced to zero above this cutoff, meaning that there is no interaction.

While this method works for limiting the van der Waals interactions of a particle, by reducing the LJ potential and switching it to zero above a certain distance cutoff from the particle, the electrostatic interactions act over large distances (> 1 nm). This is due to the slow rate of decay of the Coulomb potential, which reduces with the inverse of the distance r^{-1} rather than the inverse of the negative sixth power of the distance

r^{-6} as in the case of LJ potential. Therefore, introduction of a simple cutoff would lead to discontinuity in the system, which in turn would affect its accuracy.

Several complicated techniques have been devised to handle the treatment of long range interactions. The Particle Mesh Ewald (PME) method[157] splits the calculation of electrostatic interactions in two parts: a short-range potential and a long-range potential. While the short-range potential is calculated in real space and converges quickly, the long-range potential is evaluated in reciprocal space - a computationally expensive calculation. The PME method distributes the long-range interactions onto a finite grid and reduces the computational cost of this calculation by using the Fast Fourier Transform method[157].

2.4.4 Constant temperature and pressure simulations

The simulations in this thesis were conducted with a constant number of particles, N, at a constant temperature, T, and pressure, P, to create the isothermic-isobaric or NPT ensemble. The temperature and pressure were fixed using a thermostat and a barostat to control the fluctuations of the simulation conditions. The NPT ensemble allows closer comparison to laboratory experiments that are typically performed under constant temperature and pressure.

Different approaches were taken for coarse grained and atomistic simulations described in this thesis. In coarse grained simulations the protein, lipids and solvent were separately coupled to the Berendsen thermostat and barostat[158], while the same process in atomistic simulations was done with the V-rescale thermostat[159] and Parrinello-Rahman barostat[160].

Temperature coupling

The Berendsen thermostat[158] weakly couples the temperature of the system to an external heat bath with the temperature fixed at a desired value, thereby preventing excessive amounts of heat to enter or leave the system. The constant temperature is achieved by rescaling the velocities at each time step, since temperature of the system is related to its kinetic energy. The temperature changes proportionally to the difference between the temperature in the bath, T_{ref}, and the system, T:

$$\frac{dT(t)}{dt} = \frac{1}{\tau T}(T_{ref} - T) \tag{2.16}$$

where τT is a coupling parameter that determines how tightly the system is coupled to the bath. Typically parts of the system such as protein and solvent are coupled separately to compensate wild fluctuations in one part of the system with changes in another.

Although the Berendsen thermostat is stable and simple to implement, it does not produce the correct kinetic ensemble in atomistic simulations[159], however, good results with this thermostat can be achieved in coarse grained resolution. Although the Berendsen thermostat is used during the equilibration stages of atomistic simulations, an extension of this thermostat, the V-rescale thermostat[159], is used in the production simulations at atomistic resolution. The V-rescale thermostat has an added optimised force that allows it to reach the correct kinetic energy and canonical ensemble distribution.

Pressure coupling

Pressure coupling is implemented in a similar way to the temperature coupling with an external pressure bath providing a reference value for the system and rescaling the simulation box size and atom coordinates accordingly. Two barostats that are compatible with the "leap-frog" integrator are currently available in GROMACS - the Berendsen[158] and Parrinello-Rahman[160] barostats. The Berendsen barostat does not create the correct thermodynamic ensemble in atomistic resolution, however, it is suitable for coarse grained simulations. The Parrinello-Rahman barostat considers the pressure bath an integral part of the system and therefore includes the contribution from the pressure bath in the equations of motion. Being an improvement over the Berendsen barostat, the Parrinello-Rahman barostat is considered to be more reliable for simulations at equilibrium.

For simulations of systems that include a lipid bilayer, it is important to couple the simulation box vectors semi-isotropically, with the vectors perpendicular to the bilayer normal rescaled together, and independent from the vector parallel to the bilayer normal. This allows the bilayer to maintain a stable surface area per lipid ratio.

2.4.5 Energy minimisation

Energy minimisation is used for relaxing of the simulation system by guiding it towards a local minimum in the potential energy landscape. Numerically, this involves finding the atomic coordinates where the derivative of the potential energy function is zero. This step in the simulation setup is used to reduce and remove unfavourable interactions prior to the production runs, for example, steric clashes between protein side chains and water or lipid molecules. Steepest descent energy minimisation al-

gorithm searches for the minima by moving downhill through the potential energy landscape until a minimum is reached[142]. Although this method was used in this thesis, another popular algorithm is the conjugate gradient method, which focuses on reducing the system to a very low energy[142]. Similarly to the steepest descent algorithm, this method moves towards the most negative change in potential energy at each time step.

2.4.6 General simulation setup

Each simulation system consists of one P-glycoprotein molecule embedded within lipids and solvated with water molecules on both sides of the bilayer. The system is neutralized by adding Na^+ or Cl^- ions in place of random water molecules to reach 150 mM NaCl concentration. The energy of the system was minimised using the steepest descent method to reduce bad contacts. The temperature and pressure of the parts of the system such as protein, lipids and solvent were coupled separately. The simulations were run with GROMACS 4.6.x[140] on a Linux cluster.

Coarse-grained MD simulations

The MARTINI[149;150] forcefield was used in the CG MD simulations to describe the protein and lipid. The Berendsen thermostat and barostat[158] were used for temperature and pressure coupling. Simulations were run at temperature of 310K and pressure of 1 bar. The LINCS algorithm[156] was used to constrain bond lengths with a timestep of 20 fs. During the equilibration process the protein backbone beads were constrained. The 100 ns equilibration period was continued with 10 µs production runs of normal MD simulations.

All-atom MD simulations

The production runs of all-atom MD simulations used the V-rescale thermostat[159] and the Parrinello-Rahman barostat[160] to keep the system at constant temperature and pressure. The temperature was kept at 300K or 310K for systems with POPC or complex bilayer respectively, while the pressure was 1 bar. AMBER ff99SB-ILDN forcefield[146] was used to describe protein and the Slipids forcefield[148] described the lipid molecules. The timestep was 2 fs. The LINCS algorithm[156] was used to constrain bond lengths and the long-range electrostatic interactions were calculated with PME method[157].

2.5 Free energy calculations

Free energies can be used to determine how energetically favourable a certain process, such as ligand binding, is in a biological system. The Gibbs function, G, can be used to express free energy in simulations that use the NPT ensemble and is given by:

$$G = U + PV - TS = H - TS \tag{2.17}$$

where U is the internal energy, H is the enthalpy, S is the entropy and T is the temperature of the system.

However, determining the correct values of G as well as S and U from conventional, unbiased MD simulation is not trivial. One major caveat is the lack of sampling of phase space within these simulations. Usually MD simulations of biological systems will tend to sample the many available low energy conformations, leaving the higher energy configurations inadequately sampled.

This has lead to the development of several enhanced sampling MD methods that determine a related quantity, the free energy change between two states (ΔG). The list includes free energy pertubation[127], thermodynamic integration[126], steered MD[132], metadynamics[161], adaptive biasing force[162] and umbrella sampling[163].

2.5.1 Steered molecular dynamics

Due to limitations in the available computer time, unbiased molecular dynamics simulations often cannot reach the timescale necessary to observe rare processes such as binding and unbinding of protein-substrate complexes. Steered MD (SMD)[132] uses external forces to steer the system in the desired direction in order to sample conformations that would not have been visited due to high energy barriers. Typically, an SMD simulation will apply a constraint, such as a harmonic potential, that will move the system along a prescribed path defined by the reaction coordinate, for example, the distance between two atoms or molecules.

Care must be taken when choosing parameters such as the strength of the harmonic potential or the pulling speed. Since SMD is an nonequilibrium method, introducing too much force in the system can cause disruptions and extremely unphysiological states, however, reducing these constants would increase the simulation time. Therefore, a balance between the parameters must be struck in order to observe physiologically relevant processes in a reasonable timescale.

2.5.2 Potential of mean force

Potential of mean force (PMF) can be used to calculate free energy changes as a function of a coordinate of the system[142]. In biochemical systems this, often geometrical in nature, reaction coordinate can be relatively simple such as distance or

angle between atoms or a more complicated function of the Cartesian coordinates of a system[164]. PMF along a chosen reaction coordinate ξ can be defined by:

$$\omega(\xi) = \omega(\xi^*) - k_B T \; ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \tag{2.18}$$

where ξ^* is an arbitrary constant of the coordinate and $\langle \rho(\xi) \rangle$ is the average distribution function, which is obtained from the Boltzmann weighted average:

$$\langle \rho(\xi) \rangle = \frac{\int dR \delta(\xi'[R] - \xi) e^{-U(R)/k_B T}}{\int dR e^{-U(R)/k_B T}} \tag{2.19}$$

where $U(R)$ is the total energy of the system as a function of the coordinates R, while $\xi'[R]$ describes the degrees of freedom of the reactive coordinate and $\delta(\xi'[R] - \xi)$ is the Dirac delta function for coordinate ξ.

Since the PMF is generally calculated for physically achievable processes, it differs from other techniques such as free energy pertubation calculations that often describe mutations along non-physical pathways. Therefore, rate constants can be derived from the highest energy barrier in PMF profile, which describes the transition state of the process.

However, it is often impossible to calculate the PMF from regular MD simulations, in particular, if the energy barriers along the reaction coordinate are too large to efficiently sample within the available computer time. Special sampling methods have been developed to solve this issue and the umbrella sampling method[163], which is the technique used in this thesis, will be described in more detail.

2.5.3 Umbrella sampling

Umbrella sampling employs an artificial biasing potential that is applied to the atom or molecule of interest and restrains it within small interval of the reaction coordinate. This approach ensures adequate sampling along the reaction coordinate by forcing the system to achieve more efficient conformational sampling in this region[163].

In practice, independent simulations, also known as windows, are used to confine the sampling at regularly spaced intervals along the reaction coordinate. The biasing potential used in this thesis was a harmonic restraint of force constant k:

$$\omega_i(\xi) = \frac{k}{2}(\xi - \xi_i)^2 \tag{2.20}$$

where ξ_i is the centre of the potential for the i^{th} window. If the intervals between windows have been carefully considered, this method provides a set of overlapping biased probability distributions of the reaction coordinate. However, adequate sampling of all degrees of freedom, including those not restrained by the biasing potential, must be ensured to provide a converged profile of the PMF. In this thesis, it was checked whether the calculated PMF changes with the simulation length[165] or within different subsets of simulation windows[166], thereby assessing the sampling.

2.5.4 The weighted histogram analysis method (WHAM)

The main difficulty with umbrella sampling is to unbias and combine the individual windows. One strategy to obtain the free energy constants is to adjust the $\omega_i(\xi)$ of adjacent windows either manually or automatically (by least-squares procedure), in the overlapping regions until they match. However, the overlapping data is discarded in production of the PMF profile that spans the reaction coordinate.

A more sophisticated method is the weighted histogram analysis method (WHAM)[167] that uses all the information from the umbrella sampling simulations without discarding the overlapping regions. It calculates the total unbiased distribution function $\langle\rho(\xi)\rangle^{unbiased}$ as a weighted sum of the unbiased distribution functions:

$$\langle\rho(\xi)\rangle^{unbiased} = \frac{\sum_{i=1}^{N_\omega} n_i \langle\rho(\xi)\rangle_i^{biased}}{\sum_{i=1}^{N_\omega} n_j \, e^{-[\omega_j(\xi)-F_j]/k_BT}} \tag{2.21}$$

while the free energy constants are determined using the optimal estimate for the distribution function:

$$e^{-F_i/k_BT} = \int e^{-\omega^i(\xi)/k_BT} \langle\rho(\xi)\rangle^{unbiased} d\xi \tag{2.22}$$

The Equations 2.21 and 2.22 are solved iteratively until F_i values lie within a predetermined threshold tolerance.

2.6 Homology modeling

Homology or comparative modelling of protein three-dimensional structures aligns target sequence to known structures of related proteins, also called templates. The MODELLER software allows one to build homology models by satisfying spatial restraints derived from the known structures and their alignment with the target sequence[168]. Additionally, MODELLER can build and optimise *de novo* loops in protein structures. This feature was used to build the flexible linker region in the human P-glycoprotein models used in this thesis.

The QMEAN Z-score was used to compare homology models from different templates[169]. It is an absolute measure for the quality of protein models that assesses the

"degree-of-nativeness" of a model by comparing it to high-resolution crystal structures.

3

Homology Modelling and Atomistic Simulations of P-glycoprotein

3.1 Homology modelling of human P-glycoprotein

The efforts to investigate P-glycoprotein structural properties have been hindered by the lack of a human P-glycoprotein crystal structure, however, structures of related proteins have started to appear in recent years[32;40;43;46]. These can be used as templates for homology modelling to generate structure-based hypotheses, but can also

be investigated in their own right to give insight into structure-activity relationships. The homology models have been extensively explored by computational methods such as molecular dynamics and free energy calculations[96;170]. In this chapter, many of the available templates were made into homology models for human P-glycoprotein and the best models were further investigated by molecular dynamics simulations.

3.1.1 Methods

The protein sequences of crystal structures were aligned with the human ABCB1 sequence along with several other P-glycoprotein or homologue sequences using Clustal W2[171] (available at [i]). The UniProt IDs used in these alignments were P08183 (human ABCB1), P34712 (*C. elegans* ABCB1), P21448 (*C. griseus* ABCB1), P21447 (mouse ABCB1), and Q9ZR72 (*A. thaliana* ABCB1). The alignment was later manually edited. The final alignments are available in appendix A. The protein structures were acquired from Protein Data Bank (available at [ii]). Before modeling, the structures were fixed for missing side chains with the WhatIf[172] server (available at [iii]). The resulting structures were modelled with Modeller v9.10[168]. Alignment included the linker region of approximately 60 residues between NBD1 and TMD2 (see chapter 1). 100 models were generated and scored with the DOPE score[173] with the best scoring model selected.

The homology models resulting from this process were assessed by the QMEAN Z-score[169] to provide a measure of their quality. The score also was used to rank homology models from different templates by measuring the absolute quality of the protein model. The Z-score provides an estimate of the "degree of nativeness" of the structural features of the model by comparing it to high resolution experimental

[i] http://www.ebi.ac.uk/Tools/msa/clustalw2/
[ii] http://www.rcsb.org/pdb/home/home.do
[iii] http://swift.cmbi.ru.nl/whatif/

structures. The scores for membrane proteins are lower than those of similar quality soluble proteins due to solvent accessibility being used as one the scoring parameters.

3.1.2 Available Templates

Currently, there are several crystal structures that can be used as templates for human P-glycoprotein homology modelling (Table 1.1). One of the first structures that has been used for this purpose is the crystal structure of the *Staphylococcus aureus* vancomycin intermediate resistance 1866 transporter (Sav1866). This structure was solved in outward-open conformational state with either ADP[35] or a non-hydrolysable ATP analogue -adenosine-5'-(beta,gamma-imido)triphosphate (AMP-PNP)[40] bound. Although the structure of *Escherichia coli* multicopy suppressor of htrB (MsbA) is in an outward-open state[41], the vast majority of other published crystal structures from a variety of organisms have been solved in some kind of inward-open state (Table 1.1, Figure 1.3).

The first mammalian crystal structure of mouse P-gp was published in 2009 and has been used in homology modelling, often in combination with docking, to understand protein ligand interactions[32;110;111]. However, the subsequent publication of *Caenorhabditis elegans* P-gp structure highlighted some differences with the mouse structure[43]. The arginine mutagenesis data supported the *C. elegans* structure and showed that the helix register on TMH5 of the mouse structure is probably wrong[50]. Several other structures of mouse P-gp structures have been published, including a refined structure with an improved electron density map that allowed the correct helix register to be achieved[44;47].

Another potentially interesting template was published in 2014 - the crystal structure of an eukaryotic P-glycoprotein homologue from *Cyanidioschyzon merolae* which claimed to give more insight in to the transport mechanism of P-glycoprotein[46]. The

Table 3.1: Overview of human homology models of P-glycoprotein

Protein	PDB	QMEAN Z-score	Resolution (Å)	Sequence identity
Sav1866	2HYD	-2.34	3.0	34%
ABCB1 mouse	3G5U	-2.61	3.8	87%
ABCB1 *C. elegans*	4F4C	-1.9	3.4	46%
ABCB10 human	3ZDQ	-2.5	2.85	37%
ABCB1 mouse	4KSB	-2.46	3.8	87%
CmABCB1 *C. merolae*	3WME	-2.14	2.75	39%
ABCB1 mouse refined	4M1M	-1.86	3.8	87%

A QMEAN Z-score of less than -5 is taken as an indication there could be something problematic with the structure[169].

crystal structure of human ABC transporter ABCB10 (also known as ABC mitochondrial erythroid, ABC-me, mABC2, or ABCBA) has been solved in an inward-open state bound with non-hydrolysable ATP analogues[45]. This is in contrast to the nucleotide bound Sav1866[40] and MsbA[41] crystal structures, both solved in an outward-open state, and suggests that ABC transporters can exist in an inward open state with bound nucleotides.

In homology modelling of P-glycoprotein, a compromise between sequence identity and resolution of the template crystal structure must be found, as the eukaryotic structures generally (with the exception of ABCB10) have lower resolution than those of the bacterial homologues (Table 1.1). The available crystal structures were used to build homology models and their overall quality was assessed using QMEAN Z-score as shown in Table 3.1.

The most promising human homology models were based on refined mouse, *C. elegans* and *C. merolae* ABCB1 crystal strucutures. The QMEAN Z-scores of these models ranged from -1.86 (best) to -2.14. This result is comparable to a medium quality model (mean Z-score -1.75)[169], without taking into account the unfavourable solvent accessibility term. Interestingly, eukaryotic structures of ABCB1 provided better homology models than the structures of bacterial homologues or human ABCB10,

suggesting that high sequence identity is preferable for high quality models. However, this is not an exact measure, as even the templates that give the best homology models have a wide range of sequence identity.

3.1.3 Homology model from the *C. elegans* crystal structure

The crystal structure of the *C. elegans* P-glycoprotein was solved at 3.4 Å resolution (PDB 4F4C). Even though the sequence identity to human P-glycoprotein is 46%, the ATP activity of *C. elegans* P-glycoprotein is stimulated by some of the human P-glycoprotein substrates, inferring functional similarity[43]. The inward-open structure was obtained without any nucleotides, substrates or inhibitors bound. Since the separation between NBDs is quite large, measuring 40.4 Å, the structure is very open to the cytoplasm.

The structure is also peculiar since drug transport is only feasible from one side of the protein. The gap between TMH4 and TMH6 is obstructed by a helical hairpin. TMH10 and TMH12 form a gap in the intramembrane part of protein which could be used by P-glycoprotein substrates to enter the central cavity from the lipid bilayer. A truncation mutant lacking the N-terminal helical hairpin functioned similarly to the wild type *C. elegans* P-gp in a cytotoxic assay[43].

Even though the N-terminus of the *C. elegans* crystal structure is well resolved, it was shortened by 9 residues for the homology model. The N-terminus of the mouse protein was not present in the crystal structures, suggesting that the region is probably disordered[47]. Since the hairpin helix on the N-terminus is most likely not essential for function and probably not present in mammalian proteins, a smaller disordered loop was modelled in the N-terminus of the actual model.

The gap between TMH10 and TMH12 has been suggested to function as the substrate

entrance portal due to unwinding of TMH10 at the level of the inner leaflet of the lipid bilayer. The flexibility of the helix might allow it to function as a hinge for the gate to accommodate larger substrates of P-gp.

Overall, the homology model from the *C. elegans* crystal structure was more open to the cytoplasm than others due to the wide distance between NBDs. The quality of the model was predicted to be very good (QMEAN Z-score -1.90) with one of the highest absolute scores (Table 3.1).

3.1.4 Homology model from *C. merolae* crystal structure

The crystal structure of CmABCB1, a eukaryotic P-gp homologue was solved to a high resolution (PDB 3WME 2.75 Å)[46]. It is a half-sized ABC transporter existing as a homodimer, however, it shows similar functional properties to those of human P-gp. Based on these similarities and the CmABCB1 crystal structure, Kodan et al. propose a gating mechanism for transport consisting of intramembrane and extracellular gates. Similarly to the crystal structure of *C. elegans* ABCB1[43], unwinding in TMH4 (equivalent to TMH10 in *C. elegans* structure) is noted. Therefore, it was suggested that this feature of TMH4 could form a part of the intramembrane gates.

Due to the high resolution of the template crystal structure, a high quality homology model could have been expected. This model had comparable quality to the homology models from refined mouse and *C. elegans* crystal structures with a QMEAN Z-score of -2.14 (Table 3.1).

3.1.5 Homology model from refined mouse crystal structure

The initial mouse crystal structure was a very popular choice of template for homology modelling of human P-glycoprotein due to its high sequence identity and similar

substrate specificity with the human protein[32]. However, doubts were cast about the correctness of the first mouse structure, mainly concerning the helix register of TMHs 3,4 and 5[43;55]. It was also criticised for being solved in the absence of nucleotides or lipid bilayer. This and other mouse crystal structures were used as templates for homology modelling, however, the quality of models proved to be low (-2.46 to -2.61 QMEAN Z-score) (Table 3.1). Recently, a refined mouse structure was published with improved electron density maps that improved the helix register of transmembrane helices 4 and 5 (PDB 4M1M)[47]. In comparison to other models from mouse crystal structures, the homology model from the refined structure was predicted to have high overall quality with a QMEAN Z-score of -1.86 (Table 3.1). The quality of the model and the sequence and functional identity of the template made this a very promising model to carry forward in the remainder of the thesis.

3.1.6 Template selection

Template choice can be based on several factors - sequence identity to human P-glycoprotein or resolution of the crystal structure and conformational state, however, no ideal template can be identified. For example, the mouse P-gp is very closely related to human protein and has high sequence identity, but the resolution of the crystal is relatively low and previously, doubts have been cast on the correctness of the mouse structures.

The homology models from refined mouse, *C. merolae* and *C. elegans* crystal structures had the highest, very similar QMEAN Z-scores. The resulting homology models were slightly different in their conformation reflecting the differences between templates (Figure 3.1). However, all of these templates were solved in an inward-open conformation. Due to the similarity of QMEAN Z-scores of the best homology models, the choice of model to be used in further experiments could not be made by taking

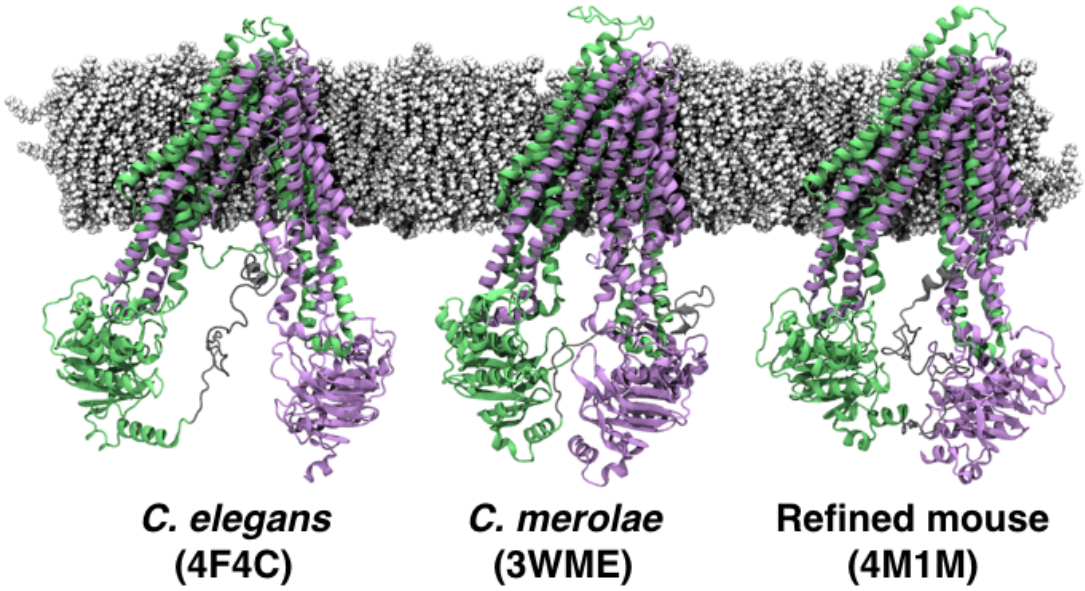


Figure 3.1: The homology models of human P-glycoprotein inserted in a POPC bilayer before molecular dynamics simulations. TMD1/NBD1 are shown in green and TMD2/NBD2 are shown in purple.

into account the quality of the model alone.

Homology models and their templates are just single snapshots of protein structures that are dynamic in reality. Therefore, a more descriptive way to compare the homology models is to assess their conformational motions is by molecular dynamics simulations. The homology models can be inserted into lipid bilayers and their dynamics can be monitored throughout the simulations to reveal the functional motion of P-gp. The next two sections of this chapter will describe the molecular dynamics simulations performed on the homology models. Similar simulations were also done with their respective template crystal structures, giving a comparative control.

3.2 Atomistic simulations of human P-glycoprotein homology models

Homology modelling of human P-glycoprotein resulted in several high quality models, however these are static snapshots of protein structure and do not give any information about the dynamics. As the homology models from *C. elegans*, *C. merolae* and refined mouse crystal structures gave the best QMEAN Z-scores, these models were inserted in to a lipid bilayer to investigate the stability and dynamic behaviour of the human homology models from these different templates.

3.2.1 Methods

The refined model (or crystal structure) was inserted into 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer. All molecular dynamics simulations were run on a Linux cluster, using GROMACS 4.6.x[140] with AMBER-99SB-ILDN forcefield[146] describing the protein and Slipids[148] describing POPC. Simulations were run at 300K and pressure 1 bar. The v-rescale algorithm was used to couple the temperature of the system with a coupling constant of 1 ps. The system pressure was coupled in semi-isotropic fashion (x and y, independent of z), using the Parrinello-Rahman algorithm with a compressibility of 4.5×10^{-5} bar^{-1} and a coupling constant of 1 ps. Particle-mesh Ewald (PME)[157] method was used to account for electrostatic interactions. The time step was 2 fs. The simulations were run on a Linux cluster in triplicate.

The overall stability and dynamics of P-gp structures were assessed by root mean square deviation (RMSD) measurements of Cα atoms, while the motions experienced by each individual domain were expressed as root mean square fluctuations and calculated per residue. The extent of NBD movements were described by measuring the

distance between the Walker A conserved lysine residue on NBD1 LYS433 and the serine (SER1177) from the Signature motif of NBD2 defined as (ATP Binding) Site 1. Site 2 was formed by the Walker A lysine on NBD2 (LYS1076) and the Signature motif serine on NBD1 (SER532). All of these measurements were made with tools available in the molecular dynamics simulations package GROMACS[140]. The structures and trajectories were displayed using Visual Molecular Dynamics (VMD)[174].

3.2.2 Atomistic simulations of the human homology model from *C. elegans* crystal structure

The Cα RMSD of the protein structure was high for a membrane protein ($\sim$6-10 Å) (Figure 3.2A), however, RMS fluctuations of individual residues suggested that it could be explained by NBD motions (Figure 3.2B), while the transmembrane domains appeared very stable and rigid. Visual check of the trajectory revealed that the high RMSD is caused by "breathing" motions of the NBDs, as previously described in the literature[124;175]. These observations correlated with measurements of the distance

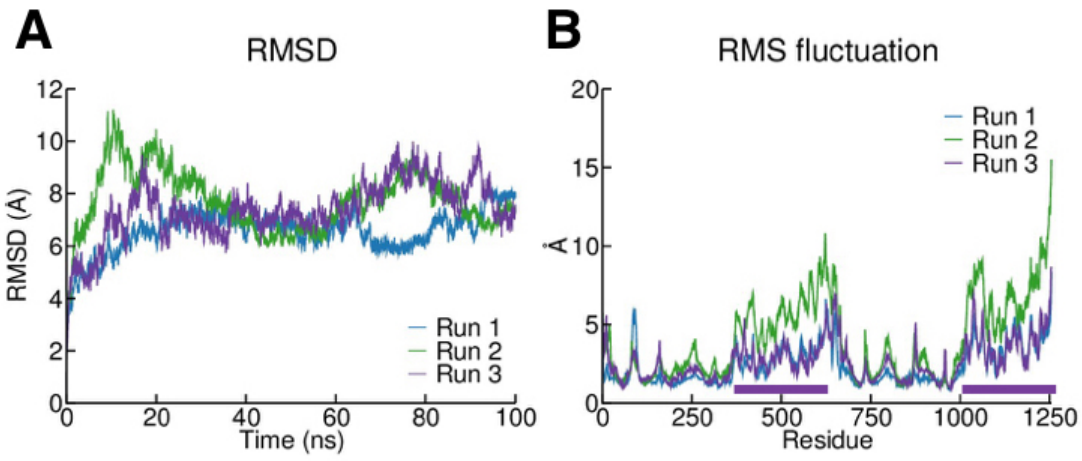


Figure 3.2: (A) RMSD of Cα atoms for run 1 (blue), run 2 (green) and run 3 (purple). (B) RMSF of P-gp residues for run 1 (blue), run 2 (green) and run 3 (purple). Bars indicate the NBD positions.

between the ATP binding sites. Interestingly, the NBD distance exhibited a slight tendency to reduce during the 100 ns simulations (Figure B.2). At the end of these simulations, it largely matched that of other homology models described in this section which did not exhibit such large degree of movement. This observation could mean that the wide opening of *C. elegans* crystal structure is unphysiological for human P-gp. That view is supported by a study which shows that P-gp can transport drugs with crosslinked NBDs[92].

3.2.3 Atomistic simulations of human homology model from the *C. merolae* crystal structure

Similarly to the results from the homology model based on the refined mouse ABCB1 crystal structure, the homology model from the *C. merolae* template appears to be very stable in a lipid bilayer during the 100 ns MD simulations according to the Cα-RMSD and RMSF plots (Figure 3.3 A,B). The nucleotide binding domains remain largely immobile with respect to one another, showing small fluctuations around the

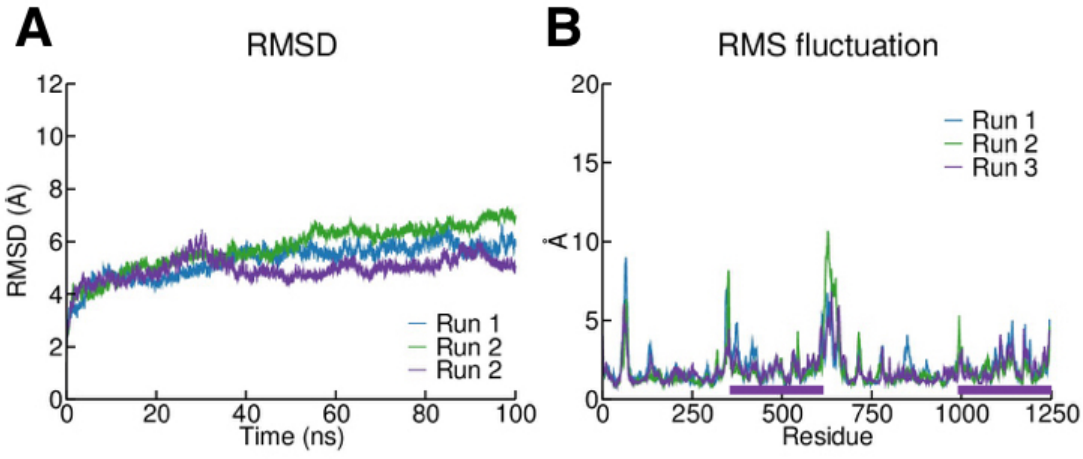


Figure 3.3: (A) RMSD of Cα atoms for run 1 (blue), run 2 (green) and run 3 (purple). (B) RMSF of P-gp residues for run 1 (blue), run 2 (green) and run 3 (purple). Bars indicate the NBD positions.

distance observed in the homology model itself (see Figure B.4).

Since this model and the homology model from the refined mouse crystal structure are stable during the 100 ns MD simulations, selection of a final model was difficult. However, although the *C. merolae* CmABCB1 crystal structure has been resolved to a high resolution, the protein itself is composed of a dimer of "half transporters" and therefore is evolutionarily more remote homologue to human P-gp compared to the mouse P-gp. Mouse ABCB1 is very closely related to human protein in sequence and function, therefore, it is more likely to give a better homology model when dynamics is concerned.

3.2.4 Atomistic simulations of human homology model from refined mouse crystal structure

In comparison to the human homology model from the *C. elegans* template, this homology model exhibits lower overall RMSD (Figure 3.4A) and less movement between domains is observed. The root mean square fluctuations (RSMF) shows that

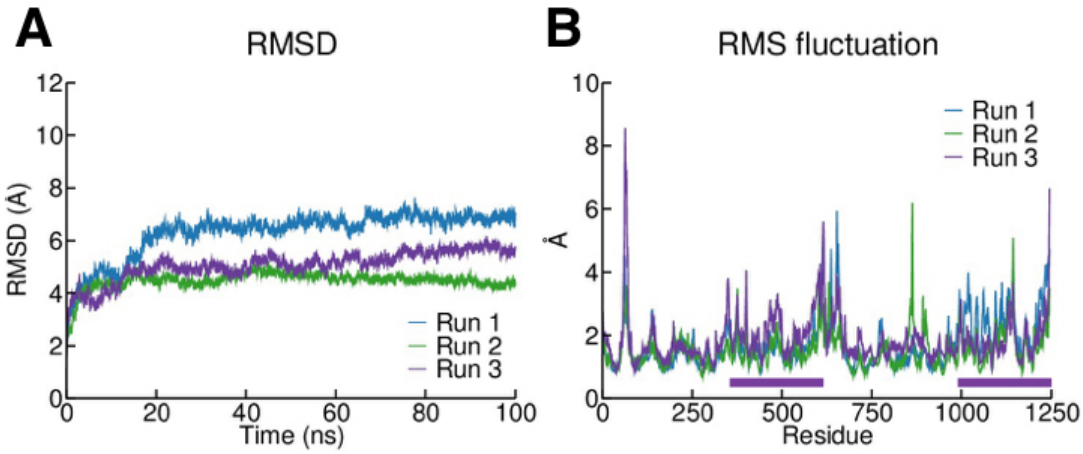


Figure 3.4: (A) RMSD of Cα atoms for run 1 (blue), run 2 (green) and run 3 (purple). (B) RMSF of P-gp residues for run 1 (blue), run 2 (green) and run 3 (purple). Bars indicate the NBD positions.

any movement reflected in the RMSD originated from the movements of flexible linker and other loops present in the P-gp structure (Figure 3.4B). Therefore, this homology model is very stable in the lipid bilayer during the 100 ns of simulation time.

The NBDs of the homology model from the refined mouse crystal structure have the smallest distance between the ATP binding sites compared to the other two homology models subjected to MD simulations (Figure B.3). Therefore, perhaps unsurprisingly, this model exhibits the least amount of movement between NBDs.

3.2.5 Atomistic simulations of *C. elegans*, refined mouse ABCB1 and *C. merolae* crystal structures

The crystal structures of *C. elegans*, refined mouse and *C. merolae* ABCB1 have been solved in various conformations of inwards-open state and exhibit slight structural differences to each other. Therefore, the dynamics of the different crystal structures of ABCB1 are of interest in their own right and as a way of comparison with the corresponding homology models.

The flexible linker was modelled in the crystal structures of *C. elegans* and mouse ABCB1, as these are full length eukaryotic ABC transporters. No linker was added to crystal structure of CmABCB1.

As might be expected, the crystal structures showed similar dynamics to the human homology models originating from them. The *C. elegans* crystal structure exhibited the largest movements of domains, reflected in RMSD and RMSF, while the refined mouse ABCB1 crystal structure moved the least (Figure 3.5).

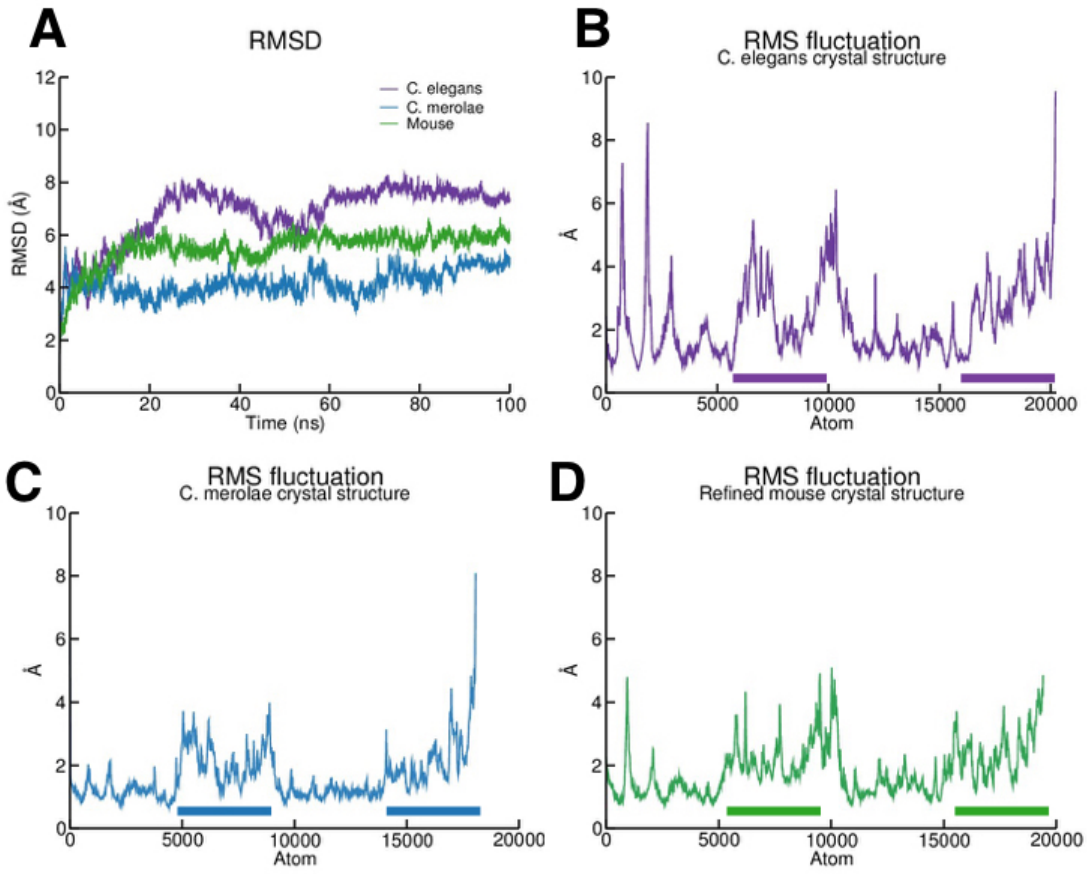


Figure 3.5: (A)The Cα-RMSD of *C. elegans* (purple line), *C. merolae* (blue line) and refined mouse (green line) ABCB1 crystal structures. (B) RMSF of *C. elegans* crystal structure. (C) RMSF of *C. merolae* crystal structure. (D) RMSF of refined mouse crystal structure. In the RMSF plots bars indicate NBD positions.

3.2.6 Justification of the mouse template

As discussed previously, deciding on the best human homology model is not a trivial task for P-glycoprotein as many factors, including resolution and sequence identity of the template, have to be taken into account. The homology models from *C. elegans* and refined mouse ABCB1 crystal structures had the best QMEAN Z-scores, implying high overall quality of the structure even if the resolution of the template crystal structures was comparatively low (see Tables 4.1 and 4.2).

Molecular dynamics simulations of the homology models revealed the dynamic motions of P-gp in a model lipid bilayer. In previous studies of P-gp in bilayer, it was observed that the main motions exhibited happened between NBDs, while the domains themselves remained rigid[175]. This so-called "breathing" motion was characterised in a study using both MD and double electron-electron resonance spectroscopy[124]. It can be observed that the homology model from the *C. elegans* crystal structure exhibits the most fluctuation during the simulation timescale of 100 ns. In contrast, the homology models from the refined mouse and *C. merolae* templates showed little movement between domains.

Taking into account all of the aforementioned factors, the homology model from the refined mouse template was selected to be used in further molecular dynamics simulations. The high quality of the protein structure and stability observed in molecular dynamics simulations suggested this model as the most appropriate. As a means of comparison we also prepared simulations of the templates themselves.

3.3 Atomistic simulations of human P-glycoprotein homology model with ATP and Mg^{2+} ions

Due to the milimolar concentration of ATP and ADP in a cell, it is unlikely that P-glycoprotein would exist in an apo, nucleotide unbound state for an extended amount of time. To gain a more physiologically relevant understanding of P-gp dynamics in lipid bilayer, an ATP molecule and a Mg^{2+} ion were added to each nucleotide binding domain.

3.3.1 Methods

The coordinates of ATP and Mg^{2+} ions were determined from the TM287/288 crystal structure crystal[42] (PDB 3QF4) crystalised with ANP (phosphoaminophosphonic acid-adenylate ester) and Mg^{2+} ions. The structure of the homology model was taken from a snapshot of MD simulation of the nucleotide-free homology model from refined mouse crystal structure in lipid bilayer after 10 ns of simulation time. The nucleotide binding domains of the homology model and the TM287/288 crystal were aligned by PyMOL[176] and the ANP/Mg^{2+} coordinates were transferred. ANP was converted to ATP by exchanging nitrogen for oxygen at the last phosphate group. ATP parameters[177] for the AMBER ff were obtained from AMBER parameter database (available at: http://www.pharmacy.manchester.ac.uk/bryce/amber/). The P-gp ATP-Mg^{2+} system in a POPC bilayer was simulated for 100 ns in triplicate.

3.3.2 Comparison between apo and ATP and Mg^{2+} bound P-glycoprotein atomistic simulations

The addition of ATP and Mg^{2+} ions led to a reduction in Cα-RMSD when compared to MD simulations of the apo homology model (see Figure 3.4A and Figure 3.6A). Visual inspection of the trajectories and RMSF of individual residues suggests that this decrease in RMSD could be caused by reduction of NBD movement relative to one another (Figure 3.6B). Two separate MD studies have observed metastable states of P-gp with bound ATP and Mg^{2+} [119;178].

Monitoring of the distance between ATP binding sites revealed a difference between apo and ATP-Mg^{2+} bound P-gp. In the simulations of apo P-gp, Site 1 shows more flexibility in distance Figure B.2, while the trend is reversed in the ATP-Mg^{2+} bound P-gp simulationsFigure B.4. The asymmetric nature of NBD movements has been observed previously in MD simulations[119].

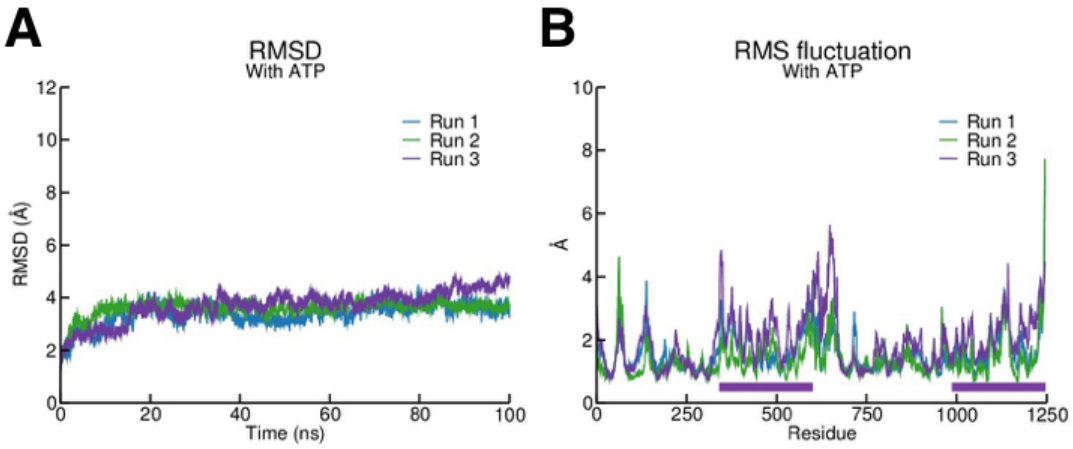


Figure 3.6: (A) RMSD of Cα atoms for run 1 (blue), run 2 (green) and run 3 (purple). (B) RMSF of P-gp residues for run 1 (blue), run 2 (green) and run 3 (purple). Bars indicate the NBD positions.

3.3.3 Behaviour of the potential substrate entrance gates in MD simulations

Recent ABCB1 structures have suggested that the gaps between transmembrane helices 4/6 or 10/12[47] could function as entrance portals for substrates. These entrance portals have access to both the inner leaflet of the lipid bilayer and the cytoplasm since the transmembrane helices that form them extend into the cytoplasm. It could therefore be argued that these portals consist from two parts - membrane and cytoplasmic sections (Figure 3.7B). The "hydrophobic vacuum cleaner" and "flippase" models support access through the gaps formed either by TMH4/6 or TMH10/12, where drugs could access the central cavity via an intra-membrane pathway. These models describe a drug molecule entrance to the binding site by partitioning into the bilayer, then laterally diffusing towards P-gp for incorporation into the central cavity[93]. This theory has been supported by experimental evidence[95] and the crystal structures of *C. elegans* and *C. merolae* ABCB1 that have a section of TMH 10 that is unwound and therefore, possibly forms part of the entrance gate[43;46]. Partial entrance of a lipid molecule from the bilayer through the gap between TMHs 4 and 6 has been observed in a MD simulation, suggesting this region to be a potential entrance portal that allows direct drug uptake from the bilayer[124]. An alternative entrance pathway from the cytoplasm has also been proposed for P-gp substrates[179]. It has also been suggested that the compounds on the hydrophilic end of the spectrum could use such a pathway[118]. Recently, O'Mara and colleagues computed potential of mean force pathways that considered access directly from the cytoplasm[122] for different drugs but concluded that the pathways were not clearly defined.

In the simulations of apo P-gp in a POPC bilayer, it was noticed that the intramembrane portions of these portals close very quickly - within 10 ns Figure 3.7C. The distance between TMH4 and TMH6 at the membrane part of the entrance portals

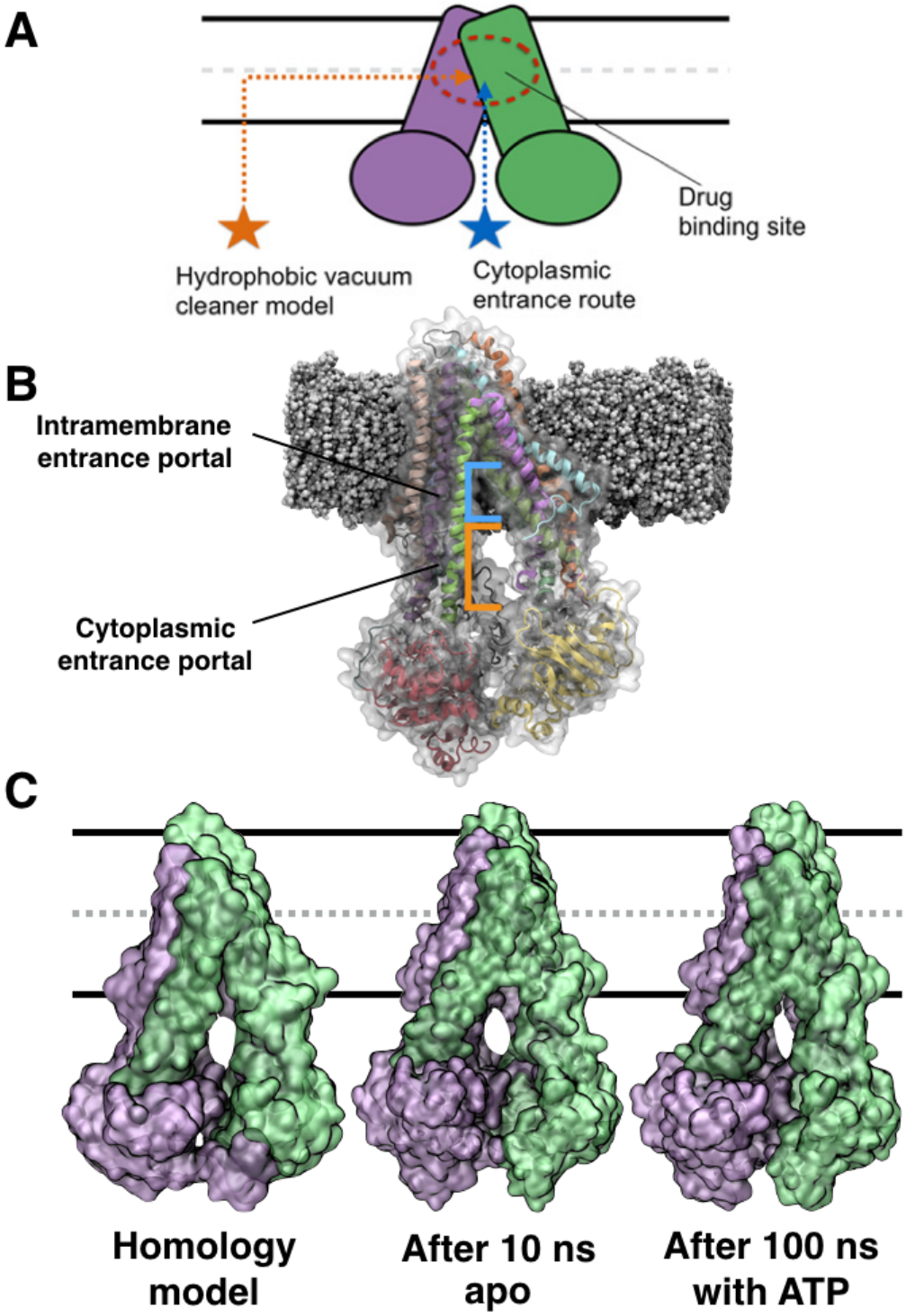


Figure 3.7: (A) Possible routes for substrate entrance illustrating hydrophobic vacuum cleaner and membrane pore models. (B) Cartoon of P-gp showing the location of potential entrance gates and the membrane and cytoplasmic sections that comprise them. (C) Molecular surface representations of P-gp illustrating that the membrane section of entrance portal closes in MD simulations.

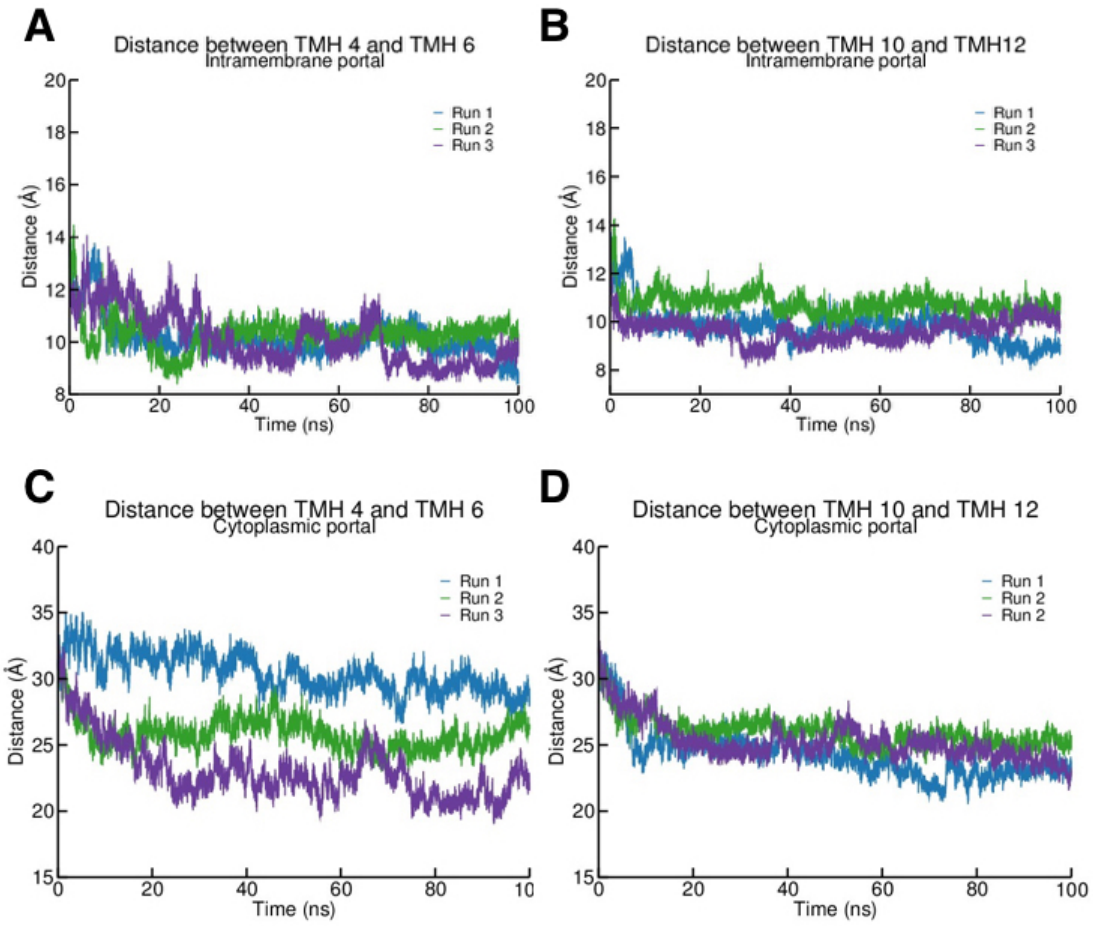


Figure 3.8: (A,B) Distances between intramembrane sections of TMH4/6 and 10/12 respectively. (C,D) Distances between cytoplasmic sections of TMH 4/6 and 10/12. The intramembrane section of helices measured from extracellular part of a helix to the point where it would reach the headgroup region of the inner leaflet of lipid bilayer. The part of a helix that would extend further into cytoplasm was considered to form part of cytoplasmic section a transmembrane helix. TMH 4 intramembrane section - residues 221-232, cytoplasmic section - residues 234 - 259, TMH 6 (residues 341-354 and residues 355-370), TMH 10 (residues 865-880 and residues 881-905) and TMH12 (residues 982-994 and residues 995-1014)).

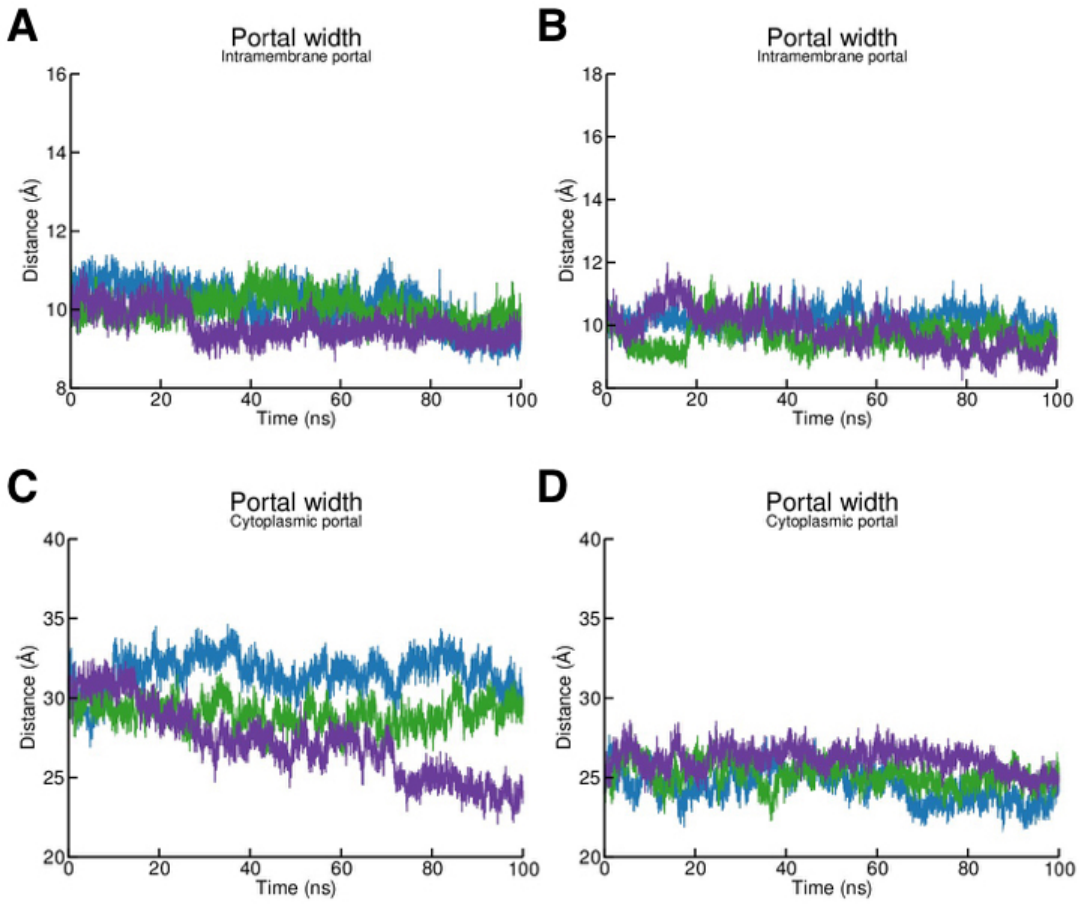


Figure 3.9: (A,B) Distances between intramembrane sections of TMH4/6 and 10/12 respectively. (C,D) Distances between cytoplasmic sections of TMH 4/6 and 10/12. The measurements were made using the same residues as in apo simulations.

decreased from approximately 14 Å to 9 Å(Figure 3.8A), and exactly the same trend was seen with the distance between TMH10 and TMH12 (Figure 3.8B). On the contrary, the cytoplasmic parts of the portals remained open during the 100 ns of the apo P-gp simulations (Figure 3.8C,D). The distance between TMH4 and TMH6 in the cytoplasmic section of the entrance portals varied between 20 to 30 Å during the simulations, while the distance between TMH10 and TMH12 converged at roughly 25 Å. Unsurprisingly, ATP-Mg^{2+} bound P-gp behaved in a very similar way with the membrane sections of entrance portals remaining closed (Figure 3.7C). The distances between cytoplasmic and membrane sections remain mostly converged over the 100 ns of the MD simulations (Figure 3.9).

3.3.4 Conclusions

ATP-Mg^{2+} bound P-gp in the model lipid bilayer behaved similarly to the apo protein, however, a greater degree of stability could be observed from RMSD measurements of protein Cα atoms. Therefore, it appears that binding of ATP stabilises the P-gp structure. This observation is supported by previous MD studies of P-gp that revealed metastable states with ATP and Mg^{2+} ions bound.

The exact pathway that P-gp substrates use to reach their binding sites is still largely unknown. The most likely entrance portals drugs could use are formed by TMH4/6 or TMH10/12 and, judged by crystal structures of various inwards-open proteins closely related to P-gp, provides access from the inner leaflet of the lipid bilayer as well as the cytoplasm. The atomistic simulations of homology models of human P-gp lead to an interesting observation - the membrane sections of proposed entrance portals closed during the MD simulations. Therefore, how compounds access the central drug-binding cavity remains unclear and will require much more detailed investigation.

4

Investigation of P-glycoprotein substrate entrance pathways

As discussed in chapter 3, despite multiple models that have been proposed in the literature, the exact pathway P-gp substrates explore to enter the central cavity and reach the binding site, is still largely uncharacterised and subject to extensive discussions. Due to the atomistic resolution, molecular dynamics simulations could shine light on this controversy and point towards experimentally testable hypotheses.

4.1 Investigation of access to the central cavity of P-gp

Access gates to the central cavity of P-gp are characterised as the gaps formed by either TMH4/6 or TMH10/12. The distance between these transmembrane helices in crystal structures suggests that they could provide access from the inner leaflet of the lipid bilayer[47]. However, these transmembrane helices also extend into the cytoplasm, potentially allowing drug uptake from the aqueous environment as well.

The relationship of P-gp with substrates and inhibitors has been previously explored using molecular dynamics simulations, however, only a few have concentrated on the aspect of drug entrance. Random accelerated molecular dynamics and metadynamics has been used to obtain egress (unbinding) pathways of paclitaxel and daunorubicin[121]. The authors concluded that the most likely unbinding pathway was through the cytoplasmic part of TMH4/6 gate. The entrance pathways supporting either hydrophobic vacuum cleaner or membrane pore model have also been investigated by enhanced MD simulations[122;123].

The simulations of apo and ATP-Mg^{2+} bound P-gp were discussed in chapter 3. In these simulations closure of membrane parts of the potential access gates was observed. This prompted the question of whether the substrate entrance pathway could deviate from the one suggested in the literature and by the crystal structures. Investigation of substrate access to their binding site is unfeasible with unbiased MD simulations due to the limits of timescale of all-atom simulations. Therefore, biased MD techniques such as metadynamics[180] and steered MD[132] have to be used to explore these rare events.

Steered MD (SMD) was used to pull amitriptyline to the central cavity of P-gp through membrane and cytoplasmic sections of the entrance gates. SMD allows us

to calculate the work energy required in a certain process defined by the collective variable, therefore, allowing one to compare different potential entrance pathways.

The starting position of amitriptyline was obtained from nuclear magnetic resonance (NMR) and MD data[181]. The two end points in the central cavity of P-gp were chosen to reflect the point of entrance through the transmembrane helices - at the level of inner leaflet of the lipid bilayer or in the cytoplasm. Since P-gp is pseudosymmetrical, the access routes can be taken through TMH4/6 or TMH10/12, therefore both sides of the protein were investigated (Figure 4.1 A,B). Amitriptyline was placed in the inner leaflet of the lipid bilayer approximately 30-35 Å from the center of the cavity.

4.1.1 Methods

A 50 ns snapshot from an ATP-Mg^{2+} bound P-gp simulation was used as a starting point for SMD simulations. An amitriptyline molecule was inserted on either TMH 4/6 or TMH10/12 side of P-gp in the cytoplasmic leaflet of POPC bilayer by alchembed method[182] and equilibrated within the bilayer for 1 ns. The SMD simulations were calculated using PLUMED 2.0 plugin[136] together with GROMACS v4.6.x. The collective variable used in the SMD simulations was distance between center of mass of amitriptyline and center of mass of 4 residues of P-gp. The spring constant was 1000 kJ/mol and the velocities were 2.5, 1.25 and 0.625 Å/ns. Amitriptyline parameters for GAFF[183] were obtained by antechamber tool[129] from AmberTools14 and ACPYPE[184].

4.1.2 Results

The SMD simulations were run with three different velocities of pulling and the pathways obtained from the separate runs were very similar (e.g. Figure 4.1 C). The

Table 4.1: Work required to enter the central cavity

	Pulling speed (Å/ns)		
	2.5	1.25	0.625
Path 1	513 ±64	534 ±84	490 ±65
Path 2	245 ±12	128 ±12	134 ±19
Path 3	711 ±20	583 ±28	670 ±205
Path 4	299 ±50	391 ±49	358 ±73

Work from the triplicate repeats is expressed in kJ/mol with the estimated standard error of the mean (SEM) .

simulations strongly favoured the cytoplasmic entrance route. The intramembrane pathway required twice as much work energy to reach the central cavity. There is a slight preference for the TMH4/6 side, which, interestingly, has been suggested as favourable by the studies of paclitaxel and doxorubicin egress pathways. The cytoplasmic sections of the entrance sites remained 25-35 Å wide during the simulations, therefore, the distance between transmembrane helices was large enough to accommodate substrates.

Investigating the trajectories further, it was noticed that amitriptyline quickly reached the protein surface, defined as within a 3 Å distance from the closest residue and the work required to reach the protein surface was comparable in all pathways (see Figure 4.2 A-D). Amitriptyline was deemed to have reached the central cavity once it was 5 Å away from the designated destination point in the middle of the cavity and these parts of the work profiles also show similarity between pathways. Therefore, the interactions between amitriptyline and the surface of P-gp accounted for most of the work energy required in the entrance process and were very different between the pathways which took intramembrane versus cytoplasmic routes (see Figure 4.2).

The results of steered molecular dynamics simulations have presented us with a novel entrance pathway consisting of three distinct parts. The drug starts in the inner leaflet of the lipid bilayer and upon closing to the P-glycoprotein surface moves into

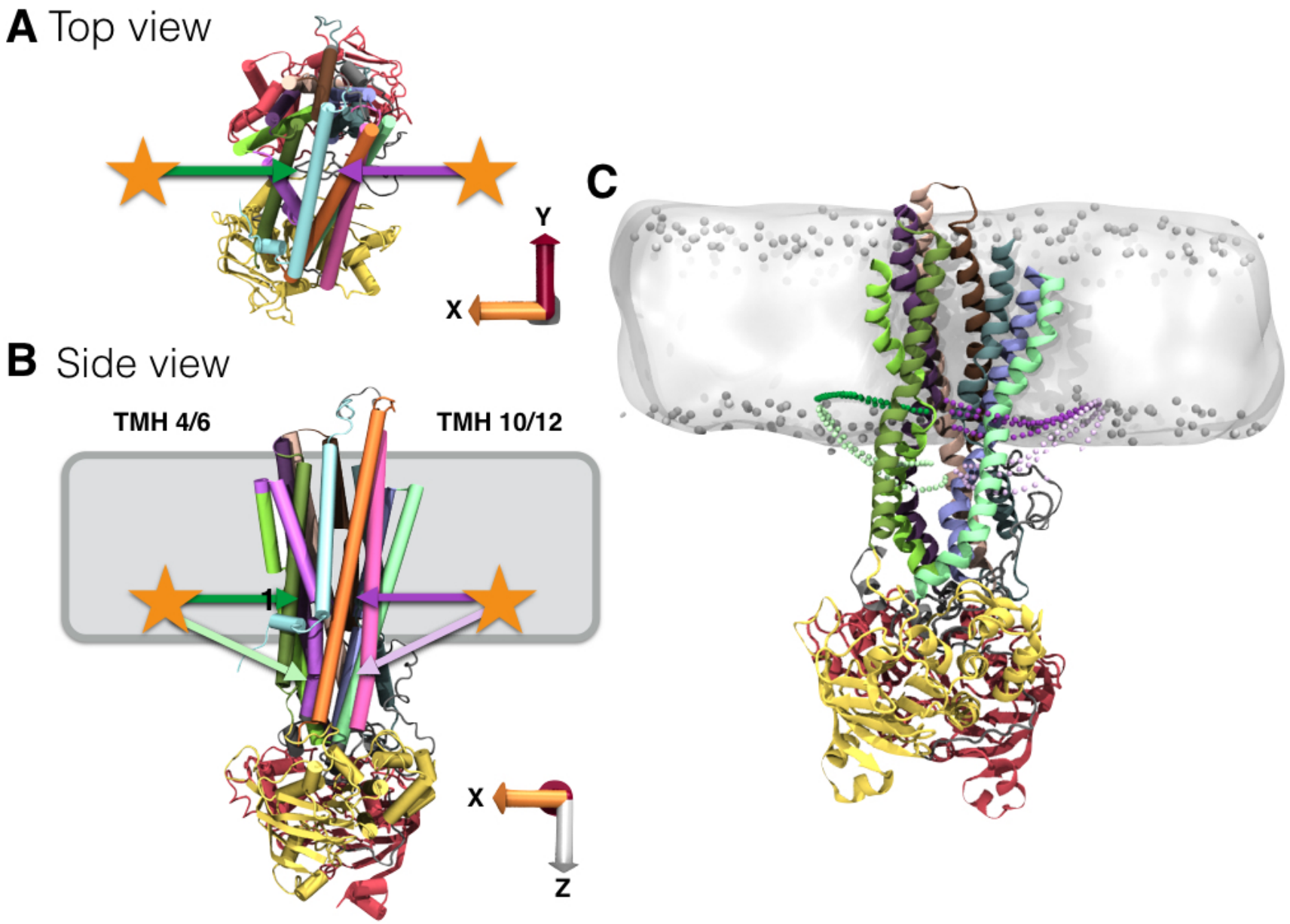


Figure 4.1: Schematics illustrating the pathways investigated by SMD simulations in top (A) and side (B) view with respect to lipid bilayer. Amitriptyline is represented as an orange star and the arrows indicate the direction of pulling by SMD. The approach to TMH 4/6 is shown with light and dark green arrows and the approach to TMH 10/12 shown with light and dark purple arrows. The dark arrows show the proposed pathway through intra-domain entrance portal, while the light coloured arrows show the pulling vector for the cytoplasmic entrance. The arrows are labeled with numbers respective to the collective variables described in the main text. The position of the bilayer is shown in grey. (C) The pathways extracted from the SMD

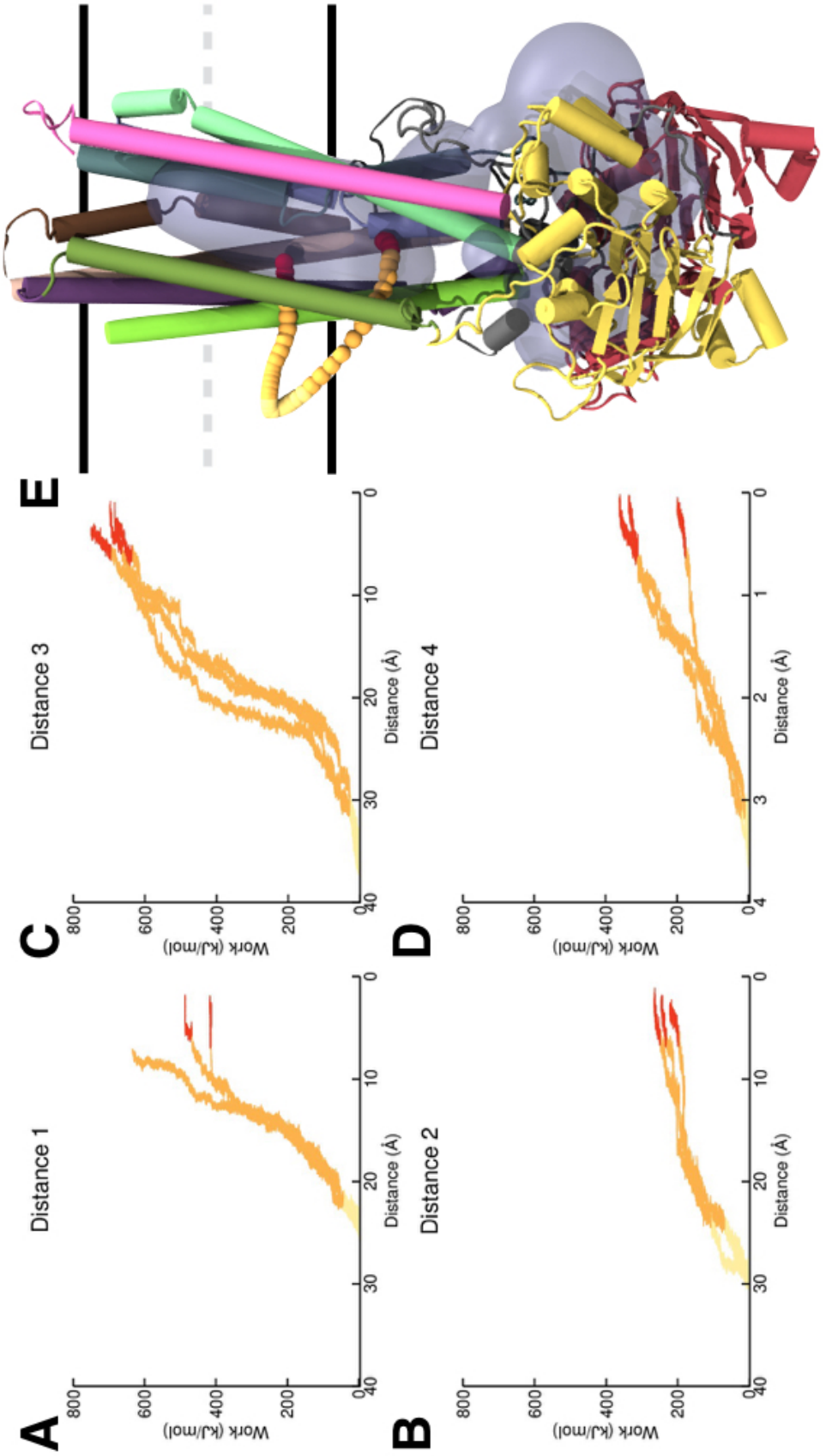


Figure 4.2: (A-D) Work as a function of distance along the pathways 1-4 at 2.5 Å/ns pulling speed(see Figure 4.1 B). (E) Comparison of pathways using the intramembrane or cytoplasmic section of entrance gate. The colouring in parts A-E reflects the position of amitriptyline along the entrance pathway: in yellow, the work profile of the drug within the lipid bilayer or water up until 3.0 Å from the protein surface. The orange part of the work profile describes interactions with the protein, while the profile in red represents the work in the central cavity.

the cytoplasm to enter the central cavity though the entrance gate. These results and previous studies of substrate unbinding suggest that the preferred entrance gate is formed by TMHs 4 and 6.

Ferreira *et al.* have calculated the potential of mean force for substrate and modulator entrance through a gate formed by TMH10/12 and showed that no activation energy is required to pass through this gate[123]. This disagreement in results could be explained by differences in the experimental methods.

4.2 The binding site of amitriptyline

P-glycoprotein has been co-crystalised with various inhibitors, which gives us structural data about the binding site, however, no P-gp crystals with substrates or modulators apart from nucleotides have been published. The binding site of amitriptyline is not known, but evidence does suggest that it is within the central cavity. Therefore, we performed unbiased simulations of amitriptyline from the endpoints of the SMD runs to investigate if there is a preferred position within the central cavity which could be compared to a potential binding site. In a similar study, the diffusion of paclitaxel and doxorubicin to the core of central cavity has been previously observed in MD simulations[120].

4.2.1 Free diffusion of amitriptyline in the central cavity

After the steered molecular dynamics simulations discussed in Section 4.1., amitriptyline was positioned in the middle of the central cavity which is formed like an upwards pointing cone with the binding site proposed to be close to the top of the cavity (see Figure 4.2 E). The endpoints of steered MD simulations were taken and amitriptyline was allowed to explore the central cavity by performing 20 unbiased MD simulations.

In one of them amitriptyline diffused upwards and towards the top of the central cavity (see Figure 4.3 A).

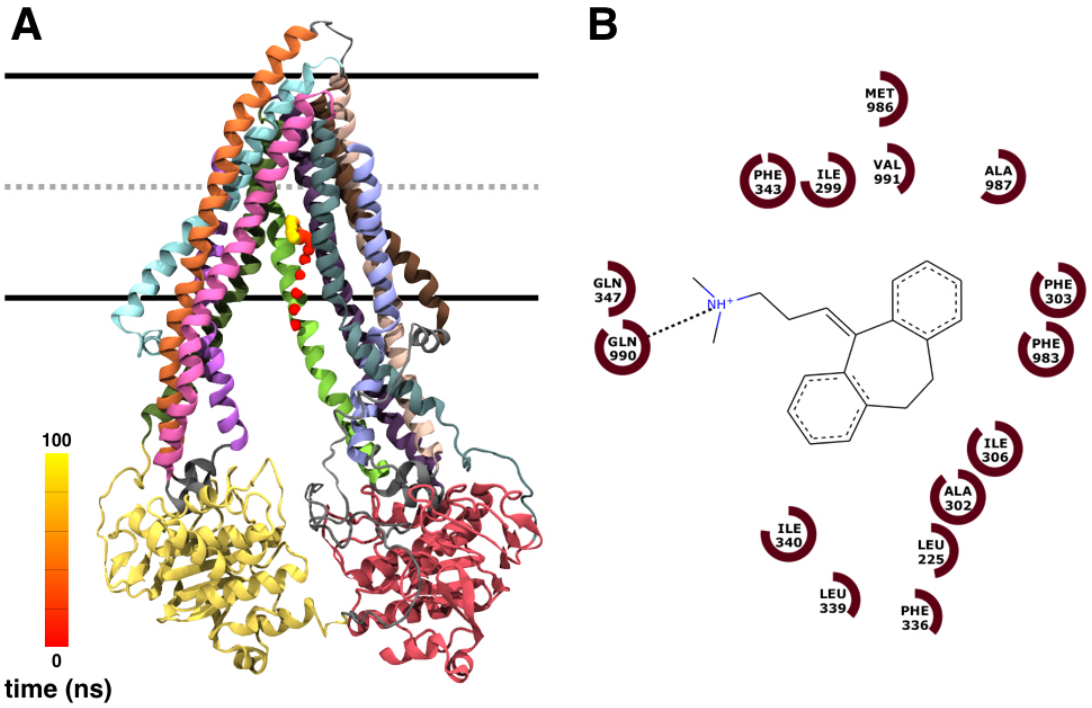


Figure 4.3: (A) The path taken by the centre of mass of amitriptyline indicated as spheres, coloured according to the simulaton timestep (0 - 100 ns as shown by the colour bar). (B) Amitriptyline interactions with P-glycoprotein residues. Each "clock" diagram represents a P-gp residue that had spent at least 30 ns within 3.5 Å of amitriptyline and the length of the interaction is depicted by the fraction of the circle that is coloured.

The diffusion process was fast at the start of the simulation ($\sim$ 25 Å in 10 ns), however, it slowed significantly once the top of the cavity was reached (see Figure 4.3 A). The last 75 ns of the simulation were analysed using LINTools[185]. The results suggest a stable potential binding site with many interactions lasting for more than 50% of the analysed simulation time (see Figure 4.3 B). Most of these residues have been previously proposed to interact with rhodamine[55;186;187]. The protonated tail of amitriptyline also formed a hydrogen bond with the polar residue GLN 990, previously suggested to be critical for substrate and inhibitor binding[32;55].

4.2.2 Methods

The ATP-Mg^{2+} bound P-glycoprotein unbiased trajectories were clustered with the g_cluster tool from the GROMACS package according to their RMSD to obtain 10 divergent protein conformations. Amitriptyline was docked in these confirmations with AutoDock Vina[188] and GOLD[189] software.

The protein and amitriptyline structures were prepared for docking with AutoDock Vina using AutoDock Tools, adding Gasteiger partial charges and polar hydrogens. A grid box of 40 x 40 x 40 Å surrounding the central cavity was used in all docking runs. In the case of docking with GOLD, the GOLD Suite program Hermes visualiser wizard was used to prepare P-glycoprotein and amitriptyline. The search radius was 30 Å with the same center coordinate as in the Vina docking, leading to exploration of the same space within P-glycoprotein using the CHEMPLP[190] scoring function.

The RMSD descriptor from Vina was used to calculate differences in coordinates obtained from GOLD or AutoDock Vina. A pair of Vina and GOLD poses were considered as a consensus pose if their RMSD was less than 2 Å, since this cutoff has been found be to optimal to equilibrate the balance rate of false positives and false negatives. Finally, the resulting consensus poses were clustered according to distance so that any consensus pose within 2 Å of another was assigned to the same cluster. The clusters were analysed by averaging their ranking in scores. The highest scoring cluster was further analysed with unbiased MD simulations. The interaction diagrams were created using LINTools (chapter 6 and[185]).

4.2.3 Consensus docking

The results of the unbiased MD simulations agree with previously published binding sites of rhodamine[55;186;187]. To explore this potential binding site further, consen-

sus docking was employed. This approach has been shown to improve pose prediction[191;192]. Ten different conformations of P-glycoprotein were extracted from MD simulations of the apo human homology model after clustering the states by RMSD (see 4.2.2). Amitriptyline was then docked into the central cavity using AutoDock Vina and GOLD. Clusters of poses were ranked by combined scores from both softwares (see in 4.2.2). The highest scoring cluster was positioned at the top the of the central cavity, similarly to many other previously investigated P-gp substrates (see Figure 4.4 A).

The stability of this pose was investigated by unbiased MD simulations. Amitriptyline remained within the upper part of the central cavity with the RMSD from initial position not exceeding 2 Å (see Figure 4.4 C). Overall, amitriptyline remains stable in its predicted binding site and does not leave during the MD simulations. Different patterns of interactions with the binding site residues were observed in the triplicate repeats (see Figure 4.4 B). Of the 13 interactions present, seven (MET 69, PHE 336, ILE 340, PHE 728, PHE 983, MET 986) were maintained for more than 50 ns in all three simulations. Similarly to the interactions observed in the binding site seen after free diffusion, this binding site exhibited similarity to that of rhodamine, as seen from cysteine and arginine scanning mutagenesis data[55;186;187].

In the simulations after SMD amitriptyline did not reach the binding side seen in consensus docking experiments (see Figure 4.4 A). However, both binding sites share 3 residues (ILE 340, PHE 983, PHE 986), suggesting that amitriptyline might have reached the lower boundary of the "docked binding site". Alternatively, it is possible that amitriptyline has two alternative binding sites, as this phenomenon has been observed with other P-gp substrates[59]. This is further supported by the fact that both putative binding sites show similar interactions to those of the rhodamine binding site[55;186;187].

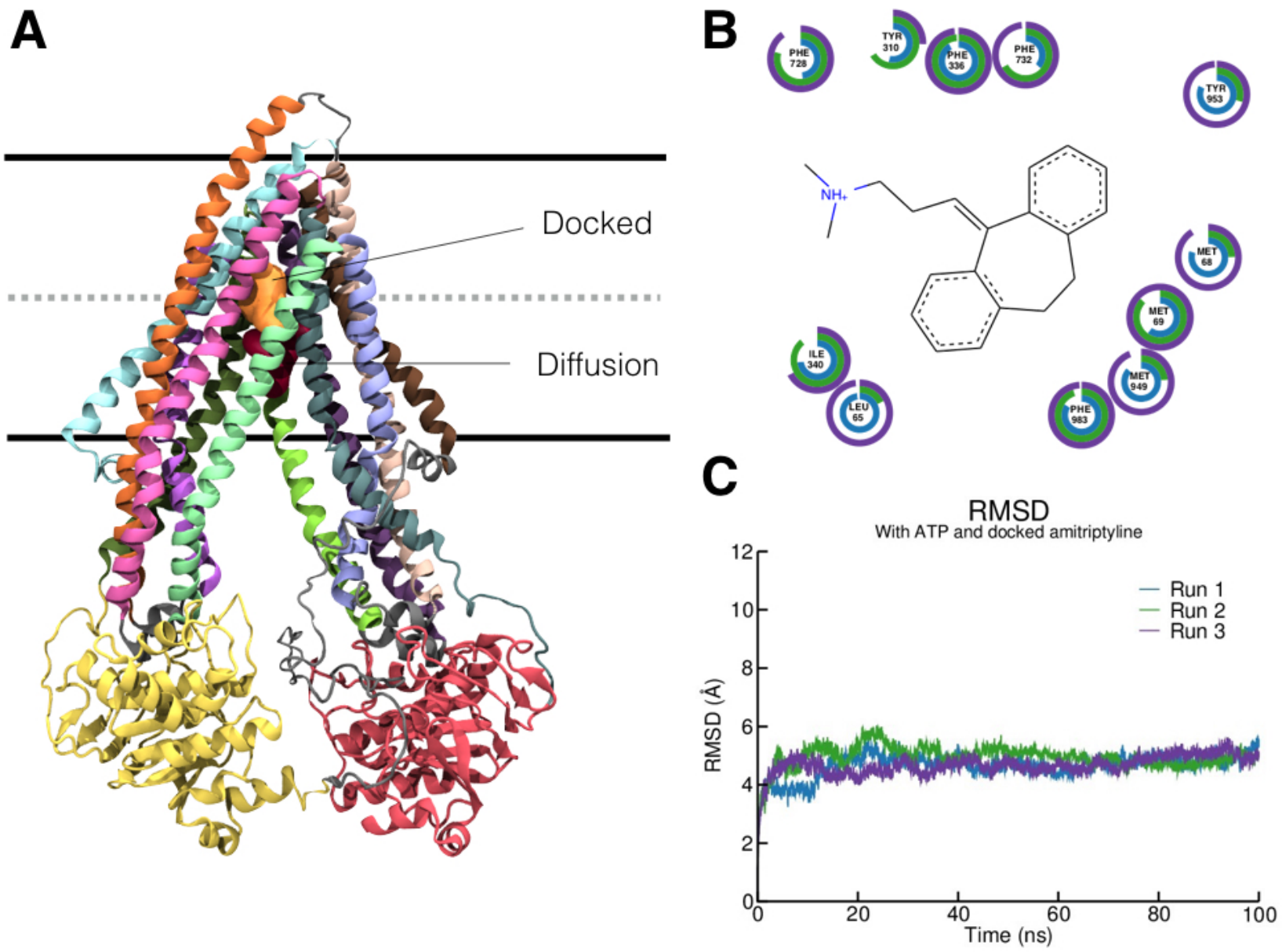


Figure 4.4: (A) The comparison of consensus docking pose with the binding pose observed after free diffusion from the cytoplasmic level of central cavity. (B) The interactions of amitriptyline with P-gp during the 3 x 100 ns simulations. The diagram represents each residue that has spent more than 50 ns within 3.5 Å of amitriptyline and the length of this interaction during each simulation run. (C) RMSD of Cα atoms of P-gp during the simulations with docked amitriptyline present.

4.2.4 Summary of amitriptyline binding sites

One of the simulations of SMD endpoints lead to free diffusion of amitriptyline towards the top of the central cavity and the putative binding site. The final position was equivalent to the level of acyl chains of the outer leaflet of the lipid bilayer. Similar observations were made by Zhang *et al.* regarding paclitaxel and doxorubicin[120]. Even though consensus docking did not uncover the same binding site seen after free diffusion, it can be argued that amitriptyline is exhibiting positional variation within the central cavity. This is further supported by the fact that the sites are sharing common residues. It is also possible that P-glycoprotein substrates do not reach their binding sites by one specified path within the central cavity. Similar observations have been made by O'Mara *et al.*[122].

4.3 The unbinding of amitriptyline

The unbinding (egress) pathways of paclitaxel and doxorubicin were calculated with random acceleration molecular dynamics (RAMD) and it was concluded that these drugs most likely exit through the cytoplasmic section of gate formed by TMH4/6[121]. Protein Energy Landscape Exploration (PELE) was used to investigate unbinding of amitriptyline from the docked pose and the binding site seen after free diffusion (see Figure 4.4 A).

4.3.1 Methods

The structure files from docking or the endpoint of simulations were prepared with the WhatIf server[172] and then the pdb2gmx tool from GROMACS by adding hydrogen atoms according to the OPLS-AA forcefield[143] used by PELE[193]. For each state, 10

independent simulations were produced using 16 processors. The simulation explored the protein energy surface so that if one process moved more than 4 Å away to explore a new area, another process would join it and continue to explore this area together. The simulations were stopped once amitriptyline travelled 60 Å from its point of origin. The simulations were run on the PELE web server[194].

4.3.2 Unbinding pathways of amitriptyline

To gain more support to whichever binding pathway is the most probable, unbinding of amitriptyline was calculated using PELE. It is a stochastic Monte-Carlo approach, which allows a more efficient way of exploring ligand binding and unbinding.

The unbinding of amitriptyline was explored with 40 independent simulations exploring four states - the initial docked pose and the endpoints of 100 ns simulations of docked amitriptyline. The main difference between these states was that the intramembrane sections of TMH 4/6 and TMH 10/12 were open in the initially docked model, but closed after the 100 ns MD simulations.

From the unbinding simulations of PELE, two similar pathways of amitriptyline exit could be distinguished, resulting in the drug either escaping from the central cavity or remaining within it. The orange pathway in Figure 4.5 shows that amitriptyline exits the central cavity through the cytoplasmic section of TMH4/6. This pathway was observed in 65% of simulations. In other cases amitriptyline could not escape the central cavity and remained within it, as depicted by the red pathway in Figure 4.5. However, it could be expected that the substrate could diffuse into the cytoplasm through the cytoplasmic section of one of the portals or the NBDs could separate, allowing access to the cytoplasm straight from the central cavity.

The unbinding pathways observed from PELE simulations give further support to



Figure 4.5: The red pathway shows amitriptyline remaining in the central cavity, while the orange pathway depicts amitriptyline exit into the cytoplasm through cytoplasmic sections of TMH4/6. The lipid bilayer is not present in the actual simulations.

the hypothesis that P-glycoprotein substrates use cytoplasmic sections of entrance portals to reach the binding site. This pathway was observed even if the initially the intramembrane sections of TMH 4/6 and TMH10/12 were open. The protein was not constrained by the model bilayer and could explore any potential conformation state through stochastic calculations of Monte-Carlo simulations, including a more inwards-open state. Binding and unbinding pathways show the same tendency to use the cytoplasmic section of TMHs 4 and 6 to reach or exit the binding site.

4.4 Amitriptyline pathway to its binding site is not linear

Several models for P-glycoprotein substrate entrance have been proposed and the main difference between those models is whether the access is gained from cytoplasm or the lipid bilayer. A lot of studies strongly suggest that P-glycoprotein substrates prefer the hydrophobic lipid bilayer environment. However, closure of the intramembrane section of potential entrance gates was noticed in the molecular dynamics simulations. Therefore, a novel pathway that uses the cytoplasmic section of TMH 4/6 to enter the central cavity was proposed and explored with various molecular dynamics methods.

Steered MD was used to explore substrate entrance through intramembrane and cytoplasmic sections of potential entrance gates of P-gp. The resulting work profiles suggested that substrates experience significantly less difficulty entering through cytoplasmic sections. The work required to enter through the closed intramembrane sections of TMH 4/6 and 10/12 was at least twice as high when compared to a similar pathway through cytoplasmic sections. The work profiles also suggested preference to the gate formed by TMH 4 and 6.

Once in the central cavity after the SMD experiments, amitriptyline could freely diffuse towards the top of the central cavity until it reached a stable position which could be regarded as a potential binding site. Such a path would be hard to define due to the stochastic nature of diffusion and possible hydrophobic interactions with the residues surrounding amitriptyline. Consensus docking suggested a slightly different binding site, which shares residues with the site amitriptyline occupies after free diffusion through the central cavity. This site is positioned slightly higher in the central cavity and some conformational changes could be required to reach it.

Similarly, it is impossible to exclude a possibility that this is an alternative binding site for amitriptyline.

The pathway that is proposed from this study starts by drug diffusion in the inner leaflet of the lipid bilayer. Upon reaching a close distance to P-gp the drug would enter the cytoplasm to enter the central cavity through the cytoplasmic sections of transmembrane helices that form an entrance gate, most likely TMH 4/6. The binding site can be reached by free diffusion through the central cavity, interacting with the hydrophobic residues that line it and a drug could potentially have more than one binding site depending on the conformation of the central cavity.

From the molecular dynamics simulations it is apparent that amitriptyline can reach its binding site through a non-linear pathway that explores various environments moving between hydrophobic interactions and the aqueous environment.

4.4.1 Further experiments

One of the promises of molecular dynamics is to guide experiments through supply of testable hypotheses derived from structural and dynamic data. With regards to P-gp, this approach has been used to understand the ATPase activity[195] and predict the behaviour of P-gp substrates and non-substrates in the lipid bilayer[181].

In the apo and ATP-Mg^{2+} bound simulations of P-glycoprotein, decrease in distances between nucleotide binding sites was observed, most notably in the homology model from *C. elegans* crystal structure, which had the widest opening. These results agree with the experimental evidence for ABC transporters[53;92] that suggests the large separation of NBDs might be an artefact caused by crystallisation / expression in detergents[52].

The simulations described in this chapter lead to the proposal of a novel, previously

undescribed P-gp substrate entrance model. This model can be tested with experimental methods to confirm or reject the hypothesis since the data acquired from simulations can give a clearer idea where to look for evidence.

In many instances the distances between NBDs have been measured by crosslinking[92;195] or luminescence resonance energy transfer (LRET) spectroscopy[53] experiments to understand the ATP hydrolysis cycle of ABC transporters. However, no detailed investigation has been done on the degree of opening of TMDs which could further the understanding of substrate import and export. Similar crosslinking or spectroscopy experiments could be done with the residues in the intramembrane sections of the entrance gate forming transmembrane helices, where the numbers and trends could be compared with MD data.

Amitriptyline is relatively poorly explored as a P-gp substrate and not much is known about its interactions with P-gp or indeed where it binds. Due to its pharmacological profile as a tricyclic antidepressant, amitriptyline would interact with P-glycoprotein at the blood-brain barrier. As such, amitriptyline could be important for research of P-glycoprotein behaviour in the blood-brain barrier. Potential experiments that could support the MD data available in this chapter are liquid-state NMR or mutagenesis experiments to gain more information about the interactions in the binding site.

The biased MD methods used in this chapter are relatively coarse and do not give the full picture of the entrance pathway. Similarly, the binding energy was not measured. More sophisticated molecular dynamics methods such as metadynamics and free energy pertubation can be used to gain additional insight once the initial data is obtained. In this case, the pathway of substrate entrance has been coarsely characterised and can be used to build a path collective variable to be used with metadynamics simulations. The two potential amitriptyline binding sites can be assessed with alchemical methods such as free energy pertubation to get a good estimate of

the binding energy.

5

P-glycoprotein in a complex lipid bilayer

The plasma membrane of human cells is a complex assembly of proteins and lipids that protects the cell from the outer environment. Lipid membranes consist of hundreds of different lipids and their distribution within membranes can be highly assymetric. In molecular dynamics simulations of membrane proteins, very simplistic model bilayers consisting of single lipid species are typically used to provide an appropriate hydrophobic environment. However, such an approach gives only limited information about the protein-lipid interactions and the influence of different lipid species on protein behaviour or vice versa. This problem is also important in the context

of P-glycoprotein which is known to be influenced by phospholipid and cholesterol content of the bilayer (for a review see[29]).

In this chapter P-glycoprotein is inserted in a complex asymmetric bilayer with a composition reflecting that of human brain epithelial cell membranes. Multiscale molecular dynamics simulations (coarse grained and atomistic) are employed to investigate protein-lipid interactions and to characterise the bilayer itself in a separate study. P-gp interactions with sterols are studied in greater detail as there is much uncertainty about the nature and importance of these interactions (e.g.[102;105]). In particular, the potential location of the cholesterol binding sites was tested with a biased MD technique - umbrella sampling. This approach allows one to assess the binding affinity of cholesterol to certain sites. The overall trends observed in coarse grained simulations were further investigated by a comprehensive set of atomistic scale simulations of P-gp in the same complex bilayer environment. Following the discussion started in chapter 3, the behaviour of potential substrate binding sites is also examined to gain information of the potential role of lipids in the substrate transport and P-gp conformational changes.

5.1 Blood-brain membrane lipid bilayer

It is important to study P-glycoprotein behaviour in a physiologically relevant environment as much as it is possible and feasible within a molecular dynamics simulation setup. In chapter 3, the difference between simulations with and without ATP and Mg^{2+} was discussed and it was deemed to be of importance to include the nucleotides in the system. In this chapter, we investigate the role of complex asymmetric bilayer in the stability and conformational changes of P-gp and have designed an asymmetric lipid bilayer with similar composition to that of blood epithelial cell membrane that

should replicate the bilayer at the blood-brain barrier.

However, since this particular composition of a bilayer has, to our knowledge, never been used in MD simulations before, it is of importance to first check whether it shows similar properties to previously investigated bilayers. In this case, we looked at the diffusion of each individual lipid species and the distribution of cholesterol. These characteristics were tested at both coarse grained and atomistic resolution and compared to the available data for bilayers with similar compositions.

The design of the bilayer depended on two factors - similarity to an experimentally determined brain epithelial cell bilayer and the availability of lipid parameters in both Martini (coarse-grained forcefield) and Amber (used in atomistic resolution) force fields. As a result, the complexity of the bilayer described in this chapter is defined by its composition from several phospholipid species, including anionic lipids, as well as sphingomyelin and sterol (see Table 5.1). The identity and total percentage of lipid headgroups closely replicated the experimentally determined composition of human brain epithelial cell membrane[196]. The initial lipid distribution between leaflets was asymmetric with PC and sphingomyelin appearing mostly in the outer leaflet and PE and PS being the dominant lipid species of the inner leaflet (Table 5.1) as predicted for a mammalian cell bilayer[197]. Phospholipid hydrocarbon tails were of similar length and in all cases were composed of one saturated and one unsaturated fatty acid (Figure 5.1).

The bilayer behaviour was investigated with multiscale molecular dynamics simulations due to the larger timescales achievable with this method. Coarse grained simulations were run in triplicate for 10 µs, while the atomistic simulations were run in triplicate for 100 ns. The difference in resolution allows comparison both between methods and with available experimental data. Two descriptors were analysed in both sets of simulations - lipid diffusion and distribution across leaflets.

Table 5.1: The composition of the complex asymmetric bilayer

Lipid	Outer leaflet	Inner leaflet	Total bilayer in simulation
POPC	45%	25%	35%
POPE	15%	35%	25%
POPS	0%	20%	10%
Sphingomyelin	20%	0%	10%
Cholesterol	20%	20%	20%

The complexity of the bilayer reflects the composition of blood epithelial cell membrane[196]. The asymmetry was determined by trends observable in mammalian cells[197]. The percentage lipids of total bilayer in simulations reflects the total percentage of the particular lipid headgroups in both leaflets of the bilayer.

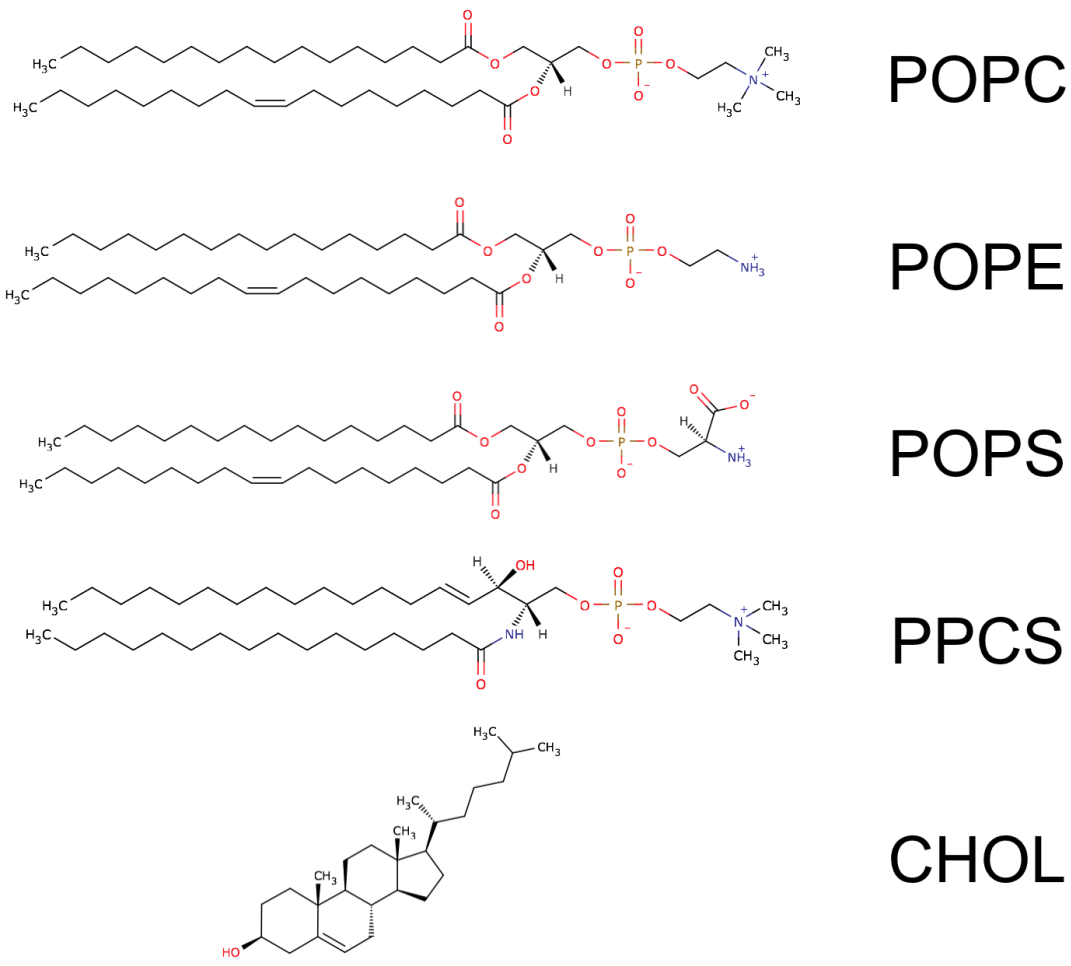


Figure 5.1: Illustration of the lipids present in the complex asymmetric bilayer used in this chapter. Marvin was used for drawing the chemical structures of lipids (MarvinSketch 6.1.3, 2013, ChemAxon, http://www.chemaxon.com).

5.1.1 Methods

The initial bilayer contained 500 POPC lipid molecules and was created by coarse grained self-assembly. It was then converted to a complex asymmetric bilayer by "editing" the POPC bilayer[198;199]. The coordinates of phospholipids (POPE, POPS, PPCS) with the same or shorter chain length as POPC were transferred from a randomly selected POPC molecule, followed by relabelling and particle type alteration. Cholesterol was superimposed on the head group and 3 of the hydrophobic tail beads of a POPC molecule. The coarse grained simulations were run according to the description in chapter 2. The diffusion of lipids was measured using "Diffusion analysis in MD simulations" script[200]. The mean square displacement (MSD) of various lipid species was calculated separately. To improve the speed of calculations only every 100th frame was analysed. The window sizes for MSD calculation are measured in frames, not time values as simulation output frequencies and timesteps may vary. The window sizes were set up in such a way that the separation between values was small at the start and larger when 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, MSD was also fitted to an anomalous diffusion model with Equation 5.1 to investigate the nature of diffusion of the lipids[201]. 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 subdiffusion where the mean square displacement will grow less than linear in time. The values obtained from this analysis were plotted using Matplotlib[202].

$$MSD = ND_{\alpha}t^{\alpha} \tag{5.1}$$

The distribution of lipids was determined by g_density tool from Gromacs. The ratio of cholesterol in leaflets of lipid bilayer for coarse grained simulations were determined

using in house scripts.

The bilayer was converted to an atomistic resolution using the Backward tool from the Martini toolkit[153].

5.1.2 Description of complex asymmetric bilayer by multi-scale MD simulations

Anomalous diffusion (subdiffusion) was observed for all lipid species in both coarse-grained (CG) and atomistic simulations (see Figure 5.2), as could be expected from experimental data of anomalous dye molecule diffusion in lipid membranes[203]. In coarse grained simulations the diffusion constant of all lipids was 2873 ± 360 Å^2/µs (SD) with cholesterol diffusing slightly faster than the phospholipids (3085 ± 316 Å^2/µs (SD)). The diffusion constant is within close agreement with coarse grained simulations of a large patch of idealized mammalian plasma membrane[204]. A lower diffusion constant (1829 ±1060$\mathring{A}^2$/µs (SD)) and, therefore, slower diffusion of lipids was observed in the atomistic resolution simulations, however, cholesterol was still faster than phospholipids (see Figure 5.2 A,C). Similar trends and diffusion constants have been observed in atomistic simulations of various phospholipid bilayers with 20% cholesterol[205].

Even though at the start of CG simulations cholesterol is evenly distributed across the leaflets, an asymmetric distribution with more cholesterol molecules found in the outer leaflet is observed (see Figure 5.3 A). This trend is even more pronounced in the atomistic simulations, partially explained by the increased amount of cholesterol found within the hydrophobic core (see Figure 5.3 B). Observations of trajectories revealed flip flopping of cholesterol, the extent of which can be estimated from the cholesterol distribution within the core region (see Figure 5.3). Both asymmetric

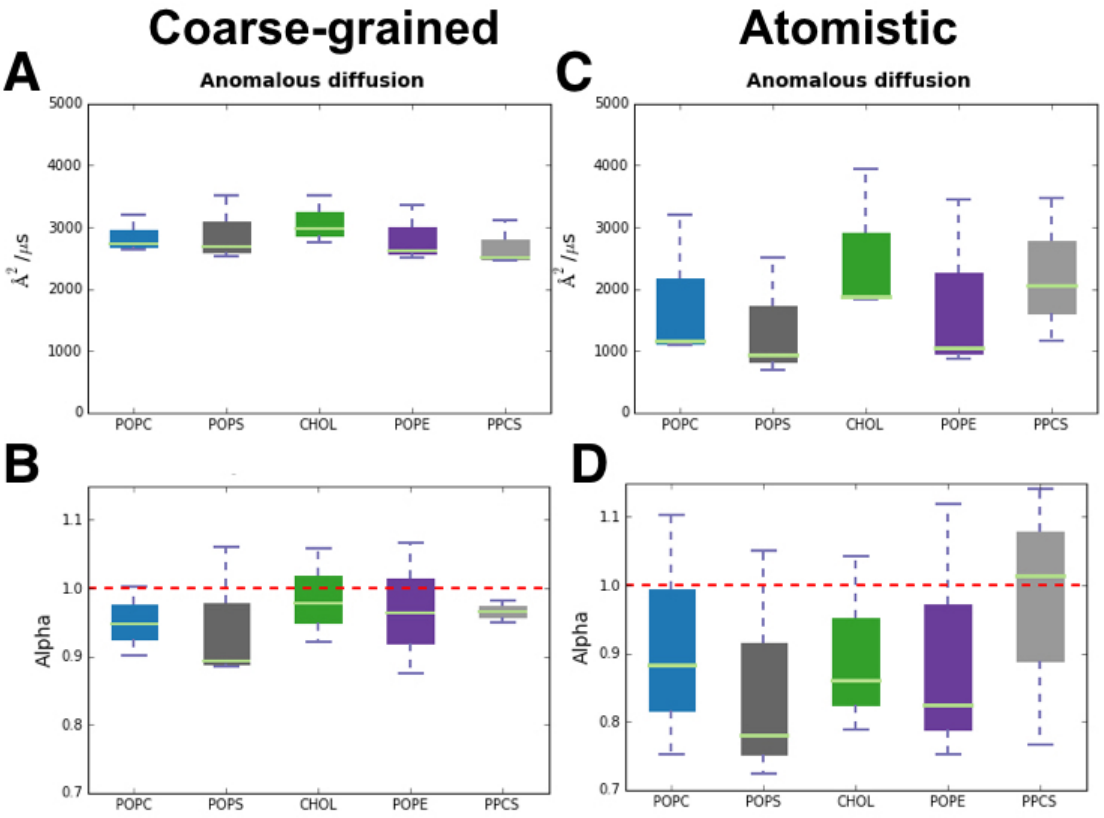


Figure 5.2: The lipid diffusion constants for coarse-grained (A) and atomistic (C) MD simulations of the asymmetric lipid bilayer patch shown as boxplots. Boxes show the interquartile range (IQR), error bars show the full range excluding outliers (points) defined as being more than ± 1.5 IQR outside the box. The results for MSD fitting to anomalous diffusion model for CG (B) and atomistic (D) simulations suggest that subdiffusion was observed for all lipid species, as the α value from fitting is lower than 1 [201].

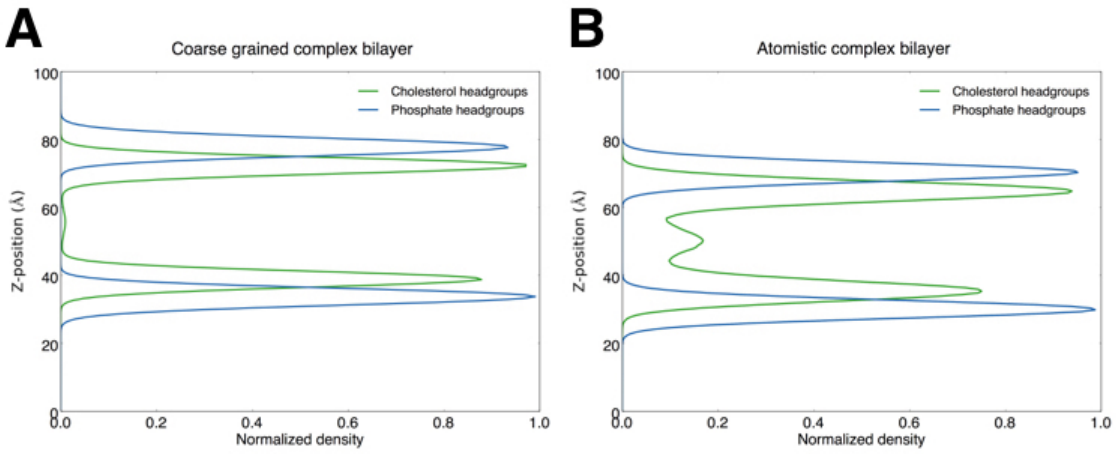


Figure 5.3: Distribution of phospholipids (blue line) and cholesterol (green line) headgroups in CG (A) and atomistic (B) MD simulations show asymmetric distribution of cholesterol.

distribution and fast flip flop rates have been observed in coarse grained simulations of cholesterol containing bilayers previously[204].

5.2 Coarse grained MD simulations of P-gp in complex bilayer

The multiscale studies of the asymmetric, brain cell epithelium-like lipid bilayer described in the previous section were satisfactory, as they agreed with previous experiments done on membranes with similar and higher complexity[204;205]. This composition was therefore used in all following simulations of a P-glycoprotein-bilayer system shown in Figure 5.4 A. Two inwards-open states were investigated in this chapter - the initial human homology model from mouse refined template with separation between center of mass of NBDs - 44.1 Å, further annotated as "wide conformation" and the same homology model with reduced separation between NBDs (COM distance 36.7 Å) - "narrow conformation" in the rest of text (see Figure 5.4 B). The latter conformation was obtained from a snapshot of one of the P-gp simulations discussed

in chapter 3. This was done to gain more information on the conformational changes in the complex bilayer when compared to pure POPC bilayer, mainly in respect to substrate entrance gates.

Similarly to the lipid bilayer simulations described in the previous section, the protein-bilayer complex was also investigated by multiscale MD simulations with similar lengths, however, in this section only coarse grained MD simulations are discussed. We investigated the protein-lipid interactions to understand the impact of P-gp on the bilayer environment. 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. The relationship between P-gp and anionic lipids or cholesterol was further examined by the visualisation of density maps of the lipid head group spatial occupancy around the protein during the simulations. Several potential cholesterol binding sites were observed using this method and their stability during the simulations was determined.

5.2.1 Methods

All simulations were run using the MARTINI CG force field for lipids[150] and proteins[149]. The models were converted to a coarse grained resolution using martinize.py script (available from [i]) which creates input files for proteins compatible with Gromacs and Martini forcefield[206]. The coarse grained protein was inserted in a lipid bilayer consisting of 500 POPC molecules using the self-assembly method. After 25 ns self-assembly simulation the POPC bilayer was formed completely around the protein. It was converted to a complex asymmetric bilayer[198;199]. Since the conversion provides different lipid coordinates after every run, this procedure was done 10 times to make 10 independent repeats. After conversion, a 100 ns equilibration run for each repeat

[i] http://md.chem.rug.nl/index.php/tools2/proteins-and-bilayers/204-martinize

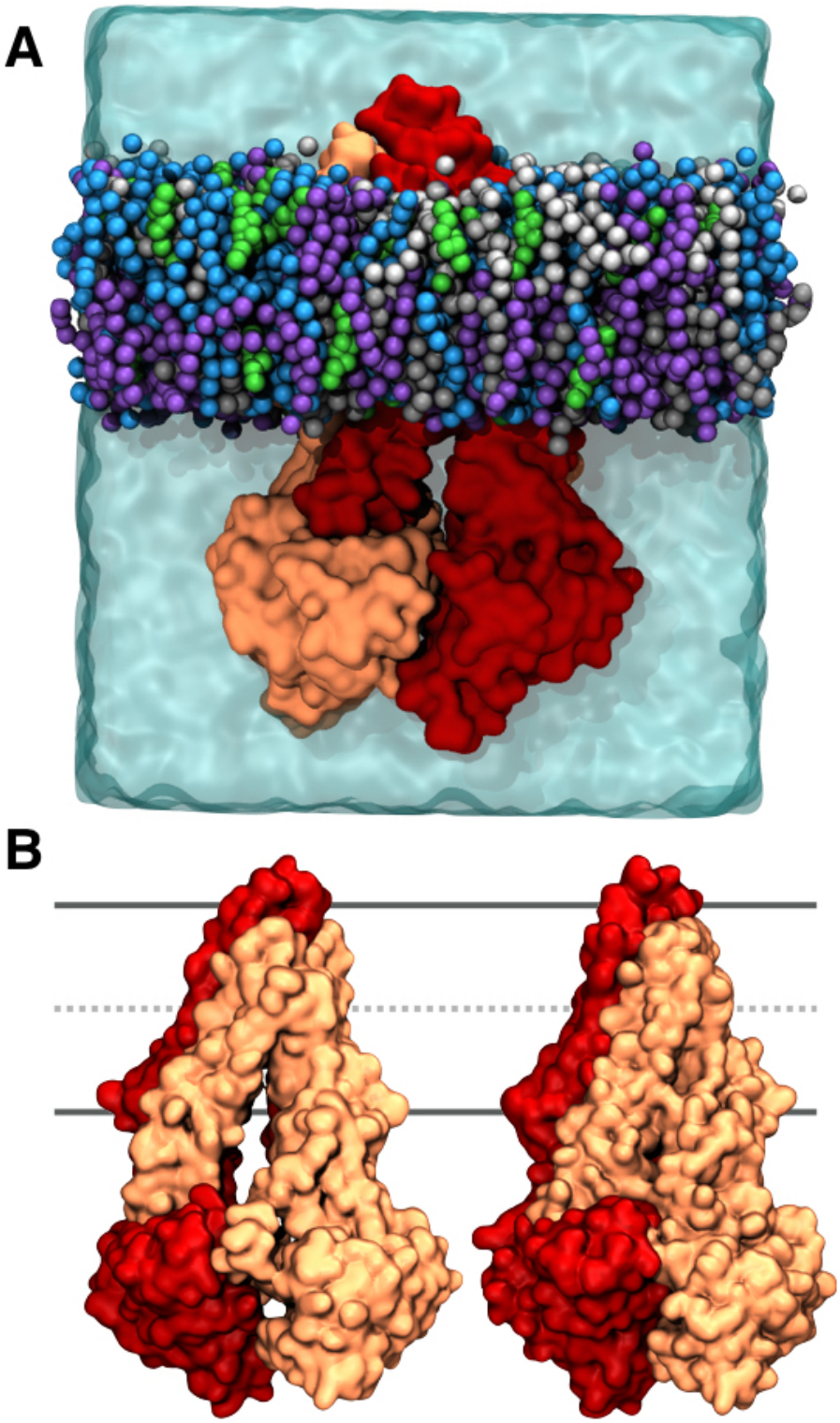


Figure 5.4: (A) P-glycoprotein in a box of water and inserted in the complex asymmetric bilayer. (B) On the left side of panel the "wide conformation" of P-gp homology model is shown with regards to the lipid bilayer borders, while the "narrow conformation" is shown on the right side of the panel. TMD1/NBD1 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 (PPCS) in light gray and cholesterol in green. This color scheme is used throughout the chapter.

was done with protein backbone beads positionally restrained. The production runs were 10 microseconds long. The radial distribution was analysed using g_rdf function from Gromacs package. Lipid densities were analysed and displayed with VMD VolMap plugin. The cholesterol binding sites were determined using MDAnalysis[207] and in-house Python scripts.

5.2.2 P-glycoprotein strongly affects the local bilayer environment

A distinct pattern of lipid ordering around P-glycoprotein was observed from the spherical radial distribution functions (see Figure 5.5). For all lipid species, the highest probability distance is observed around the first interaction shell at the distance of 5 Å. However, other peaks with decreasing probability can be observed at 10, 15 and sometimes 20 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 for much smaller systems such as juxtamembrane domains[208] and is therefore unsurprising for a larger, more complex protein. P-gp incorporation in PC:PE bilayer lead to altered fluidity and modified packing organisation of fluorescent lipid probes[101].

The visualisation of density maps of different lipid headgroup spatial occupancy in Figure 5.6 A,B can reveal further details of lipid ordering. POPS is observed to be the most abundant lipid in the first interaction shell of the inner leaflet, forming an annular ring around the positively charged residues of P-gp (Figure 5.6 B,C). In comparison, such a formation was not observed on the outer leaflet - lipids appeared to be more evenly spread with no particular lipid species exhibiting preference to P-gp which can be explained by lack of POPS in the outer leaflet (Figure 5.6 A). Similar affinity for anionic lipids has been observed in other ABC transporters[209]. This feature is proposed to be important for the protein function as delipidation

causes a decrease in ATPase activity[209]. In a mass spectrometry study of purified P-gp the apparent K_d values suggested more preferential binding of negatively charged lipids[99] and the lipid species that were associated with purified P-gp were mostly PS and PE with small quantities of PC[98].

Diffusion constants of lipids in bilayers with P-gp were lower in comparison to the previously analysed bilayer patch (see Figure 5.7). Additionally, a significant degree of slowing down is observed for POPS when compared to other phospholipids. These differences can be easily explained by insertion of the protein in the bilayer and the annular ring formed by POPS. Lipids are still expected to diffuse anomalously, albeit the α constant is increased, suggesting that insertion of P-gp has an influence on the 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[210]. The ring-like lipid ordering discussed above could also contribute to the lower diffusion constants.

From the analysis of averaged lipid density maps over the cumulative 100 microsec-

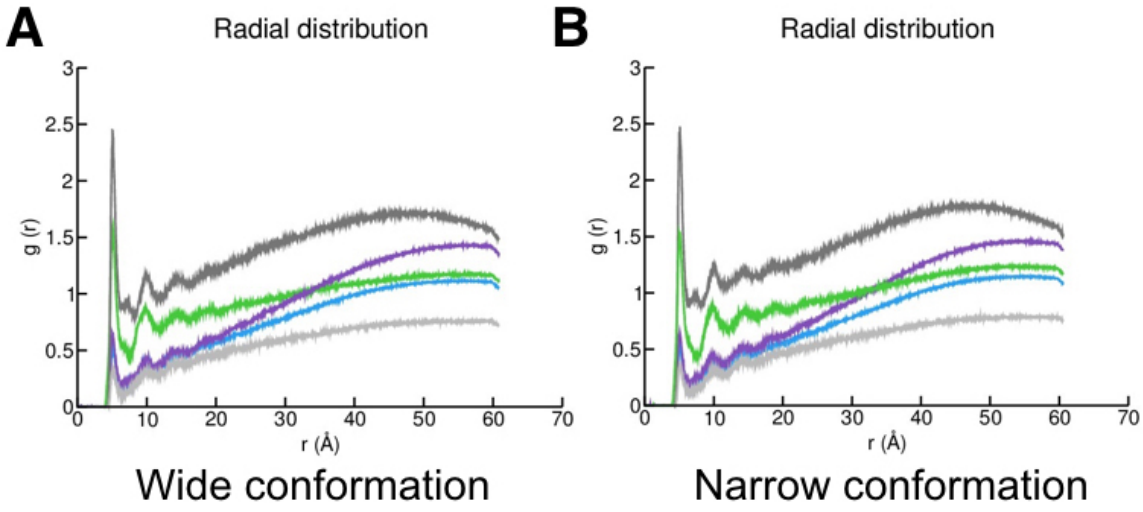


Figure 5.5: The spherical radial distribution functions for lipids for CG simulations of P-gp in a complex bilayer for (A) "wide" and (B) "narrow" conformations of P-gp. POPC is represented as a blue line, POPE as a purple line, POPS - dark grey, sphingomyelin (PPCS) as a light grey line and green line for cholesterol.

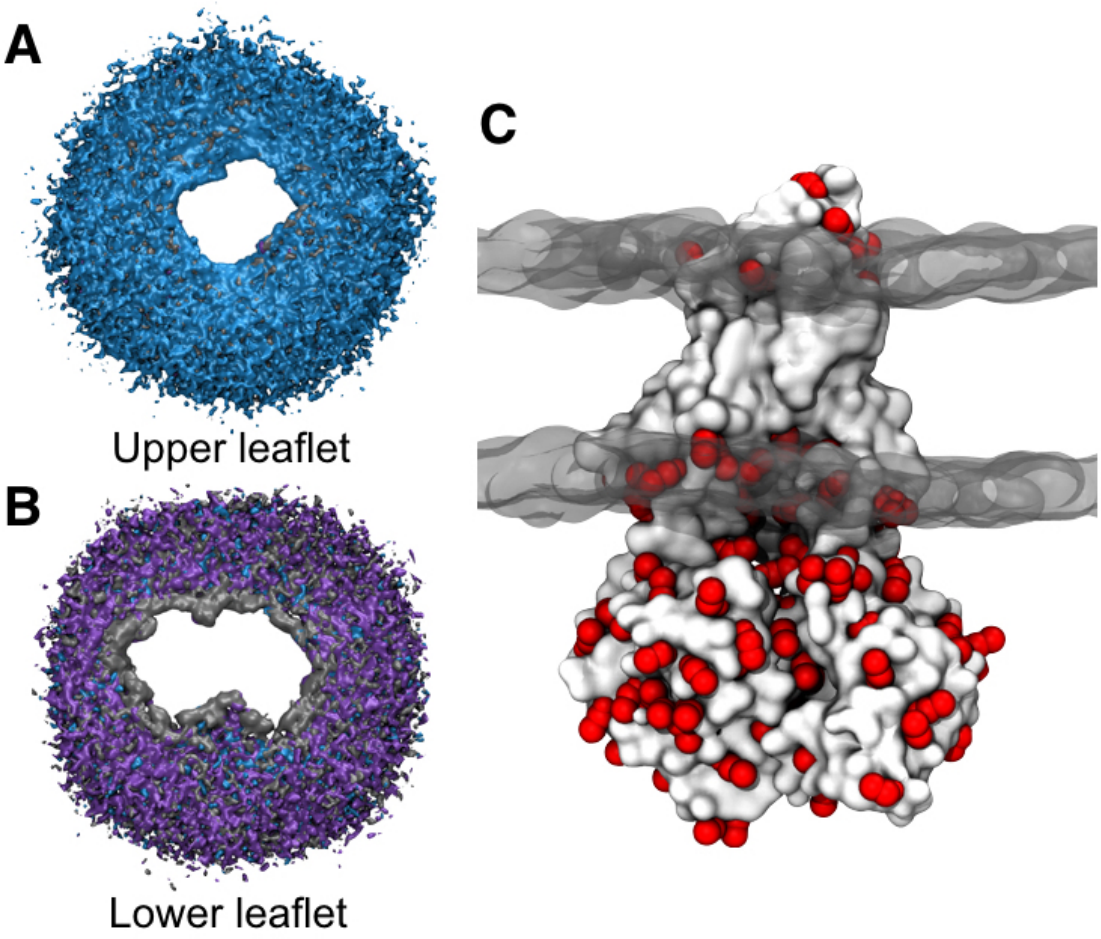


Figure 5.6: The density maps of phospholipid headgroup spatial occupancy for upper (A) and lower (B) leaflet of the complex asymmetric bilayer with POPC (blue), POPE (purple), POPS (dark grey) and sphingomyelin (light grey) headgroup densities shown. (C) Representation of positive residues (lysines and arginines) in red on P-gp surface. Phospholipid headgroup regions depicted as grey surfaces.

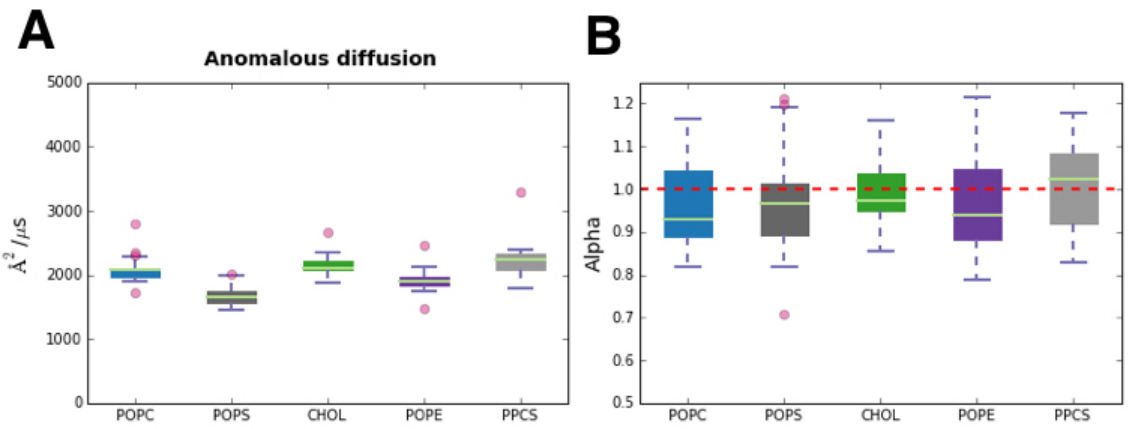


Figure 5.7: (A) The lipid diffusion constants for coarse-grained MD simulations of P-gp inserted in asymmetric lipid bilayer patch shown as boxplot. Boxes show the interquartile range (IQR), error bars show the full range excluding outliers (points) defined as being more than 1.5IQR outside the box. (B) The results for MSD fitting to anomalous diffusion model suggest that subdiffusion was observed for all lipid species, as the α value from fitting is lower than 1[201]

ond simulations, cholesterol could be found in the clefts of the P-gp surface, most prominently in the gaps between TMH 4/6 and TMH10/12 - the putative intramembrane entrance gates. Such a lipid binding arrangement, known as nonannular surface lipids, is not unusual for membrane proteins that exhibit clefts, cavities or pores on their surface[211]. In general, two conditions have to be fulfilled for lipid interaction to be considered stable - the lipid headgroup forms hydrophilic interactions with the protein, while the tails have a nonpolar surface of the protein available for interaction[211].

The simulations were analysed to define cholesterol binding sites by scoring the binding site residues that remain within close distance (< 6 Å) of cholesterol for at least half of the simulation length. By far the most stable potential binding site was positioned in the gap between TMH10/12 (see Figure 5.8). It was visited in 65 % of all simulation time and more than three quarters of residues associated with this binding site remained within cholesterol during most of the simulation, strongly suggesting this as a potential high affinity binding site.

A site situated between TMHs 7/8 was also of interest because it was shown to be relatively stable (cholesterol remained bound for at least half of the simulation time) in the few simulations it was present and it was close to a CARC motif, suggested to form a cholesterol binding site for P-gp[105]. Another site that exhibited similar behaviour was present in close vicinity of TMH 9/12. All of these binding sites fulfilled the criteria for stable nonannular surface lipid interactions with P-gp, as their headgroups were interacting with basic residues (LYS or ARG) and the tails generally took up the non-polar surface of transmembrane domains (see Figure 5.8).

Insertion of P-glycoprotein in the complex bilayer leads to changes in local environment, mainly exhibited by ordering of lipids around the protein in ring-like patterns and the decrease in diffusion rates of lipids. The most affected lipid species are the anionic POPS, which forms annular ring around the positively charged residues of P-gp, and cholesterol that appears to bind P-gp in specific sites formed by the grooves and clefts of the protein surface. The potential cholesterol binding sites determined from the simulations need more detailed investigation to understand the nature and strength of the interactions.

5.2.3 Lipid access to the central cavity

P-glycoprotein has been proposed to function as a lipid flippase and there is substantial evidence for this. An altered distribution of fluorescently labelled phosphatidylcholine (PC)[212;213], phosphatidylethanolamine (PE)[213], sphingomyelin (SM)[214] and phosphatidylserine (PS)[215] has been observed in cells overexpressing P-gp. Inhibitors of P-gp block the lipid transport in a dose dependant manner, similarly to that of known P-gp substrates. This might suggest similarities in the transport of lipids and drugs.

In the coarse grained simulations of the human homology model of P-gp in the wide

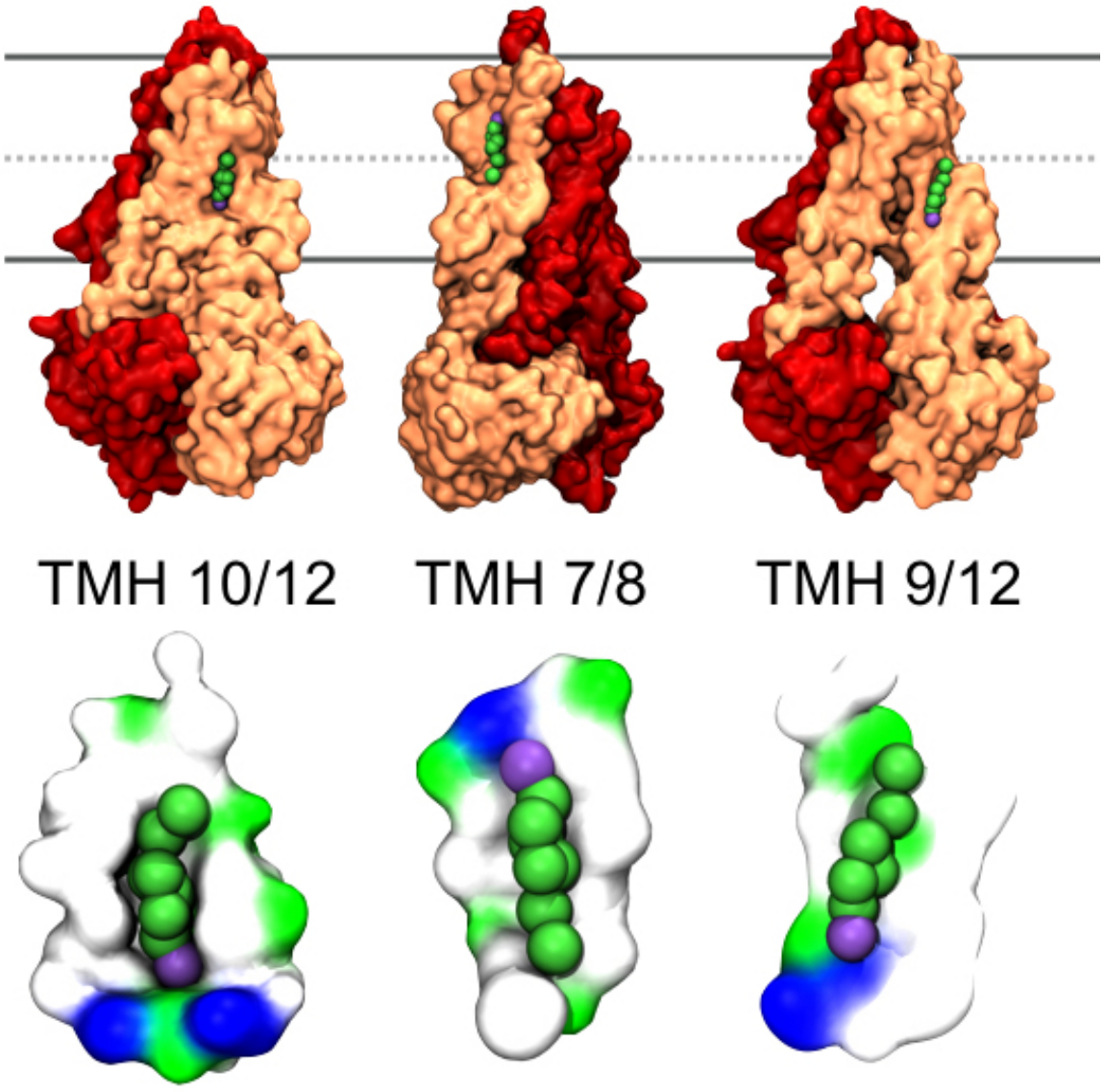


Figure 5.8: The top panel shows the location of cholesterol (green spheres) binding sites on P-gp surface with TMD1/NBD1 colored in red and TMD2/NBD2 colored in orange. Cholesterol headgroup bead is purple. The cholesterol binding sites are shown in more detail on the lower panel, where P-gp residues within 10 Å of cholesterol are shown as surface with positive residues in blue and polar residues in green. Cholesterol headgroup bead (purple) is interacting with positive residues in all binding sites.

conformation several cases of partial lipid entrance were observed. In all cases the putative intramembrane substrate entrance gate formed by TMHs 10 and 12 was utilized by the lipids. Similar events have been captured in atomistic resolution MD simulations of mouse crystal structure of P-gp, where a partial entrance of lipid tail was observed[124]. However, in this case a model bilayer made from POPE was used and the lipid tail entered through the substrate entrance gate formed by TMHs 3, 4 and 6.

Partial entrances of one or more lipids were observed in 6 out of 10 coarse grained simulations of the wide conformation of P-gp. The lipid species observed to enter were POPC, POPE, CHOL and POPS - all species that are present in the lower leaflet of the complex bilayer. Sphingomyelin is a notable exception, especially since it has been shown to be flipped by P-gp, however, it can be easily explained by the fact that sphingomyelin is present exclusively in the outer leaflet of the bilayer used in simulations. While fluorescent species of phospholipids are flipped to the outer leaflet by P-gp, it is unlikely that cholesterol is the target for P-gp flippase action[102]. Cholesterol entrance could be explained by the "cholesterol fill-in" model, where the lipid occupies the empty space in the substrate-binding pocket[103;104].

In most cases, the headgroup of a lipid enters the central cavity of P-gp, however, the tail of POPS has also entered the potential binding site. Phospholipids often enter accompanied by cholesterol which binds or enters TMH 10/12 either before or during phospholipid entrance. Cholesterol would normally use part of the entrance portal that is closer to the hydrophobic core region, while the phospholipid headgroups would use the portal through a gap close to the headgroup region. Such behaviour is very unsurprising if lipid headgroup distribution is taken into account.

5.3 Potential of mean force simulations of potential cholesterol binding sites

Cholesterol lipid binding sites have been attributed to sequence interaction motifs such as CARC/CRAC motifs. These motifs are also present in P-gp and have been suggested as potential cholesterol binding sites[105]. The notion that P-gp might directly interact with P-gp came from experiments where P-gp co-eluted with [^{3}H]cholesterol[103].

Lipids that bind to specific sites tend to position themselves in the grooves of the protein surface or in the spaces between domains, therefore potentially stabilizing the structure. In a recent study cholesterol interaction motifs within nicotinic acetylcholine receptor transmembrane domain were identified using computational methods[216]. These sites agreed with the structural data from cryo-EM, as they formed gaps in the original structure. When the role of cholesterol in these sites was examined with MD simulations, absence of cholesterol caused collapse of the structure, while the presence of bound cholesterol allowed it to maintain the original conformation.

Some potential cholesterol binding sites were observed in coarse grained simulations of P-gp in the complex asymmetric bilayer (see Figure 5.8). These binding sites were positioned in the clefts and crevices on the P-gp surface and a closer look revealed interaction patterns characteristic to surface nonannular lipids. The stability of cholesterol in these binding sites was assessed by measuring the residence time during the simulations. However, this gives only a very indirect indication of the binding site strength. Clay *et al.* suggested that the cholesterol binds P-gp with 3-9 µM affinity[105]. Umbrella sampling simulations were used to calculate the potential of mean force of cholesterol binding for each of the sites proposed in Section 5.2.

5.3.1 Methods

A frame with cholesterol bound to the site of interest was extracted from a 10 µs production run. The protein-lipid complex was inserted in a POPC bilayer by a 50 ns long self-assembly simulation. P-gp and cholesterol were restrained during the self-assembly run using restraints of 1000 kJ/mol/nm^2. Self-assembly was followed by a 10 ns equilibration run during which protein was restrained with 400 kJ/mol/nm^2 restraints on the XY axis to prevent rotation and translation while cholesterol and POPC lipids were unrestrained. Configurations along the collective variable (distance between cholesterol headgroup and center of mass of residues within 7 Å of cholesterol when bound) were created using SMD simulations. These configurations ranged from lipid-bound to lipid-unbound state where cholesterol is present in the bulk lipid bilayer and the free energy profile plateaus. Cholesterol was pulled at a rate of 1Å/ns over a distance of 30 Å with a force constraint of 1000 kJ/mol/nm^2. The protein was restrained from positional and rotational movements by placing position restraints on SER237, ALA354, ALA883 and ALA1006 backbone beads on the XY 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.

Umbrella sampling simulations were run with conformations obtained from 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 1 Å spacing with larger collective variable values (distances over 10 Å) (see Figure C.1 in appendix C). 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 SMD simulations. The same positional restraints used in SMD simulations were also applied in the umbrella sampling simulations. Within each window temperature and pressure were

controlled using the Berendsen thermostat and Berendsen barostat.

The g_wham tool - GROMACS implementation of the Weighted Histogram Analysis Method (WHAM) was used to combine and unbias the umbrella potentials. Bootstrapping was used to estimate 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 (see Figure C.2 in appendix B).

5.3.2 Cholesterol binds tightly in the groove between TMH 10/12

Umbrella sampling simulations were performed to obtain the PMF of cholesterol binding in several potential binding sites. The free energy obtained from these simulations suggests cholesterol binds most tightly to the TMHs 10/12. In this case the free energy profile shows a broad energy minimum close to the initial binding site. Visual inspection of the umbrella sampling windows provides a potential explanation of the free energy profile. The groove formed by TMHs 10 and 12 is rather deep and cholesterol movement through the groove was reflected in the broad minimum of the free energy. During this stage cholesterol was still interacting with the transmembrane helices. The lowest energy state (-27.9 kJ/mol) was obtained once the cholesterol had almost left the groove rather than at its deepest point (-25.5 kJ/mol), however, the difference is not big - 2.4 kJ/mol (Figure 5.9 A). A secondary binding site (-5.4 kJ/mol) on the surface of P-gp next to the TMH 10 was explored during the umbrella sampling simulations. The lowest free energy from this site corresponds to a binding constant of 10 µM cholesterol binding - a result close to that suggested by experiments[105].

The binding sites on the surface of the protein achieved very similar scores - cholesterol bound with -6.4 kJ/mol to the site between TMHs 7 and 8 (Figure 5.9 B) and with -4.4 kJ/mol between TMHs 9/12 (Figure 5.9 C). These values are similar to that seen for a secondary binding site from the umbrella simulations of TMH 10/12 binding site and the lower energy values suggest weaker binding. Alternatively, the binding affinity could depend on the depth of the groove cholesterol is binding to. In case of superficial binding to the surface of P-gp the binding affinity would be lower than if cholesterol is positioned relatively deeply in a crevice or another type of surface indentation. Of course, that could also signify that P-gp conformational changes would have great effect on cholesterol binding.

5.4 Atomistic simulations of P-gp in complex bilayer

Multiscale approaches to molecular dynamics simulations allows one to investigate protein-lipid interactions in greater detail and circumvents the individual downsides of each technique. In this case, coarse grained simulations give access to a much longer timescale and allow convergence in umbrella sampling simulations to be achieved with lower computational cost. However, the finer details of the interactions are lost and the protein movements are constrained by the elastic network of bonds. This information can be recovered by conversion of coarse grained topologies to an atomistic resolution and further simulations. In this section, several interesting topologies observed after 10 µs CG simulations were chosen and converted to atomistic resolution for a further run of 100 ns at the more detailed resolution. Examples include the narrow conformation of P-gp with cholesterol bound in potential binding sites and wide conformation with and without lipid entrance to the substrate binding site. This

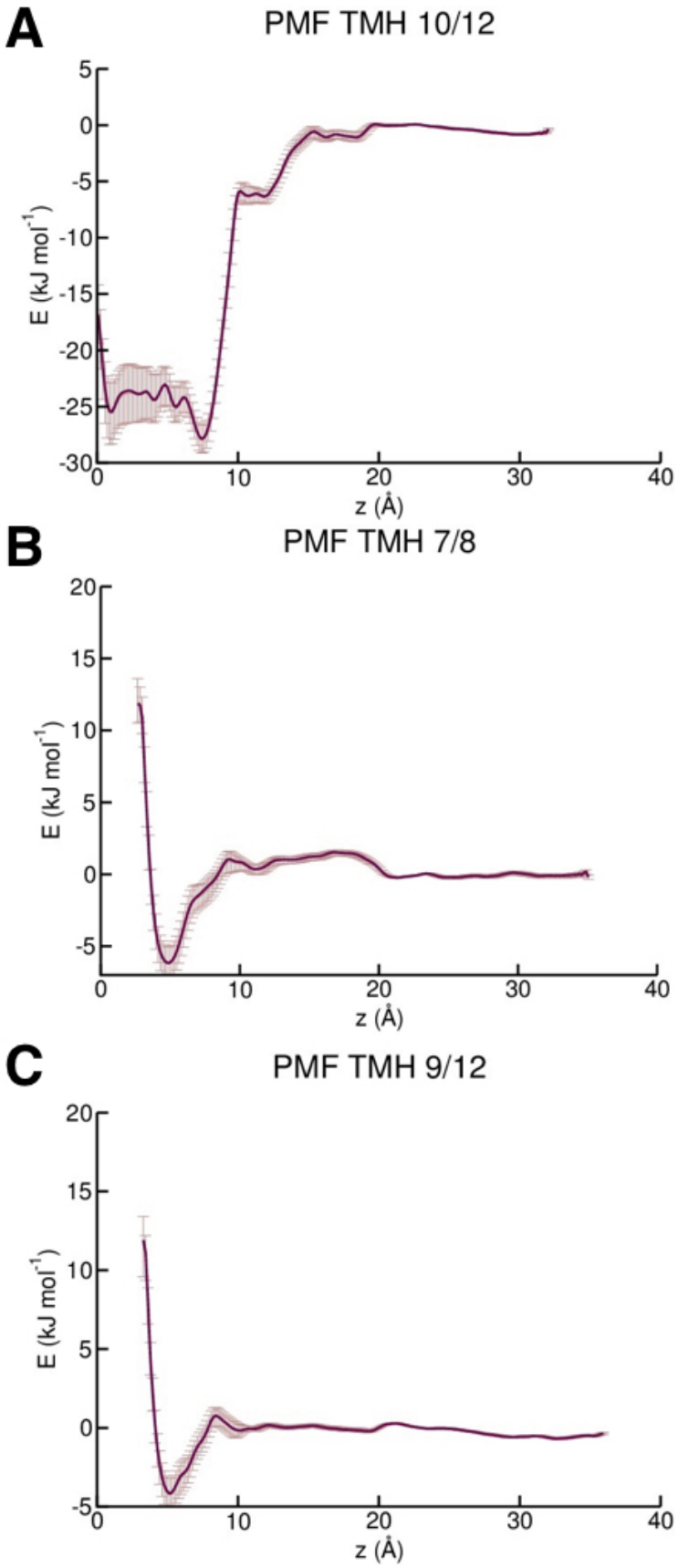


Figure 5.9: The PMF profiles of cholesterol binding sites at (A) TMH10/12, (B) TMH7/8 and (C) TMH9/12. The error bars in these graphs show error as calculated by bootstrapping.

selection was made to investigate the widest range of observed behaviour in the most efficient way. Three cases of each type were simulated to give an accurate estimate of the sampling.

5.4.1 Methods

The starting points of atomistic simulations were acquired from the last frames of selected CG simulations. Lipid molecules were then converted to atomistic resolution using the Backward tool from the Martini toolset [153] and the cg2at approach was used for the protein [152]. Alchembed was used to reduce potential steric clashes [182]. The systems are resolvated and 150 mM NaCl is added. After steepest descent energy minimisation, a 1 ns long equilibration simulation was run with heavy protein atoms restrained. After the equilibration, the production runs were 100 ns long.

5.4.2 Results

Analysis of lipid diffusion in the system with P-gp in a complex bilayer shows the same trends that were observed in coarse grained simulations of the same system - slowing of phospholipids in comparison to cholesterol and asymmetric diffusion (see Figure 5.10). The asymmetrical diffusion is more likely as the α value is well below 1 and the diffusion constants are lower than in similar coarse grained simulations, as could be expected from an atomistic resolution system.

The stability of the P-gp structure was assessed by measurements of Cα atom RMS deviations and fluctuations and overall results are shown in Figure 5.11. There was no significant difference between the different conformations considered in this section and RMSD was similar to that observed from simulations of apo and ATP-bound P-gp in POPC bilayer described in chapter 3. Similar conclusions could be reached

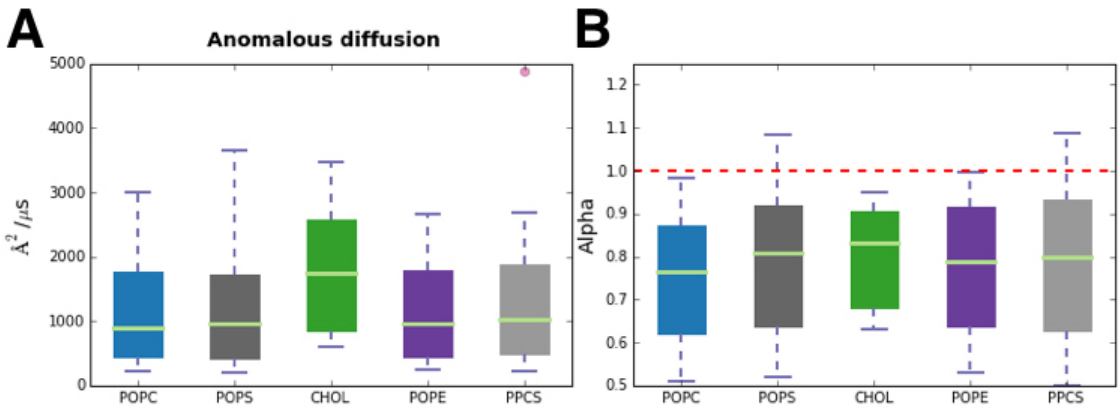


Figure 5.10: (A) Lipid diffusion in atomistic simulations of P-gp in complex bilayer shown as boxplots. Boxes show the interquartile range (IQR), error bars show the full range excluding outliers (points) defined as being more than 1.5IQR outside the box.(B) The results for MSD fitting to anomalous diffusion model suggest that subdiffusion was observed for all lipid species, as the α value from fitting is lower than 1[201]

about the RMS fluctuation - no significant movement was observed between the main P-gp domains. However, the "narrow" conformation had lower RMSD (Figure 5.11 A) than the "wide" conformation (Figure 5.11 C,E) which could be explained by reduced "breathing motion" exhibited by slight motions of the NBDs (see[175]).

Since lipid bilayer has been shown to be important for P-gp function and conformation[29], the behaviour of the putative substrate entrance gates was compared between simulations of model POPC bilayer and the asymmetric bilayer. Different conformations investigated in this section provided an additional layer of complexity and sampling in a more physiologically relevant setting.

The intramembrane entrance gates were of particular interest due to the closure of these gates in the apo simulations of P-gp in a POPC bilayer as discussed in chapter 3. The addition of other types of phospholipids and cholesterol was hypothesised to have an effect on the dynamics of P-gp movements. For both entrance gates, the distance between helices ranged from 8 - 12 Å during the simulations in the model POPC bilayer (blue lines in Figure 5.12 A,B). This profile is consistent with

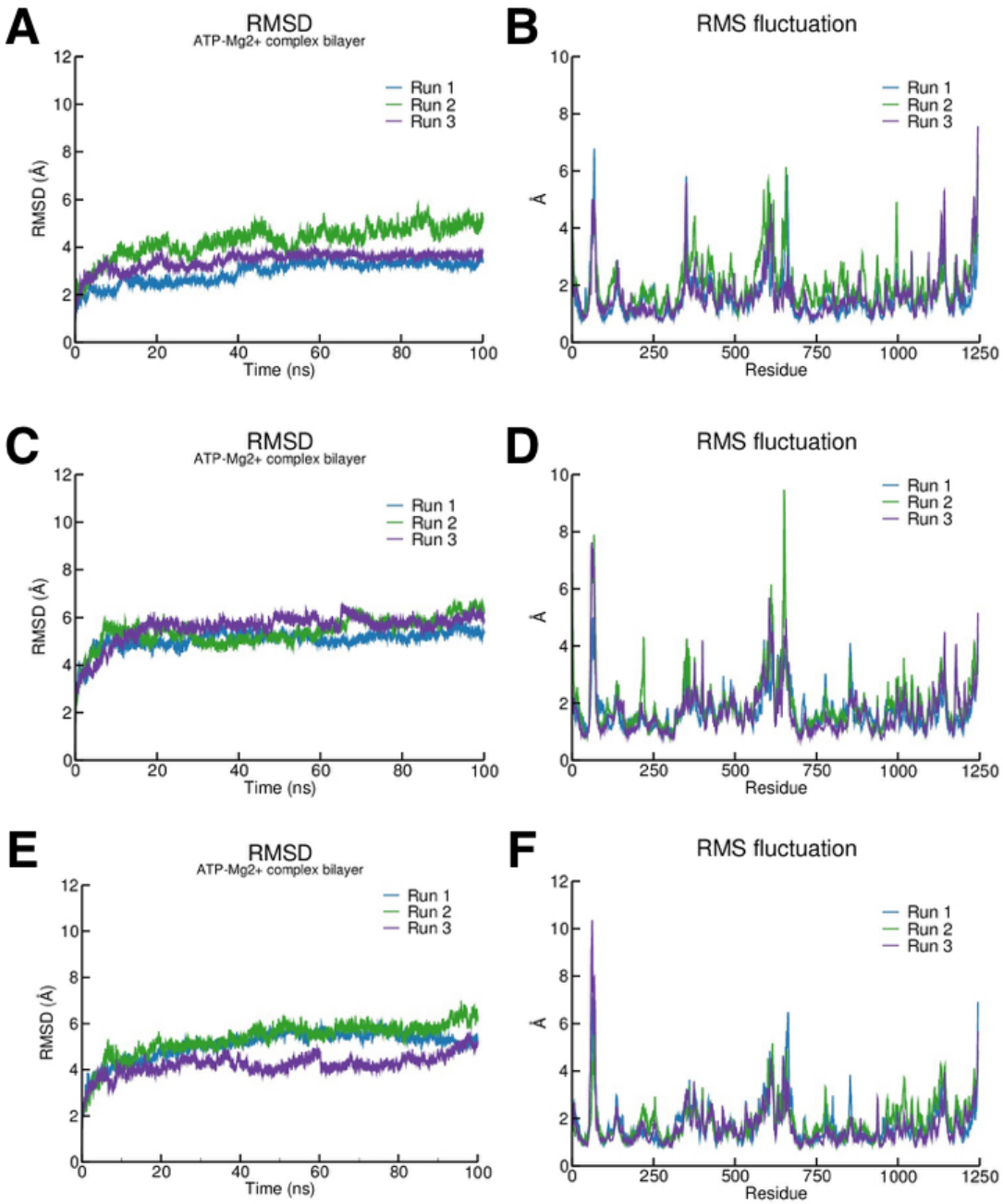


Figure 5.11: (A) RMSD of Cα atoms for the "narrow" conformation of P-gp. (B) RMSF of P-gp residues for "narrow" conformation. (C) RMSD of Cα atoms for the "wide" conformation of P-gp with no lipids present in the substrate entrance gates. (D) RMSF of P-gp residues for "wide" conformation with no lipids present in the substrate entrance gates. (E) RMSD of Cα atoms for the "wide" conformation of P-gp with lipids present in the substrate entrance gates. (F) RMSF of P-gp residues for "wide" conformation with lipids present in the substrate entrance gates. In each graph the first repeat is blue, second - green and third - purple.

closure of the gate as observed by visual inspection of the trajectory. The "narrow" conformation was taken from one of the simulations with closed entrance gates and the entrance gate-forming helix distance profile (dark orange lines in Figure 5.12 A,B) remains similar to P-gp in a POPC bilayer. This observation suggests that the mere presence of the complex lipid bilayer does not cause opening of either TMH4/6 or TMH10/12 intramembrane substrate entrance gates. In the case of the "wide" conformation the situation is more complicated. Lipid entrance through the gates causes them to remain open through the simulations, as evident in the wider distance profiles. This could be caused by parts of the lipid molecules residing in the entrance gates, thereby preventing them from closing. More interesting is the behaviour of the "wide" conformation when no lipids are present in the entrance gates. For TMH10/12 entrance gate (Figure 5.12 B) the distance profile was very similar to that of POPC bilayer, while the TMH4/6 entrance gate appeared to be more flexible in both closing and opening (Figure 5.12 A). In the case of the cytoplasmic entrance gates the distance profiles between POPC and the more complex bilayer were very similar (see Figure 5.12 C,D).

5.4.3 Lipid bilayer effect on P-gp

The atomistic simulations of P-gp in a complex asymmetric lipid bilayer allowed us to investigate the effect a more physiologically relevant environment has on stability and dynamics of the protein. This was further explored in more detail by looking at the conformational changes caused by lipid entrance through the TMH10/12 substrate entrance gate. While the overall stability and dynamics of the P-gp remained similar to that observed from simulations in a model single lipid (POPC) bilayer, the intramembrane portions of putative substrate entrance gates showed more flexibility. The cytoplasmic parts of these gates, however, maintained a similar profile to that

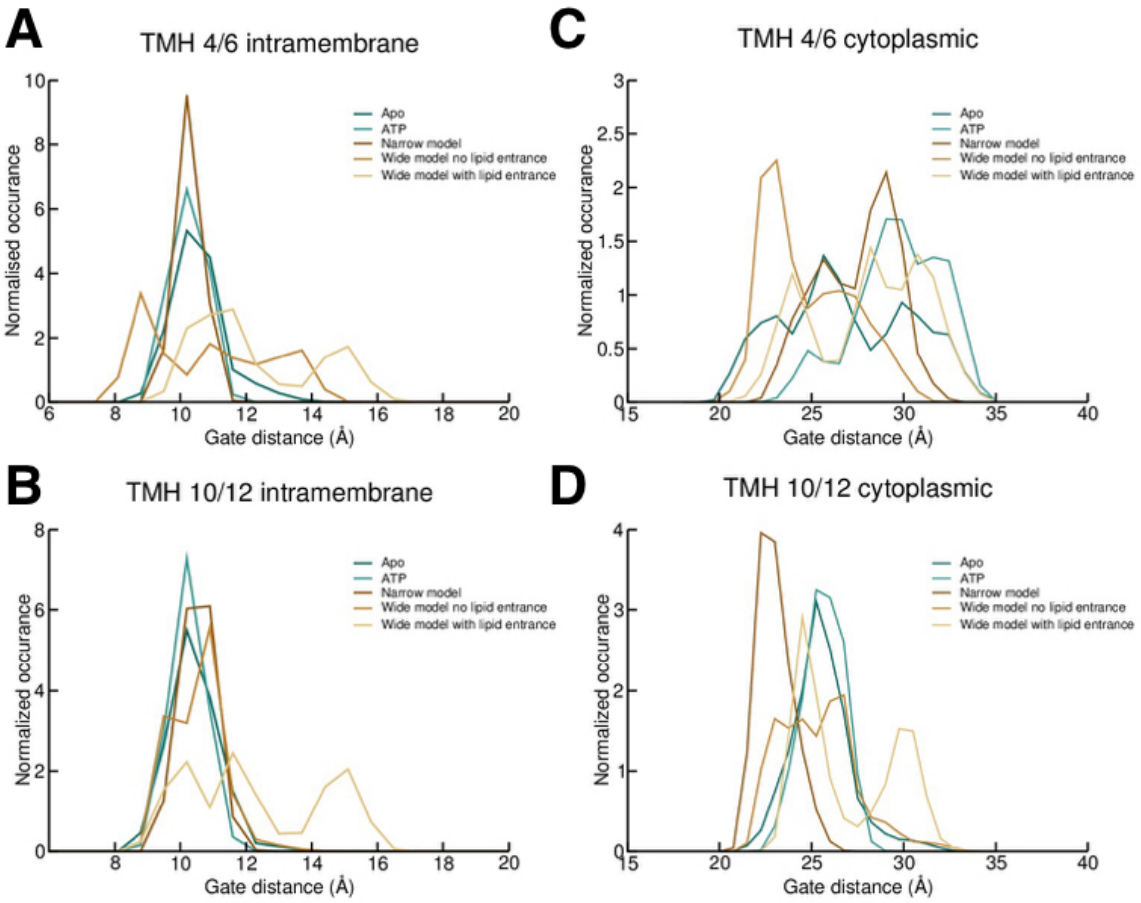


Figure 5.12: The distance profiles of intramembrane and cytoplasmic portions of potential substrate entrance gates. (A) Histograms of distances between COM of residues 221-233 (TMH4) and 341-354 (TMH6) which constitute the intramembrane part of TMH 4/6. (B) Histograms of distances between COM of residues 865-880 (TMH10) and 982-994 (TMH12) which constitute the intramembrane part of TMH 10/12. (C) Histograms of distances between COM of residues 234-259 (TMH4) and 355-370 (TMH6) which constitute the cytoplasmic part of TMH 4/6. (D) Histograms of distances between COM of residues 881-905 (TMH10) and 995-1014 (TMH12) which constitute the cytoplasmic part of TMH 10/12. In blue hues are the histograms for simulations of apo (darker blue) and ATP-bound (lighter blue) P-gp in POPC bilayer. In orange hues are the histograms for the simulations of "narrow" conformation (dark orange), "wide" conformation with no lipid entrance in the seed CG simulations (orange) and "wide" conformation with lipid entrance in the seed CG simulations (light orange).

observed in simulations with a model bilayer (see Figure 5.12 C,D). The intramembrane entrance gates remained closed in case of "narrow" conformation, however, the gates remained open to some degree (see Figure 5.12 A,B) if parts of lipid molecules had entered the central cavity through the intramembrane entrance gates. Interestingly, the TMH 4/6 intramembrane gate was relatively open even if no lipids had entered this or the other (TMH10/12) gate (see Figure 5.12 A,B). These observations hint to a potential function of lipids as stabilisers of P-gp structure, however, true to the concerns discussed in Section 4.5.1, experimental analysis of transmembrane helix behaviour in a bilayer is required to validate and support the MD data.

5.5 Conclusions

Since P-gp is a membrane protein, the surrounding hydrophobic environment has considerable effect on its structure and function[29]. The extent of this dependancy has been recently shown by two groups - Zoghbi *et al.*[53], as well as Loo and Clarke[52]. The first study voiced concerns over the physiological relevance of crystal structures acquired from an environment that consists of detergents[53], while the latter revealed a requirement for lipids to activate a switch at the transmission interface between TMD2 and NBD1[52]. 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[105].

In this chapter, we used a multiscale MD simulation approach to analyse a bilayer patch with a composition previously uncharacterised by MD. The composition of this bilayer was designed with the human blood-brain barrier in mind, taking into account the limitations presented by MD, namely, the availability of parameters for the lipid molecules in the desired forcefields. The main factors for analysis were

lipid diffusion constants and lipid distribution between leaflets. For this bilayer, the lipid diffusion constants appeared to be consistant with data available for lipid bilayers with different compositions in both CG and atomistic resolutions. After some time required for equilibration, the cholesterol distribution changed from an initially symmetric distribution towards a preference for the upper leaflet.

The behaviour of the lipid bilayer agreed with other available data and it was decided to proceed with P-gp insertion into the bilayer patch. Again, a multiscale approach was used. Coarse grained simulations provided information on the effects P-gp has on the surrounding lipid bilayer, while atomistic MD simulations were used to investigate the conformational changes of P-gp in a more physiological lipid bilayer and to contrast these results to a simple POPC bilayer (chapter 3).

Lipid ordering in ring-like structures was caused by the insertion of P-gp and this effect was most profound for the negatively charged POPS which formed an annular ring around P-gp in the lower leaflet, most likely attracted by the increased amount of positive residues in the headgroup region of the lower leaflet. A more specific type of binding was observed for cholesterol molecules which appeared to bind strongly to the crevices and cavities of P-gp, in particular, the cavities formed by TMH4/6 and TMH10/12 which form the putative substrate entrance gates.

Three of the sites were investigated by umbrella sampling simulations to obtain the potential of mean force. The binding site at TMH10/12 was the strongest, potentially because the cavity was deeper than the other sites in the umbrella sampling study. All relatively shallow binding sites had similar free energy, which could suggest less specific or non-specific binding on the surface of P-gp.

Another interesting lipid behaviour that was observed in the coarse grained simulations was the full or partial entrance of lipids in the central cavity through a gap between TMH10/12. Such behaviour agrees with P-gp function as low-activity phos-

pholipid flippase, however, it also suggests that the behaviour of substrate entrance gates might be lipid dependant. This was further explored in the atomistic resolution simulations which can reveal more detailed insight in P-gp behaviour in a complex lipid bilayer.

Although the analysis of P-gp dynamics did not show any difference from the simulations in POPC bilayers, the substrate entrance gates showed more flexibility, suggesting that lipids can cause widening of the substrate entrance gates. However, this effect was not observed in all simulations and in cases where a widely open conformation did not have bound lipids in these sites, the intramembrane substrate entrance gates closed similarly as in simulations with pure POPC bilayer. Therefore, no definite conclusions can be made regarding the intramembrane gate behaviour and either longer simulations or experimental data is needed to validate these hypotheses.

In conclusion, although insertion in a more complex bilayer did not create noticeable differences in P-gp conformation at atomistic resolution, it does create a more physiologically relevant environment and the effects might be subtle. Alternatively, they might be revealed by simulations with longer timescales. However, the benefits are more noticeable in a coarse grained resolution where the lipid interactions with P-gp can be observed on a long timescale.

A clear definition of potential cholesterol binding sites was achieved in the coarse grained simulations with the binding site formed by TMHs 10/12 showing the highest probability to be a high affinity cholesterol binding site. These results can be further tested with techniques such as liquid-state NMR which could show if the cholesterol actually binds to the sites proposed in this chapter. The detailed description of residues involved in each binding site also provides information for future mutagenesis studies. Therefore, simulations in this chapter could serve as an example of MD-lead hypothesis and discovery.

5.5.1 Further experiments

The cholesterol binding sites described in this chapter have not been previously characterised, since no specific cholesterol binding sites for P-gp have been experimentally determined, although some suggestions based on sequence have been made[105]. With the data available from MD simulations, mutation studies and NMR of cholesterol analogues can be guided to reveal the P-gp-specific cholesterol binding sites.

6

LINTools - dynamic representation of protein-ligand interactions

6.1 Introduction

It has been recognised that proteins undergo dynamic interactions with their ligands due to structural changes and allostery. This process is of fundamental biological interest and understanding of protein-ligand interactions is essential for structure-based drug design[217]. Lessons from the drug discovery process state that drug efficacy

is affected not only by binding affinity but also by the kinetics of the protein-ligand complex, particularly, the length of time a ligand spends in the binding site[218;219].

Molecular dynamics is a computational method that has rapidly become an important tool in drug discovery efforts (for a review see[220]). While X-ray crystallography has had great success as a major source of protein and protein-ligand complex structural information, these are but snapshots of a dynamic process. The structural data acquired from this and other methods can be utilised by MD to provide insight to alternative conformational states and the transitions between them.

Recent technological development has made it possible for MD to observe events in the microsecond range, making it more relevant and closer to the experimental timescales. The spatial dimension is also expanded, with routine simulations involving membrane proteins embedded in their native environment which is also becoming more complex and physiologically relevant (e.g. simulations in chapter 5 of this thesis). The advances in the field and exponential growth of available structural information has led to significant increase in reported MD simulations of protein-ligand complexes (see review[220]).

Molecular docking is often a method of choice for virtual screening of large libraries and determining the binding site of a drug molecule. It is a fast and computationally inexpensive method, however, the binding free energy estimate obtained from this process is not very accurate[221] and determination of the correct binding pose can be a non-trivial task[222]. MD is often used to analyse binding site interactions after docking, as this provides a temporal resolution to the static state. The information that can be obtained from MD simulation includes a measure on the stability of the complex and distances between ligand atoms and residues of interest.

A lot of effort has been put into visualisation of the ligand binding sites to provide an easy to understand representation of the binding site, its properties and protein-

ligand interactions. Schematic 2D diagrams are widely used to visualise interactions with ligands and a plethora of software packages are available for this purpose. LIGPLOT[223] and its successor LigPlot+[224] are commonly used by the academic community, while PoseView[225] has been implemented in the PDB webserver. Modules of commercial software, for example, MOE (Molecular Operating Environment)[226] and Maestro (Schrödinger Inc., www.schrodinger.com) also generate schematic diagrams of protein-ligand interactions. Newer additions to the field include LeView[227].

These programs provide a comprehensive view on the different types of protein-ligand interactions such as hydrogen bonds, aromatic bonds, salt bridges and interactions with waters in the context of the binding site. The input is typically either a crystal structure with ligand bound or a protein-ligand complex acquired by docking. However, all of these programs are made to show only the static state of the complex and do not accept any information of the dynamics of the interactions. While this information cannot be acquired from docking or crystal structures of protein-ligand complexes, it is provided by MD simulations.

The recent expansion in the use of the Python language[228] in biochemistry has resulted in the creation of libraries that can analyse MD trajectory data. Most common examples are MDAnalysis[207] and MDTraj[229]. These libraries allow the import of MD trajectory data through the Python environment, opening a wide range of opportunities, for example, analysing data from any simulation package with the same script - a feat that is normally impossible since each simulation package has its own analysis scripts and accepts only certain types of input files. Therefore it is possible to create universal analysis tools and due to the open source nature of the libraries, disseminate them easily through academic and commercial communities.

In this chapter I introduce Ligand Interaction Network Tools or LINTools - time-resolved graphical representation of ligand-protein interactions. This open-source

Python script analyses topologies and trajectories of proteins and ligands and produces 2D representation of 3D interaction sites with an added time dimension. The dynamic representation of protein-ligand interactions using data from trajectories sets LINTools apart from other tools looking at the protein-ligand interactions.

6.2 LINTools workflow

The workflow of LINTools process consists of four main steps. First, the topology and trajectory files are parsed and analysed with MDAnalysis[207] in a Python environment. Second, the information obtained from the first step is plotted for protein residues of interest. Third, RDKit[230] and Shapely[231] is used to obtain a 2D drawing of the molecule and the 2D placement of protein residues. Finally, an SVG (Scalable Vector Graphic) image is compiled and saved.

Table 6.1: LINTools options and flags

Flag	Description
-t, –topology	Topology file
-x, –trajectory	List of trajectory files
-o, –outname	Output name
-rmsf	If selected, RMSF analysis on ligand molecule will be performed.
-c, –cutoff	Cutoff distance (in angstroms) for residence time analysis, default is 3.5 Å.
-ro, –residueoffset	Number of offset residues for the protein, default is 0.
-ac, –analysis_cutoff	The percentage of simulation time that a certain feature (e.g. hydrogen bond) has to exceed in order to be shown in the final image. Used in hydrogen bonding and residence time analysis.
-df, –domain_file	Input file for user-defined protein domains.
–no_hbonds	Hydrogen bonding will not be analysed.
–debug	Functions for debugging
–no_resinfo	The residue information files will not be produced.
-mol2	MOL2 file

6.2.1 Data input and analysis

The input for LINTools is a topology file (-t or –topology flag) with the associated trajectory file or multiple files (-x or -trajectory flag) in any format supported by MDAnalysis (for this and other options see Table 6.1). The trajectory files are optional and it is possible to obtain the 2D representation of a binding site from a topology file alone. This option can be useful for viewing docking results. An optional flag (-ro or –residueoffset) allows one to control the renumbering of protein residues in the case of mismatch between the topology and an accepted numbering of residues, for example, after homology modelling . The input files are parsed into and stored as a MDAnalysis Universe, which allows the Python interface to access information such as residue and atom names and coordinates.

6.2.1.1 Obtaining coordinates of the ligand

As part of the input, the user is asked to select a ligand from a list for further analysis. The list is compiled by residue names that are not part of protein or solvent. The remaining residues are shown by their name, number and segment. The selected residue coordinates are extracted from the MDAnalysis Universe and saved as a PDB file. However, a PDB file does not contain all the information that is required for successful 2D drawing of a molecule, crucially, the information about bonds is missing. To circumvent this problem, the PDB file with ligand coordinates is converted into MOL2 type file that contains informations about the bonds and formal charges of the ligand molecule (script written by Dr Teresa Paramo Prieto).

6.2.1.2 Determine protein residues of interest

Only the residues that have been determined to be of interest will be plotted in the final figure. This selection is defined by the cutoff distance from the ligand and residues within it for a certain amount of time are considered to interact with the ligand. It is set up with a -c or –cutoff flag and by default is set to 3.5 Å. Thefore, the selection is straightforward in case no trajectory files have been submitted - any residue that has any atom within the cutoff distance of any ligand atom will be plotted as part of the binding site. Once the residues are determined, distance measurements by MDAnalysis are used to determine the closest ligand atom for each residue. This information will be used to determine the 2D positioning of that residue. Only heavy ligand atoms are considered in this analysis as hydrogen atoms are not present in the 2D depiction of the ligand.

The matters become more complicated once trajectory data is analysed. An additional time dimension is added to the analysis and must be considered for best representation. Each frame of a trajectory is analysed to determine the residues within cutoff distance and each residue receives a score that consists of the number of frames this residue was within cutoff distance. The analysis cutoff (-ac or –analysis_cutoff) flag is used to determine the percentage of simulation time a residue must spend within the cutoff distance. For example, the analysis cutoff with the default value of 30 would allow to select residues that have spent at least 30 out of 100 ns within the cutoff distance. This allows one to dismiss residues that either do not interact with the ligand or appear only transiently (e.g. sudden flip of a side chain), resulting in reduced noise for analysis and the final image. An additional layer of selection is added to cases where more than two trajectories are submitted for analysis - the residues must fulfill a “best-of-three” or “best-of-five” requirement as well. For example, this means that a residue will be plotted only if it interacts with the ligand at least two

out of three simulations. The residues determined from this process are then assigned a heavy ligand atom that is closest to them in the topology file submitted together with the trajectories.

6.2.1.3 Determine hydrogen bonding

The residues that have been determined to be of interest in the previous steps of the analysis are further investigated for hydrogen bonding with the ligand. Since MDAnalysis cannot determine donors and acceptors of non-standard molecules (i.e. ligands), RDKit is used to find these ligand properties. Once all donors and acceptors are found, MDAnalysis HydrogenBondAnalysis class is used to analyse hydrogen bonds between protein residues of interest and ligand. A list consisting of residue - ligand atom pairs is compiled and stored for compilation in the final figure. Similarly to previous analysis, the frequency of hydrogen bonds in a simulation must pass analysis cutoff in order to be shown in the final figure. A distinction between interactions with backbone or sidechain atoms is made and reflected by line style in the final figure.

6.2.1.4 RMS fluctuations of the ligand

It is possible to check the root-mean-square fluctuation (RMSF) of the ligand within its binding site from a trajectory. RMSF is a measure of atomistic movement from a reference and therefore can give an indication which part of the molecule is most flexible during the MD run. This analysis is done using MDAnalysis RMSF class after root-mean-square fitting of the trajectory to a reference frame - in LINTools it is set up to be the file submitted under topology flag. The RMSF is averaged in case of several trajectories and each atom is assigned a numerical RMSF value which is saved in text file and plotted in the resulting 2D diagram (see Figure 6.1).

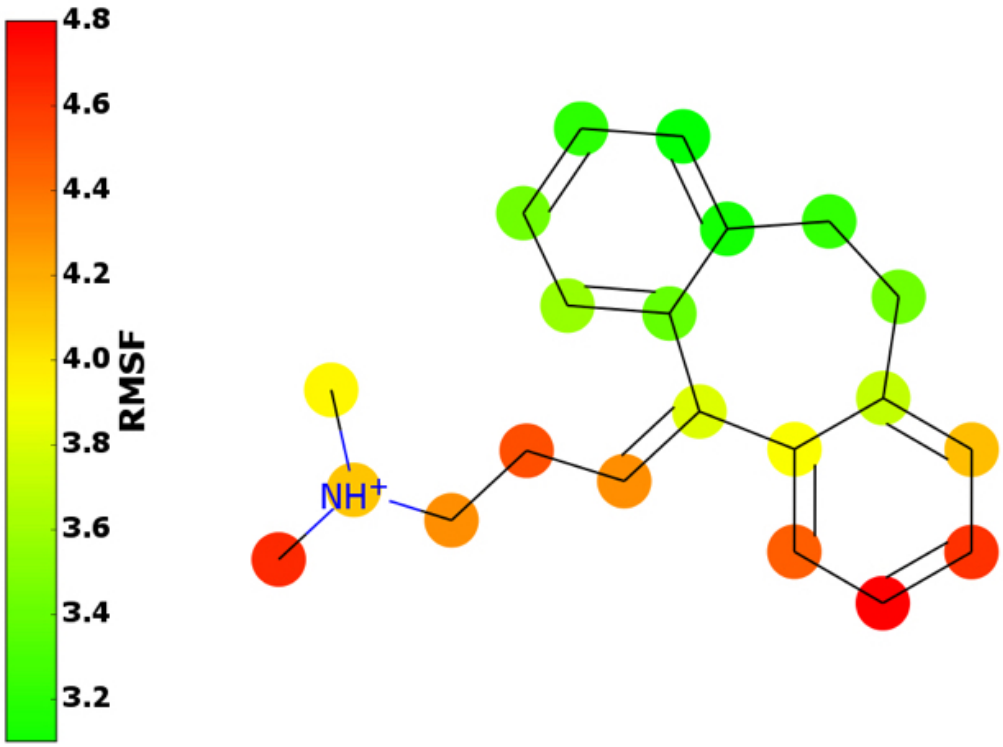


Figure 6.1: The RMSF is plotted on the ligand diagram for every atom. The results are color coded from green (low RMSF) to red (high RMSF) as seen on the colorbar on the left.

6.2.2 Plotting the residues of interest

Selection of plotting style is a part of initial user input. The input consists from selection between the available plotting styles that depend from other input parameters (see Table 6.2). The "amino" style of plots is always available, showing the type of residues based on the hydrophobicity and polarity of the amino acid. The "domain" plotting style allows greater flexibility in residue description by using user defined parameters for residue properties. The "clocks" plots are available if a trajectory is also submitted, since these show the fraction of simulation time spent interacting with the ligand for each trajectory, per residue. All of these styles are plotted using Matplotlib package and saved as separate SVG files.

6.2.3 2D representation of protein-ligand interactions

Since the input of LINTools is a 3D coordinate system, while the output is a 2D depiction of the time-resolved interactions, the non-trivial problem of dimensionality reduction must be solved (see Figure 6.2). In case of the ligand, this is achieved by using the 2D drawing utility of RDKit, however, this results in loss of dimension and distortion of some of the features present in 3D. The 2D coordinates of the ligand are obtained from the resulting drawing and the relationship between ligand atoms and the distance to the residues is used to depict the 2D model of protein-ligand interaction site.

6.2.3.1 Drawing of 2D ligand

As mentioned previously, the dimensionality reduction is a non-trivial problem, none more so in the case of 3D molecules. While it is possible to use PCA analysis to reduce a dimension for a set of 3D coordinates, the result will often be challenging to use

Table 6.2: The styles of residue plots

Style	Example	Description
"Amino"		This style is always available and displays the charge properties of the plotted residue with the acidic residues (ASP, GLU) in orange, polar residues (SER, THR, ASN, GLN, CYS, HIS) in pink, hydrophobic residues (ALA, VAL, ILE, LEU, MET, GLY, PRO) in dark green, aromatic residues (PHE, TYR, THR) in green and basic residues (ARG, LYS) in blue.
"Domain"		This style is available if a domain file (described in Section 6.) is submitted. The color and type of the domains is user defined. Two types of the domain plots are available - solid line (left) and dotted line (right). Residues that do not belong to any domains defined by the user are automatically shown in grey.
"Clocks"		This style is available if at least one trajectory file has been submitted. The results for each trajectory are shown with a separate color line in this case dark green for one trajectory and green for another, allowing one to analyse two or more repeats at the same time.

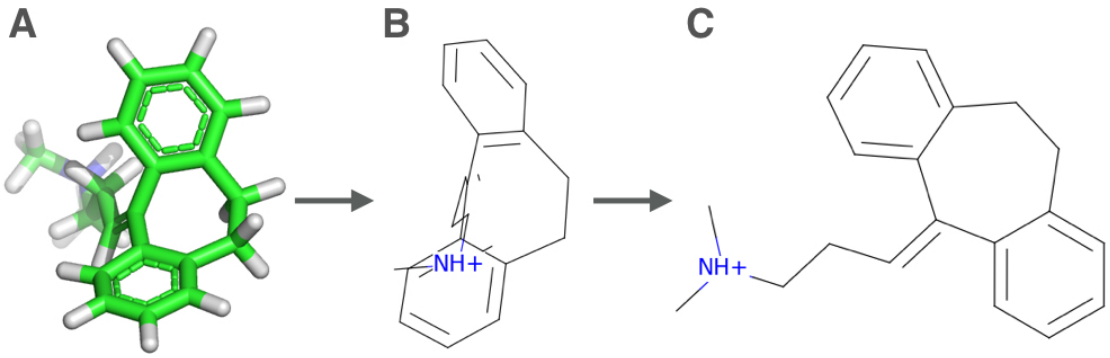


Figure 6.2: (A) Three dimensional representation of amitriptyline converted directly to (B) 2D figure show distortions caused by loss of dimension. (C) The "clean" 2D representation of amitriptyline from 2D coordinates calculated by RDKit from a SMILES string and not from the 3D coordinates. This achieves a clear representation, however, it has lost any connection to the 3D coordinates of the topology file.

in a representation and plainly unrecognisable (see Figure 6.2B). The RDKit library has a module that produces rendered 2D representations of molecules. In order to get the best representation, a SMILES string is used as an input to the RDKit. This string is obtained from the MOL2 type file created earlier. The resulting 2D drawing of the ligand is saved as SVG file (see Figure 6.2C). Since SVG files consist of textual description of how the image should look, this property can be used to parse through the lines of the file and get the 2D coordinates of each ligand atom.

6.2.3.2 Determining residue placement in 2D

The 2D representation of the ligand produced from the SMILES string can be very different from the 2D representation created by the dimensionality reduction process and as such it is impossible to directly relate dimensionally reduced coordinates of protein residues. Therefore, the relationship of distances between residues and ligand atoms is used to determine approximate placement of protein residues in regard to the ligand atoms.

A convex hull is drawn around the 2D coordinates of ligand atoms and, to provide some buffering space around the ligand, it is expanded by the diameter of a residue plot. This forms the initial shell around the ligand and ensures that residue plots do not overlap ligand atoms. One by one, plotted residues are placed on convex hulls around the ligand. First by determining the buffer of the convex hull from the distance between residue and ligand, then the point defined by 2D coordinates of the ligand atom closest to the residue is projected onto the expanded convex hull. The projected point is then interpolated from the convex hull, giving the initial set of 2D coordinates for that residue. A graphical summary of this process is shown in Figure 6.3. The geometrical problems are handled by yet another dependency - the Shapely geometry library.

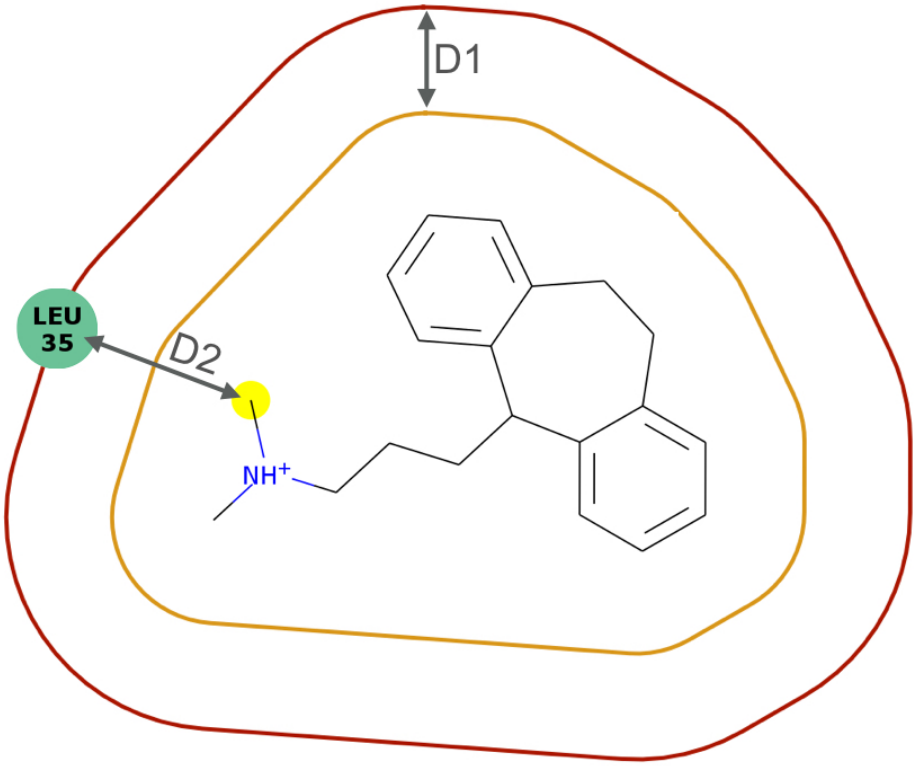


Figure 6.3: The method of residue placement starts by drawing a convex hull around the 2D representation of the ligand molecule and expanding it by 120 points (orange line). Then another convex hull (red line) is drawn for each residue (in this figure LEU 35) by expanding the initial convex hull (orange line) by a scaled value of distance (line labelled D1) between the closest ligand atom and the residue in the 3D space. The ligand atom closest to LEU 35 is shown on the ligand diagram as a yellow dot. Another point (green LEU 35 plot) is projected on the new expanded convex hull (red line) by finding the closest distance to this convex hull (line labelled D2). After interpolation the 2D coordinates of this point are found and these will be used for further assessment of overlap with other residues or the final figure.

Problems with this approach arise in case of two or more residues being close to the same ligand atom or a part of a residue plot overlaps another. It is solved by moving the coordinates along the convex hull until no overlap is detected.

6.2.4 Compilation of final image

To compile the final SVG image, the canvas of the initial drawing of 2D molecule is extended to fit all other components - residue plots, legends, etc. The SVG format is easy to work with in Python since it is a text document that can be modified to create the required images. Therefore, after centering of the 2D molecule, the residue plots are placed at the appropriate coordinates to represent the interaction site. This is achieved by modifying and concatenating the SVG plot files with the modified canvas of 2D ligand. Additional elements, such as legends, are also created as appropriate and inserted in the SVG file. Once the final image (e.g. Figure 6.4) is saved, the intermediate files are deleted.

Time coded information of residue interactions with the ligand are shown in "residue-information" files (see Figure 6.5) which are created individually for each residue.

6.3 Dependencies

LINTools require several dependencies that analyse the MD and chemical data, as well as help with drawing of the images. The input and analysis of MD data is done by MDAnalysis[207]. This and similar libraries can be used to analyse simulation data in a Python environment, thereby simplifying the analysis process and making universal analysis tools possible. MDAnalysis is used in LINTools to analyse the residence time, hydrogen bonding and RMSF of the ligand. Some chemical analysis that has not been implemented in MDAnalysis, is done by RDKit, a popular chemoinformatics

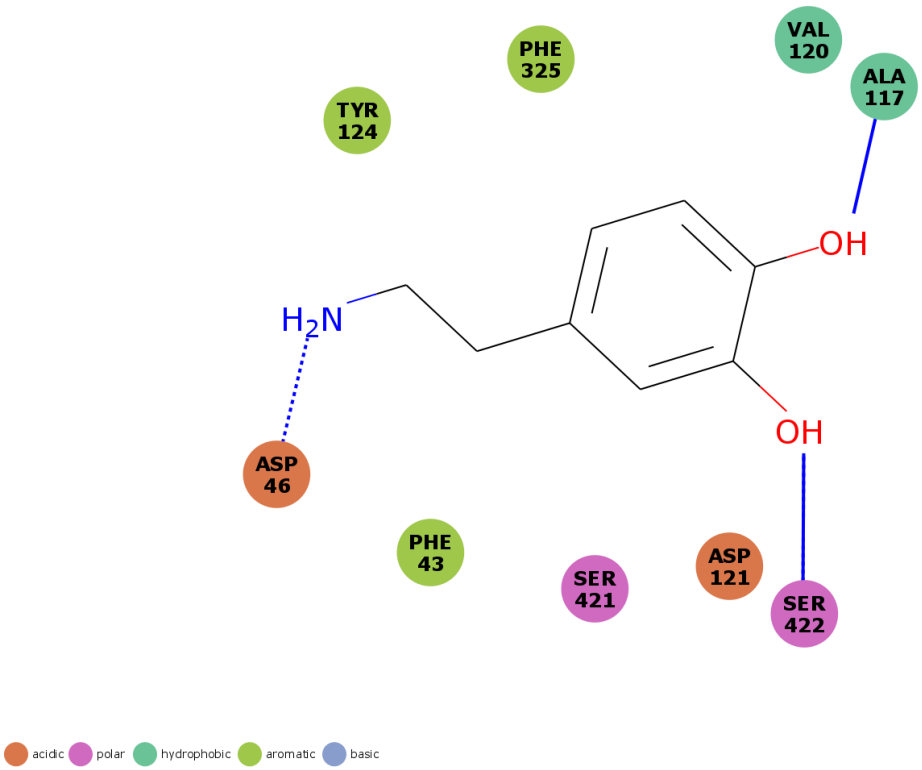


Figure 6.4: An example of "amino" style plot final image of LDP residue in PDB 4XP1 [232]. In this example we can also see the different types of hydrogen bonds where solid blue lines represent hydrogen bonds that are formed between ligand and backbone atoms of that residue, while the dotted blue lines show hydrogen bonds with sidechain atoms.

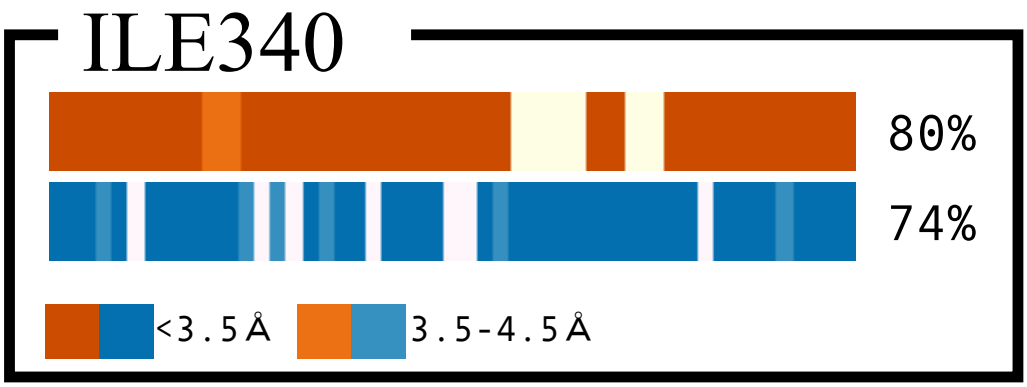


Figure 6.5: Time coded information of residue interactions (distance from the ligand) for each simulation for a single residue - in this example ILE340. Each horizontal bar represents a single simulation.

library. For example, it finds the hydrogen bond donors and acceptors in ligand molecules. Additionally, RDKit produces the 2D drawing of the ligand molecule. The geometry required for residue placement around the ligand molecule (e.g. convex hulls, projection of points on the convex hulls and interpolation of resulting points) is calculated with the Shapely geometry library. All of these dependencies are open source and freely available through package managers such as conda.

6.4 Open-source and licensing

LINTools is licensed under GNU General Public License version 3 which allows anyone to share and change all versions of the program. For the user this means they can use the software as if it was public domain - use it, change it and copy, as well as give away or sell the software as long as they follow the terms of the license. This license was selected to ensure the freedom of use of LINTools. The software also has a logo that will help to establish recognisability (Figure 6.6).

6.5 Good software development practices

In order to create software that is easily accessible, well-documented and tested several good development practices are in place. LINTools is an open-access software and as such is available from a GitHub repository (https://github.com/ldomic/lintools). A GitHub repository holds the current code and the history of development, as well as tracks user-submitted issues and wiki, making it easier to use and understand the code. The version control tool git allows quick deployment of the code, providing an ease of installation for the users.

LINTools is rigorously tested - Travis Continuous Integration tool[233] is integrated with



Figure 6.6: LINTools logo designed by Rihards Sobols.

the LINTools GitHub repository so that every time the code is updated, it is tested through Travis CI. The coverage of the testing is calculated by Coveralls.io tool[234] which is integrated with the Travis CI account for LINTools. At the time of writing (July 2016) the coverage for LINTools is 84%, meaning that most of the code has been invoked at least once during the testing. The testsuite that has been prepared looks at the end result from a particularly set-up run of LINTools and compares it with stored version. Therefore, if the new code causes unwanted or unexpected changes, those can be caught automatically without having to remember to occasionally run the tests.

The development and in-house testing is done in a Docker[235] container - a virtual machine that runs Linux operating system with very few dependencies installed, allowing strict control of the environment. It can also potentially serve as a distribution tool, since LINTools cannot work on Windows OS due to the unavailability of its dependencies, however, a virtual machine with Linux OS can solve this problem without causing potential clashes with other software libraries on the machine.

6.6 P-glycoprotein and amitriptyline binding

Section 4.2 describes simulations of amitriptyline in the P-glycoprotein central cavity after free diffusion towards the core of the protein and after consensus docking. These simulations were analysed with LINTools to gain more information about the amitriptyline binding and the protein-ligand interactions. In this section the results of this analysis will be described in more detail to showcase the functionality of LINTools.

6.6.1 Free diffusion of amitriptyline in the central cavity of P-gp

As further discussed in Section 4.2.1, amitriptyline freely diffuses towards the core of the P-gp central cavity. The amitriptyline molecule moved upwards for approximately 25 ns and after that remained largely at the same spot (Figure 6.7A). The 25 ns at the beginning of the simulation were discarded for LINTools analysis since during this time the molecule was moving upwards and this would not represent a stable binding site.

The residence time analysis of this site shows that the interactions are relatively stable since most of the residues within close vicinity of amitriptyline show long residence time (Figure 6.7B). Most of the residues with lower residence time are further away from the amitriptyline molecule. Glutamine 990, a residue previously been suggested to be critical for substrate and inhibitor binding[32;55], forms a hydrogen bond with the protonated nitrogen atom (Figure 6.7B). A closer look at the helices involved in this site reveal that it is formed by TMHs 4, 5, 6 and 12 (Figure 6.7C). 11 out of 15 residues shown on this plot have been suggested to form the rhodamine binding site[55;186;187].

6.6.2 MD simulations of docked amitriptyline

To further investigate amitriptyline binding we employed consensus docking and consequent molecular dynamics simulations. This work has been described in Section 4.2.3. The final docked pose of amitriptyline was positioned deeper in the central cavity than the final pose after free diffusion simulation (see Figure 6.8A). Therefore, LINTools was again employed to understand the protein-ligand interactions in the potential binding site revealed by consensus docking, with a hope that it could

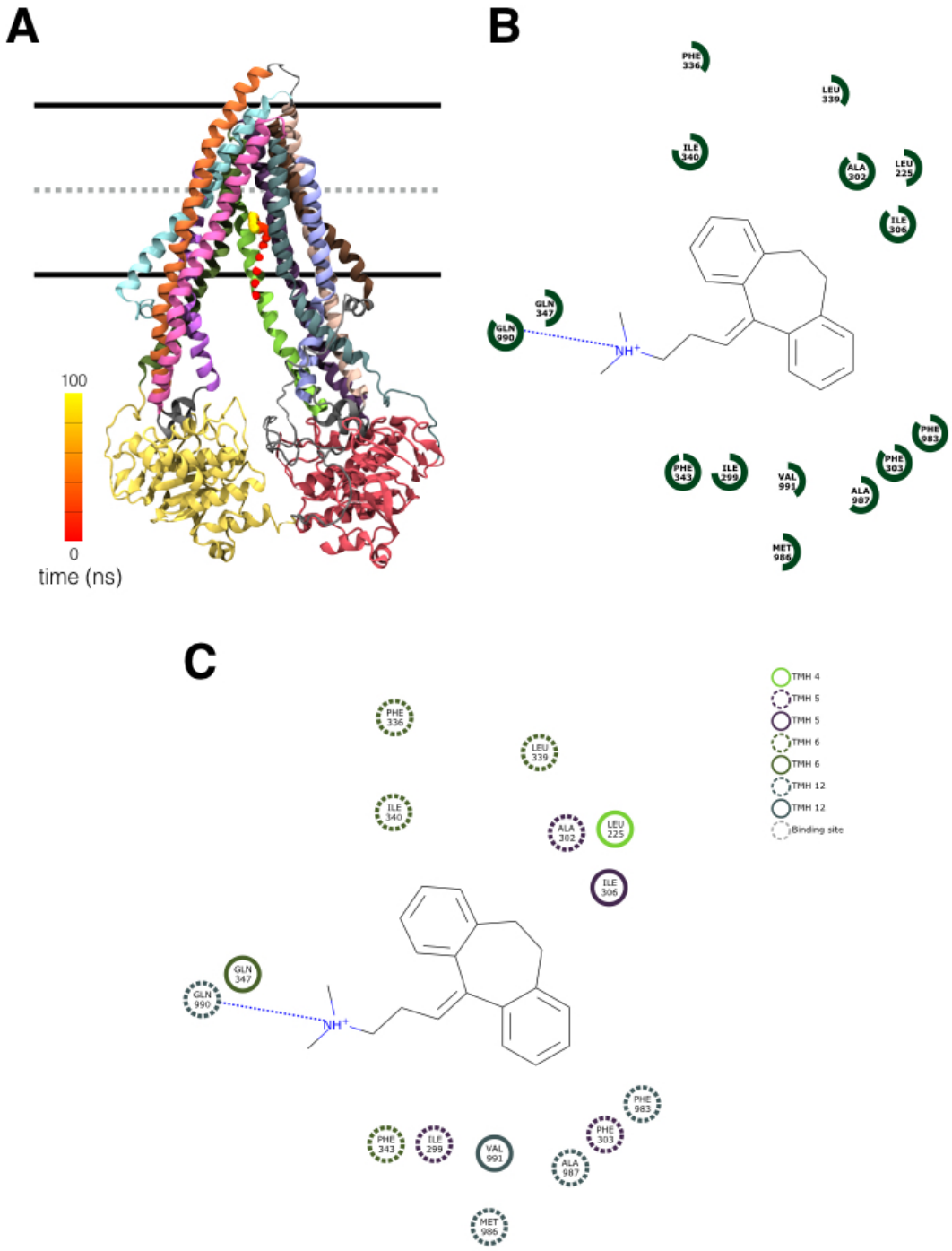


Figure 6.7: (A) The path of amitriptyline during the free diffusion through P-gp central cavity is shown with color coded (red to yellow) spheres. (B) "Clock" representation of amitriptyline interactions with P-gp during the last 75 ns of the 100 ns simulation. (C) This representation shows the identity of the transmembrane helix of each plotted residue. The color scheme is the same as used in panel A. The dotted circles show residues that have been shown to be a part of the rhodamine binding site[55;186;187].

suggest which site is more likely to be the actual binding site.

The analysis of triplicate repeats showcases LINTools ability to compare similar simulations with a simple diagram (Figure 6.8B). It is immediately apparent that although the starting point for all repeats is the same, the protein-ligand interactions are showing different profiles for each repeat. Some residues in this site seemed to be essential (PHE 336, ILE 340, PHE 983, PHE 728, MET69). Interestingly, 4 out of 5 of these residues are implicated to be involved in rhodamine binding to P-gp[55;186;187]. However, when compared to the site after free diffusion, the amount of binding site forming residues is much lower - only 5 out of 15 plotted residues have been suggested to form the rhodamine binding site (Figure 6.8C)[55;186;187].

6.6.3 Comparison of amitriptyline binding sites from LINTools analysis

Docking has often been combined with molecular dynamics simulations. Recently, a strategy for determining the correct binding pose from docking by utilizing short metadynamics simulations has been developed[222]. This study determined that the correct pose is highly likely to have more hydrogen bonds and lower RMSD than the other potential poses[222].

Similar conclusions can be drawn from LINTools analysis on the potential amitriptyline binding sites. The site observed after free diffusion in the central cavity of P-gp forms stable interactions. Most of these residues have been shown to be involved in rhodamine binding and amitriptyline is forming a hydrogen bond with a residue implied to be critically important in substrate binding. On the other hand, the pose uncovered by consensus docking, albeit stable in the 100 ns simulations, does not form so many critical interactions and shows no hydrogen bonding interactions. The

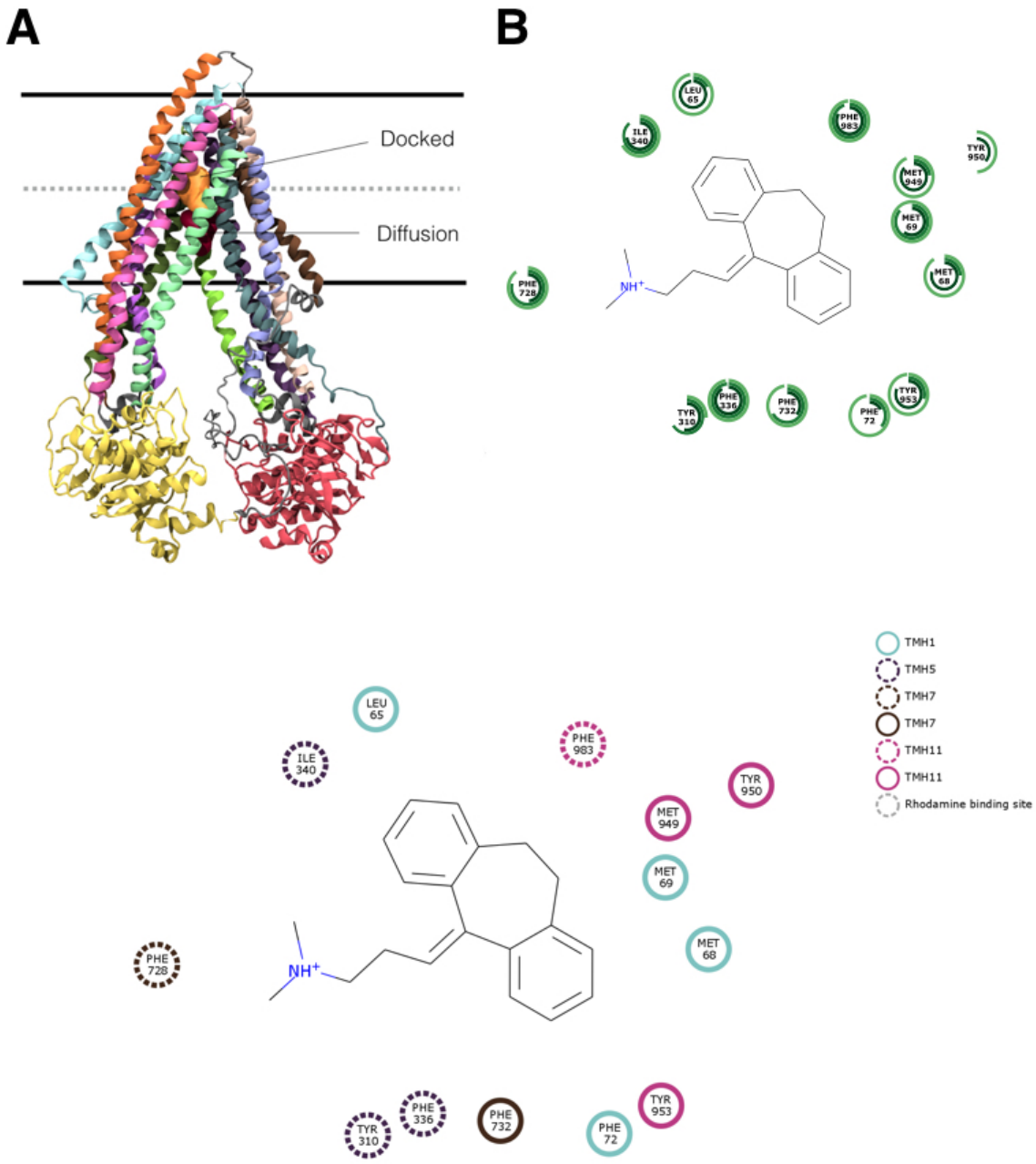


Figure 6.8: (A) The positions of amitriptyline after docking (orange) and free diffusion in the central cavity (red). (B) "Clock" representation of amitriptyline interactions with P-gp in three 100 ns repeats. (C) Transmembrane helix identity shown as "domain" representation, the color scheme is the same as in panel A. The dotted circles represent residues that have been suggested to form interactions with rhodamine[55;186;187].

strongest interactions from these simulations are the ones both poses share. All of these facts strongly suggest that the amitriptyline binding pose seen after free diffusion is more likely to be the actual binding site of amitriptyline in P-gp.

This simple analysis showcases LINTools' ability to potentially discriminate between different binding poses obtained from MD and molecular docking. This feature could be further enhanced in further development of LINTools.

6.7 Drug binding in bromodomains

Bromodomains initiate the recognition of ε-N-acetylation of histone lysine residues and as such proteins with bromodomains are involved in the regulation of transcription and have been identified to play a key role in several aggressive types of cancer[236]. Therefore, targeting these proteins could provide treatments for cancer.

All bromodomains have a conserved fold comprised of a left-handed bundle of four alpha helices linked by diverse loop regions that are flanking the substrate binding site - a hydrophobic cavity that recognises ε-N-acetylated lysine residues. A conserved asparagine residue forms a hydrogen bond with the N-acetyl group of the substrate and this interaction has been targeted during the drug discovery process for the development of acetyl-lysine mimetics.

Several 10 ns simulations of bromodomain BRD4(1) were kindly provided by Matteo Aldeghi[237] to use as one of the test sets for LINTools. In this case the same protein (BRD4(1)) is interacting with different compounds, allowing to showcase the flexibility of the tool and the ease of comparison between simulations.

6.7.1 Alprazolam

Triazolo-benzodiazepines have recently been described as potent and highly selective protein interaction inhibitors for a specific family of bromodomains and alprazolam, a trialo-benzodiazepine compound used to treat anxiety disorders, is binding to BRD4(1) with low µM affinity[238].

In the holo crystal structure of BRD4(1) with alprazolam, a nitrogen on the triazole ring of the compound forms a single hydrogen bond with the conserved asparagine 140[238]. However, the simulations reveal that this hydrogen bond can be formed with the nitrogen in either position 1 or 2 on the triazole ring (see Figure 6.9A). The chlorophenyl substituent of the diazepine ring is found strongly interacting with WP motif formed by TRP81 and PRO82 (see Figure 6.9A). This position is important since in other crystal structures together with the surrounding loops it forms the second acetylated lysine binding site[236]. Very few interactions are observed for the benzyl ring (Figure 6.9A) which is explained by the increased solvent accessibility of this group (Figure 6.9A).

6.7.2 3,5-Dimethylisoxazole derivative

The 3,5-dimethylisoxazole moiety has been shown to be an effective mimic of the acetylated lysine leading to development of a chemically distinct class of bromodomain ligands[239]. LINTools analysis for the compound present in crystal structure with PDB code 4J0S is shown in Figure 6.10A. Similarly to alprazolam (see Figure 6.9A), this compound can have hydrogen bonds with either nitrogen or oxygen atom on the 3,5-dimethylisoxazole ring (see Figure 6.10A). The benzyl group of this compound is less solvent exposed (Figure 6.10B) than the same group in alprozalam (Figure 6.9B) which is reflected in the increased number of interactions within

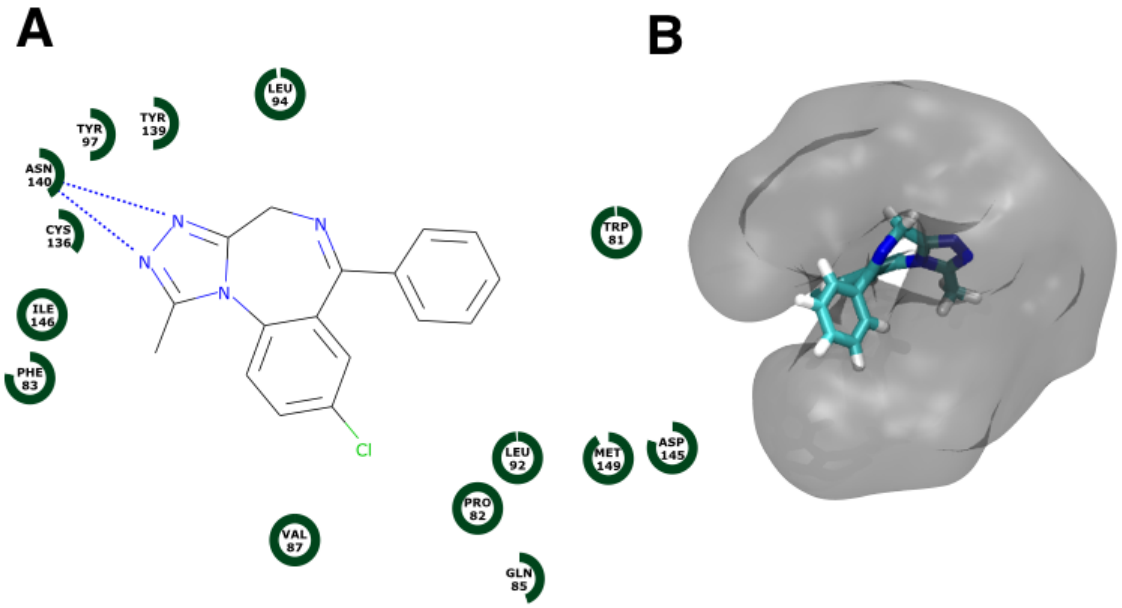


Figure 6.9: (A) Alprazolam interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue. (B) Alprazolam placement within BRD4(1) acetylated lysine binding site. For clarity, only residues within 5 Å of the compound are shown as a grey surface to observe which parts of the molecule are solvent exposed. The compound is shown in licorice representation with carbon atoms in cyan, hydrogen in white, nitrogen in blue, chloride in green and oxygen in red. The same scheme has been used throughout the section.

3.5 Å cutoff (Figure 6.10A) Interestingly, the hydroxy-phenyl ring formed a hydrogen bond with the PRO82, which forms part of the WP shelf and is known to be involved in binding of second acetylated lysine residue[236].

6.7.3 RVX-208 and RVX-OH

BRD4 is part of the bromo and extraterminal (BET) protein family and the human BET proteins are have two BRD domains. Even though these domains are conserved, sequence alignment revealed that three residues in close proximity to the acetylated lysine binding site that differ between the domains[240]. Due to their similarity, it is difficult to find inhibitors that are domain-specific, however, a recent study describes RVX-208, a quinazolone specific to the second domain of the BET proteins, while a related compound RVX-OH shows no significant specificity[240].

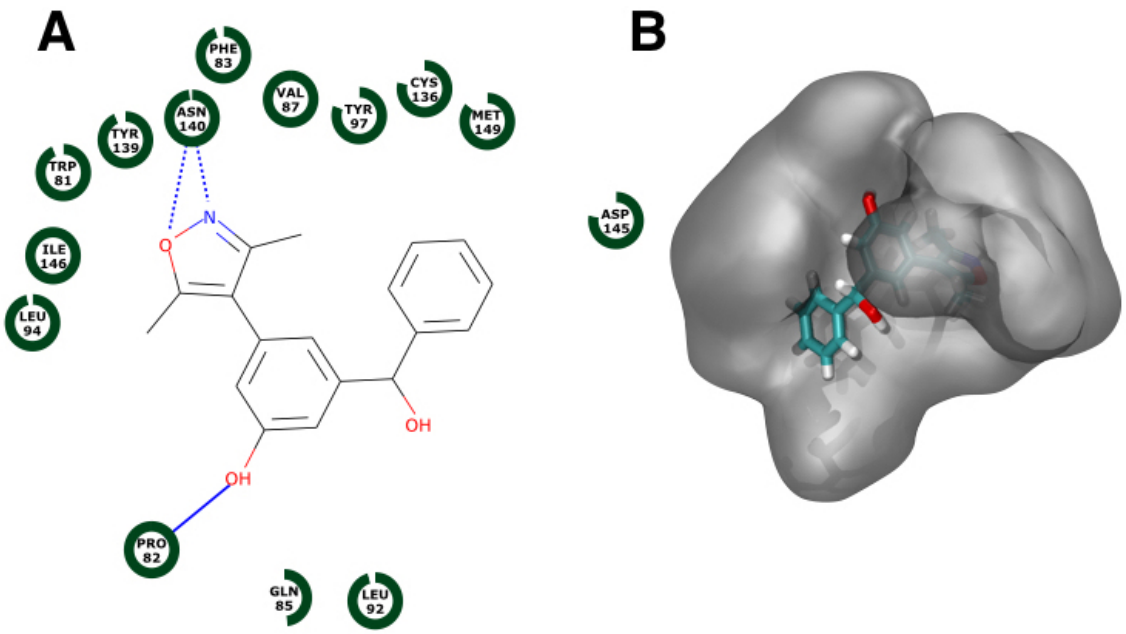


Figure 6.10: (A) 3,5-dimethylisoxazole derivative interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot, where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue and the solid blue line shows hydrogen bond with the backbone of a residue. (B) 3,5-dimethylisoxazole derivative placement within BRD4(1) acetylated lysine binding site.

Surprisingly, the binding and interaction modes of these compounds are different. The quinazolone ring system on RVX-208 behaves as acetylated lysine mimic and forms two hydrogen bonds with the conserved asparagine 140 (see[240] and Figure 6.11). The hydroxy-ethylether moiety points out of the binding site and is solvent accessible (Figure 6.11B) and only makes a few contacts with the BRD4(1) (Figure 6.11A). On the other hand, the binding mode of RVX-OH is inverted (see[240] and Figure 6.12) with the OH group on the phenyl ring forming a hydrogen bond with ASN140 and the quinazolone ring system pointing towards solvent. The 2D diagrams from LINTools (Figure 6.11A and Figure 6.12A) provide clear comparison between the binding modes of the quinazolone compounds in the acetylated lysine binding site of first BRD domain of BRD4 or BRD4(1).

6.7.4 Dual bromodomain-kinase inhibitors

Disruption of several signalling pathways is often the hallmark of cancer development. Kinases have become attractive targets for cancer therapy due to their aberrant regulation. The aim of the polypharmacologic approach is to produce drugs that can target multiple pathways to reduce the possibility of compensating signalling taking place during cancer therapy. Recently, dual bromodomain-kinase inhibitors have been discovered, making them an attractive therapy option for cancers where both target families are involved[241]. Two compounds (BI2356 and TG101348) that showed nanomolar affinity for bromodomains are especially noteworthy[241] and have been subjected to MD simulations to reveal their interactions with BRD4(1)[237].

The co-crystal structures of BRD4(1) with BI2356 and TG101348[241] were in good agreement with the plots acquired from LINTools analysis of the MD simulations (Figure 6.13A and reftgA). BI2356 formed a single hydrogen bond with the conserved asparagine residue (Figure 6.13A, while TG101348 was a closer mimic to the

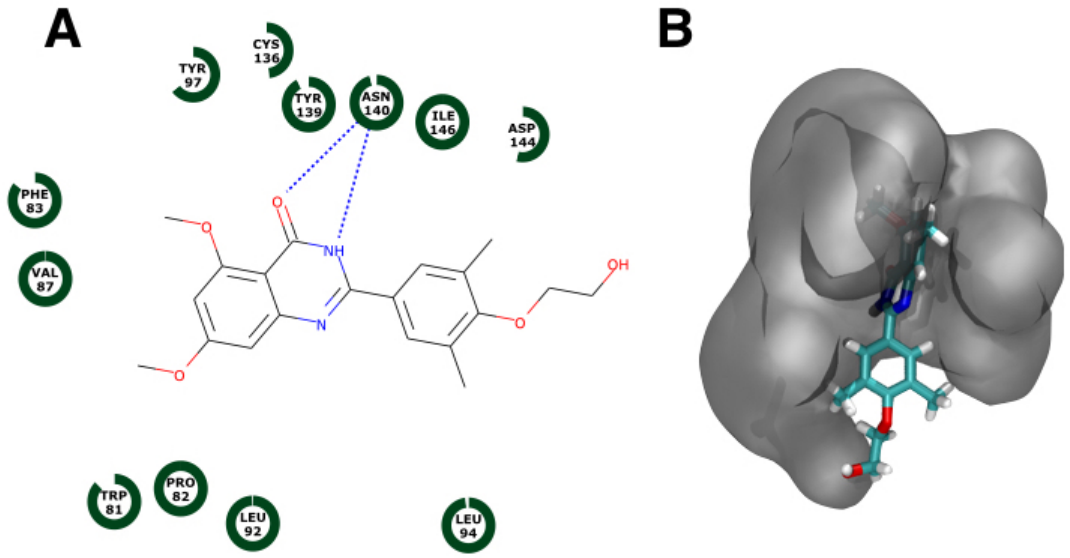


Figure 6.11: (A) RVX-208 interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue. (B) The placement of RVX-208 within BRD4(1) acetylated lysine binding site.

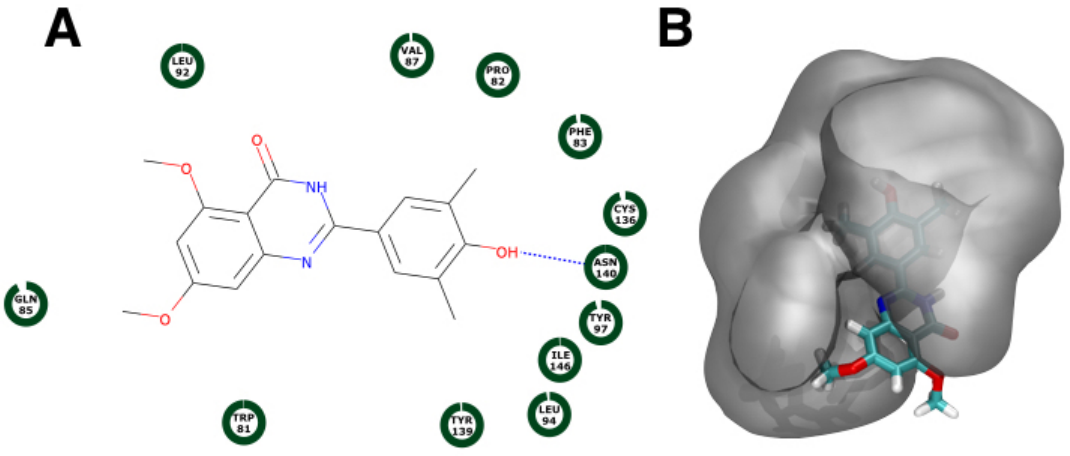


Figure 6.12: (A) RVX-OH interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue. (B) The placement of RVX-OH within BRD4(1) acetylated lysine binding site.

acetylated lysine with two hydrogen bonds formed with the N140 (Figure 6.14A). Surprisingly, both compounds also formed hydrogen bonds with the backbone of PRO82, similarly to the 3,5-dimethylisoxazole derivative discussed above. The dual bromodomain-kinase inhibitors are larger than the other compounds discussed in this section and as such it is perhaps unsurprising that parts of the molecules are solvent exposed (see Figure 6.13B and Figure 6.14A). However, while the pyrrolidine ring of TG101348 maintained at least one contact with the bromodomain surface (e.g. LEU94 in Figure 6.14A), the methylpyridium moiety in BI2356 is completely solvent accessible (Figure 6.13).

6.7.5 Summary of bromodomain interactions

LINTools was used to compare and describe interactions between BRD4(1) and several inhibitors that bind in the acetylated lysine binding site. All of these ligands formed at least one hydrogen bond with the conserved asparagine residue at position 140, which is the main hallmark of a acetylated lysine mimic. Interestingly, some compounds (e.g. TG101348 and BI2356) formed a hydrogen bond with the backbone of PRO82. This is particularly interesting since this residue is part of a WP motif which forms the second acetylated lysine binding site[236].

It was possible to distinguish which parts of the molecule were solvent exposed just by looking at the number of interactions with the particular group or moiety. However, it would be preferable to show solvent exposure of a molecule in more detail, which is one of the future directions for LINTools.

Another future direction for LINTools would be to introduce the possibility to account for water-bridged hydrogen bonds. This feature is especially important for bromodomains that have several conserved water molecules in the acetylated lysine binding site[236].

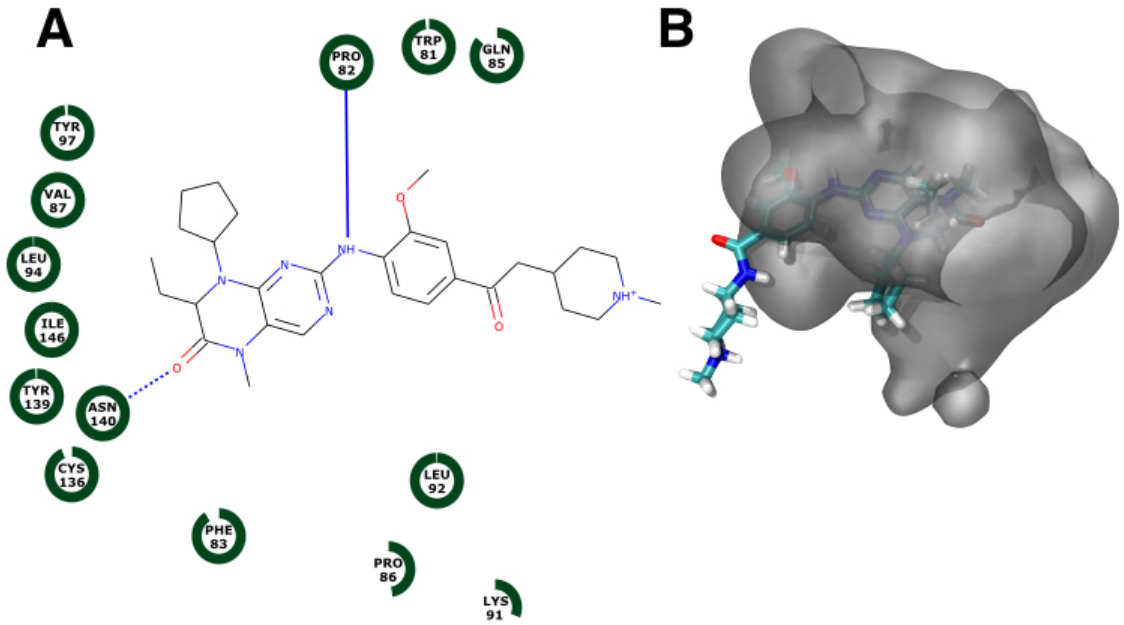


Figure 6.13: (A) BI2356 interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue and the solid blue line shows hydrogen bond with the backbone of a residue. (B) The placement of BI2356 within BRD4(1) acetylated lysine binding site.

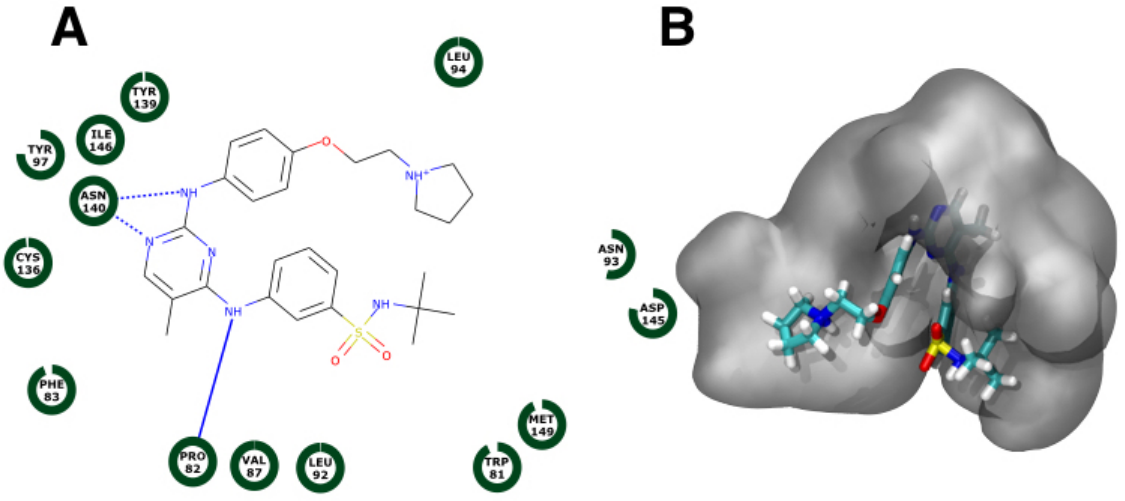


Figure 6.14: (A) TG101348 interactions with BRD4(1) from 10 ns simulation are depicted as a LINTools plot where the fraction of the circle around a residue name shows the time spent within 3.5 Å of the compound. The dotted blue lines show hydrogen bonds with the sidechain atoms of a residue and the solid blue line shows hydrogen bond with the backbone of a residue. (B) The placement of TG101348 within BRD4(1) acetylated lysine binding site.

In this section, LINTools was used to provide a quick comparison between a diverse set of molecules interacting with the same protein, in this case, the first bromodomain of BRD4. The information about hydrogen bonds and the strength and placement of interactions is clearly shown in the LINTools plots, however, for a more full picture of the interactions more features should be added.

6.8 Conclusions

Molecular dynamics has become an important tool in drug discovery and has been used to investigate the protein-ligand interactions in the both temporal and spatial dimensions. The availability of longer timescales and improvements in small molecule parametrization have made this method more relevant and available, with the free energy of ligand binding being an area of particular improvement over the classical approach of molecular docking.

Therefore it is unsurprising that the fields of computational chemistry and biochemistry have responded to the popularity of MD by producing libraries that can parse and analyse trajectory data (e.g. MDAnalysis[207]). Coupled with the increased use of Python in scientific analysis, the availability of chemical and biochemical Python libraries allow to produce universal analysis scripts for MD.

The increased interest in protein-ligand interactions has resulted in several tools that produce schematic diagrams of these interactions, however, these only consider the static state captured in a crystal structure or docked pose. In this chapter we present a novel open-source program that produces 2D schematic graphs of protein-ligand interactions in MD simulations - Ligand Interaction Network Tools or LINTools. Addition of the temporal dimension allows to observe the dynamics of the binding site, thereby further characterising the interactions.

6.8.1 Reasons to use LINTools

One of the main benefits of LINTools is its open-source nature and the consequent possibility for extension and modifications by the user base. LINTools are freely available through GNU General Public Licence and are stored in a GitHub repository. All LINTools dependencies are also freely available through variety of sources. This ethos has been very important throughout the development of LINTools with the aim of maximally increasing its availability and usefulness to the potential users.

Another important characteristic of LINTools is its flexibility as it can accept many different file formats for analysis due to the variety of file readers implemented in MDAnalysis - a property exploited by LINTools through this dependency. The output of LINTools at the moment is an image file and the choice of its format was based on several considerations - whether it is possible to mutate the file in the Python environment and if the resulting images had high quality and could be easily converted to another format if required. The SVG file format fitted all of these requirements and the users can modify the resulting file either directly with a text editor or using an image processing software.

Although it is still in development, with increase of the user base of LINTools, more uses for the software will be determined. The feedback from the users will be used for future developments and adjustments allowing natural expansion according to the interest of public. Nevertheless, several additional features and extensions have already been planed and the process will likely involve a survey of potential and existing users to decide on the best future directions and implementation strategies.

6.8.2 Future directions for LINTools development

The LINTools version described in this chapter consists of a framework that is required to make the 2D schematic representation of a ligand and finding the appropriate coordinates for interaction site residue placement. Additionally, it analyses residue residence time, ligand RMSF and the hydrogen bonds formed between ligand and protein residues. These features are contained within a Python script that also compiles and saves the final image. This tool is invoked from the command line and the options are determined either as flags or selections.

However, there is much room for improvement and additional features. These can mainly be divided in improvements to the user experience and additions to existing analysis and visualisation. Additionally, continued effort will be put in increasing the robustness of the script by developing more tests and increasing test coverage, as well as improving documentation and adding short tutorials.

Several changes are planned to improve the overall user experience while obtaining and using LINTools. First, the installation will be made easier through adding the script to the conda package manager. This will allow easy installation, since LINTools and all of its dependencies will be installed with a single command in the console. Second, the usage of the script will be made easier to navigate through adding more documentation and examples, as well as simplifying the input process as much as possible. For example, the domain file will be given a new, more intuitive file type that will not require such strict adherence to rules of input. Third, the addition of a progress bar will allow users to follow the process and immediately see if there is any problem with running the script. Last, options for numerical output of the analysis scripts will be added. This feature will create additional text files that can be plotted by conventional plotting software or viewed in a text editor as is.

There are many potential upgrades to the analysis and visualisation parts of LINTools

and these problems tend to go hand in hand - if a new type of analysis is added, it might require changes or additions in the visualisations of the ligand, protein residues or the rest of the image. First, more types of interactions will be added, such as salt bridges, hydrophobic interactions, aromatic stacking, π-cation interactions and water-bridged hydrogen bonds. Second, the ligand molecule will be more carefully characterised using some of the pharmacophore descriptors from RDKit. Third, an option for direct comparison between repeats or simulations of mutated protein can be added. Last, information about the overall state of interaction site and protein would be available (e.g. RMSD).

LINTools is a novel and promising script that combines the classical approach of 2D schematic diagrams of protein-ligand interactions and adds the information obtained from MD simulations. It has shown that the medium of a 2D diagram can be successfully used to create clear and simple representations of protein-ligand interactions not only from static 3D conformations but including the time dimension as well. However, the full potential and maximal user experience can probably be achieved through an interactive graphical interface which would allow users to choose most optimal representation of their data on the fly. This goal will require a strong underlying framework, the basis of which has been presented in this chapter.

7

Conclusions and future directions

7.1 Summary and conclusions

Due to its prominent role in multidrug resistance in cancer and CNS disorders P-glycoprotein has been the subject of many studies that have investigated its structure and function (for a review see[242]). However, disagreement between structural and experimental data has caused a lot of controversy in the field - most notably in studies concerning conformational changes of the protein and substrate binding/unbinding pathways. The structural studies are further complicated by the lack of available

human P-glycoprotein crystal structure.

With the publication of the structures of several proteins related to human P-glycoprotein, it became possible to create homology models using these structures as templates. Chapter 3 describes a study that compares homology models made from several different templates by evaluating their absolute quality. A shortlist of three homology models was extracted from this process and due to the similarity between their quality scores, further assessment was required. Therefore, the dynamics of shortlisted homology models was investigated by unbiased molecular dynamics simulations with the proteins inserted in model POPC bilayer. This analysis revealed that the recently published refined mouse P-glycoprotein crystal structure (PDB 4M1M)[47] produces a homology model of human P-glycoprotein that is stable within the simulation timescale. High sequence identity (87%) and structural stability within MD simulations were the main reasons to choose the homology model from refined mouse crystal structure for further experiments.

Substrate entrance pathway

The "hydrophobic vacuum cleaner" model predicts that the substrates gain entrance to the binding site in the central cavity of P-glycoprotein from the inner leaflet of the lipid bilayer[93]. Presumably, this process would involve movement through the transmembrane helices of P-gp and structural data suggests that the entrance gates could be formed by the TMH 4/6 or TMH 10/12 due to the pseudosymmetry of P-glycoprotein[32]. However, the closure of the intramembrane sections of the putative entrance gates observed in the MD simulations suggested a more careful investigation of the substrate entrance pathway using enhanced sampling methods required to observe rare events[96].

The substrate entrance pathway was investigated with steered MD simulations to

determine the most likely point of entrance to the central cavity. Amitriptyline was pulled though both intramembrane and cytoplasmic portions of the putative entrance gates and the resulting work profiles suggested that the cytoplasmic entrance route is strongly preferred, in particular, the path through transmembrane helices 4 and 6. The egress pathway of amitriptyline obtained from PELE simulations showed that the molecule prefers to exit through the TMH4/6 cytoplasmic gate, supporting the observations from SMD simulations.

Additional unbiased simulations were used to investigate amitriptyline dynamics within the central cavity and an event of free diffusion towards the putative binding site was observed. Further information on amitriptyline binding was obtained from consensus docking. While the binding sites predicted by the two methods were different, they shared some common features such as similarity to the rhodamine binding site.

Therefore, we have provided a more detailed description of the substrate entrance pathway with novel features such as entrance through cytoplasmic section of the previously proposed entrance gates. Additionally, amitriptyline binding in the central cavity of P-gp was explored in more detail. In conclusion, the proposed pathway starts with a substrate in the inner leaflet of the bilayer, followed by lateral diffusion across the lipid bilayer. When the substrate is close to the protein, access to the central cavity is gained through the cytoplasmic section of TMH4/6 and once inside, free diffusion can guide the compound to the final binding site.

Interactions with lipids

The surrounding bilayer environment has a critical role in P-glycoprotein function and structural stability (for a review see[29]). Therefore, P-glycoprotein was inserted in a complex asymmetric lipid bilayer designed to mimic the human blood-brain barrier to

provide a more physiologically relevant environment. The coarse grained simulations of this system investigated how P-glycoprotein affects the bilayer in its vicinity. These simulations revealed local ordering of the phospholipids and cholesterol around the protein, with the negatively charged POPS forming an annular ring. Reduction in diffusion constants for all lipids was also observed - an effect probably caused by ordering.

P-glycoprotein interactions with cholesterol have been a significant area of interest, with recent studies suggesting an intimate relationship between the two molecules[105]. Several potential cholesterol binding sites were observed in the coarse grained simulations. The pattern of these interactions fits with non-annular surface lipid binding, where the lipid binding sites are formed by cavities on the surface of the protein or by the interface between subunits. PMF calculations were used to determine the strength of cholesterol binding at the potential binding sites and the site positioned between transmembrane helices 10 and 12 showed binding affinity similar to that predicted by experiments. The same site was used for several types of lipid molecules to enter the central cavity. While this observation fits with the proposed role of P-gp as a low-level lipid flippase, it disagrees with the conclusions reached in chapter 4.

The dynamics of P-glycoprotein behaviour in the complex bilayer was also explored in atomistic simulations. Even though the presence of a complex bilayer did not cause wider opening of the entrance gates, more flexibility in the distances between gate-forming transmembrane helices was observed.

Visualisation of dynamic protein-ligand interactions

Even though MD simulations have become an important tool in drug discovery, there is a lack of specialised tools that analyse and visualise protein-ligand interactions from the additional time dimension that can be acquired from MD simulations. Specialised

libraries for analysis of MD and chemistry data have appeared in recent years with the rise of the Python coding language, making it possible to create universal analysis and visualisation tools.

LINTools is a Python script that obtains information about protein-ligand interactions from molecular dynamics files and visualises it as a schematic 2D plot. Due to the flexibility of MDAnalysis[207], the MD analysis library used in this tool, it is possible to analyse many types of files produced by different MD simulation packages. LINTools shows protein-ligand interactions in a way familiar to most chemists - a 2D diagram - and adds the time dimension to this information. There are many potential applications for a script that incorporate both analysis and visualisation aspects, for example, screening results of docking or fragment-based analysis.

7.2 Future directions

Several interesting hypotheses have been proposed from the simulations described in this thesis. The closure of the proposed substrate intramembrane gate led us to investigate other possible options for thr substrate pathway to the central cavity and, as a result, a novel pathway was proposed. However, the non-equilibrium nature of the methods used for these experiments does not describe the pathway in enough detail, therefore, it would be of great interest to explore this pathway with equilibrium methods such as umbrella sampling and calculation of PMF or metadynamics. This approach would achieve greater sampling of the process and would provide the information about the free energy changes in the system.

The data generated by molecular dynamics can and have been used to direct and guide practical experiments and the work described in this thesis has also produced several potentially testable assumptions. The distance between transmembrane helices

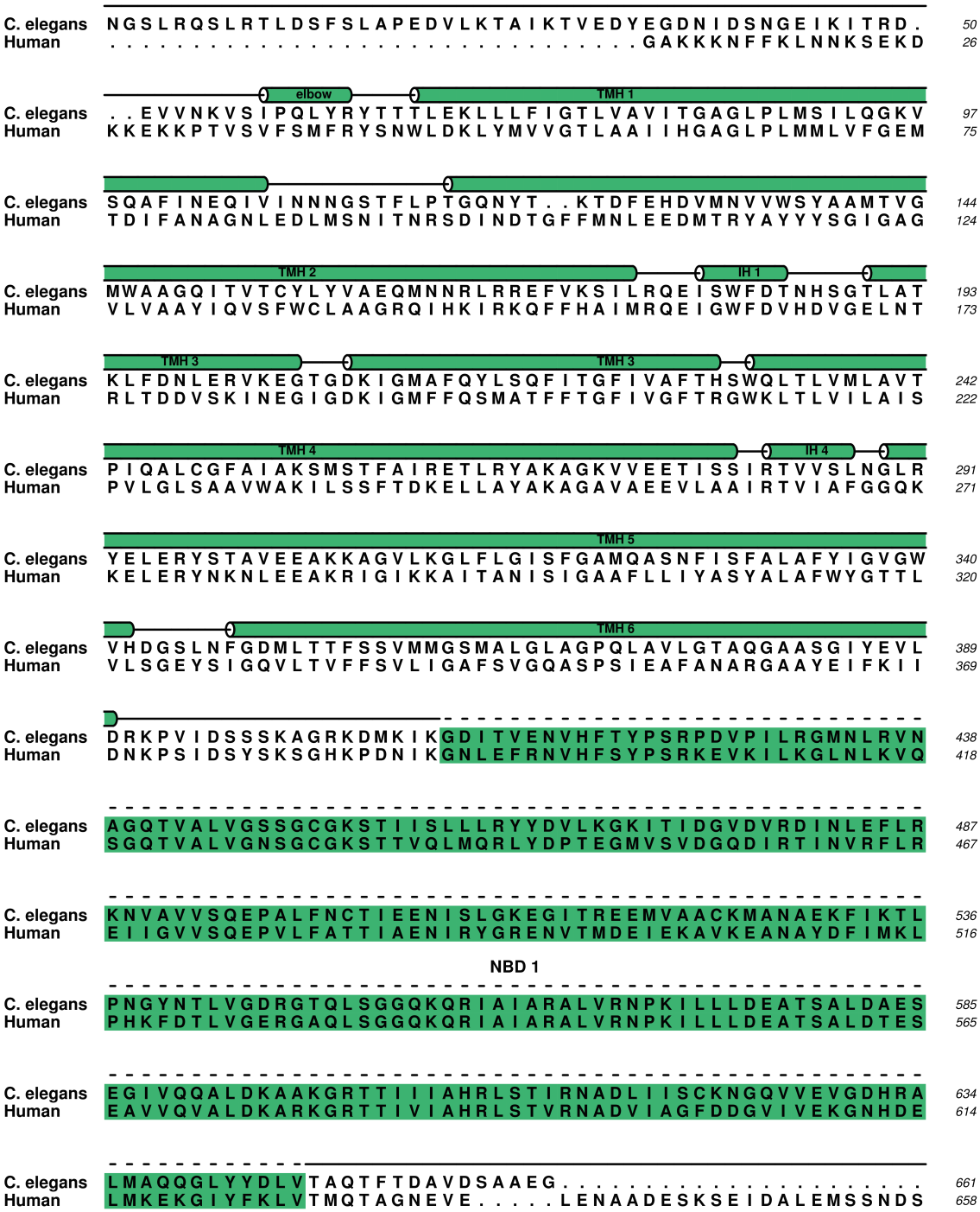
TMH4/6 and TMH10/12 decreases during most of our simulations - an observation that indirectly fits experimental evidence. However, a more direct experiment, such as measurement of distances between helices by crosslinking or spectroscopy experiments, could be used to compare simulation and experimental data. Similarly, NMR experiments could be used to determine the precise location of the cholesterol and amitriptyline binding sites and this information could be compared with the data from simulations described in this thesis.

As explained in the conclusions of chapter 6, LINTools is still under development and more testing and feedback is required to improve the tool. Additionally, new features can also be implemented, in part responding to the feedback from the existing user base, as well as trying to keep up and surpass the already existing tools for visualisation and analysis of static protein-ligand interactions.

A

Sequence alignments for homology modelling

The homology modelling of human P-glycoprotein used several different templates from crystal structures of related proteins published recently. The best homology models, as assessed by QMEAN Z-score, used *C. elegans*, *C. merolae* and refined mouse ABCB1 crystal structures as templates. The sequence alignments with these proteins have been provided in A.


elbow
TMH 1
TMH 2
IH 1
TMH 3
TMH 3
TMH 4
IH 4
TMH 5
TMH 6
NBD 1
C. elegans NGSLRQSLRTLDSFSLAPEDVLKTAIKTVEDYEGDNIDSNGEIKITRD. 50
Human .....GAKKKNFFKLNNKSEKD 26
C. elegans ..EVVNKVSIPQLYRYTTTLEKLLLFIGTLVAVITGAGLPLMSILQGKV 97
Human KKEKKPTVSVFSMFRYSNWLDKLYMVVGTLAAIIHGAGLPLMMLVFGEM 75
C. elegans SQAFINEQIVINNNGSTFLPTGQNYT..KTDFEHDVMNVVWSYAAMTVG 144
Human TDIFANAGNLEDLMSNITNRSDINDTGFFMNLEEDMTRYAYYYSGIGAG 124
C. elegans MWAAGQITVTCYLYVAEQMNNRLRREFVKSILRQEISWFDTNHSGTLAT 193
Human VLVAAYIQVSFWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHDVGELNT 173
C. elegans KLFDNLERVKEGTGDKIGMAFQYLSQFITGFIVAFTHSWQLTLVMLAVT 242
Human RLTDDVSKINEGIGDKIGMFFQSMATFFTGFIVGFTRGWKLTLVILAIS 222
C. elegans PIQALCGFAIAKSMSTFAIRETLRYAKAGKVVEETISSIRTVVSLNGLR 291
Human PVLGLSAAVWAKILSSFTDKELLAYAKAGAVAEEVLAAIRTVIAFGGQK 271
C. elegans YELERYSTAVEEAKKAGVLKGLFLGISFGAMQASNFISFALAFYIGVGW 340
Human KELERYNKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTL 320
C. elegans VHDGSLNFGDMLTTFSSVMMGSMALGLAGPQLAVLGTAQGAASGIYEVL 389
Human VLSGEYSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKII 369
C. elegans DRKPVIDSSSKAGRKDMKIKGDITVENVHFTYPSRPDVPILRGMNLRVN 438
Human DNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQ 418
C. elegans AGQTVALVGSSGCGKSTIISLLLRYYDVLKGKITIDGVDVRDINLEFLR 487
Human SGQTVALVGNSGCGKSTTVQLMQRLYDPTEGMVSVDGQDIRTINVRFLR 467
C. elegans KNVAVVSQEPALFNCTIEENISLGKEGITREEMVAACKMANAEKFIKTL 536
Human EIIGVVSQEPVLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKL 516
C. elegans PNGYNTLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATSALDAES 585
Human PHKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTES 565
C. elegans EGIVQQALDKAAKGRTTIIIAHRLSTIRNADLIISCKNGQVVEVGDHRA 634
Human EAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDE 614
C. elegans LMAQQGLYYDLVTAQTFTDAVDSAAEG..... 661
Human LMKEKGIYFKLVTMQTAGNEVE.....LENAADESKSEIDALEMSSNDS 658

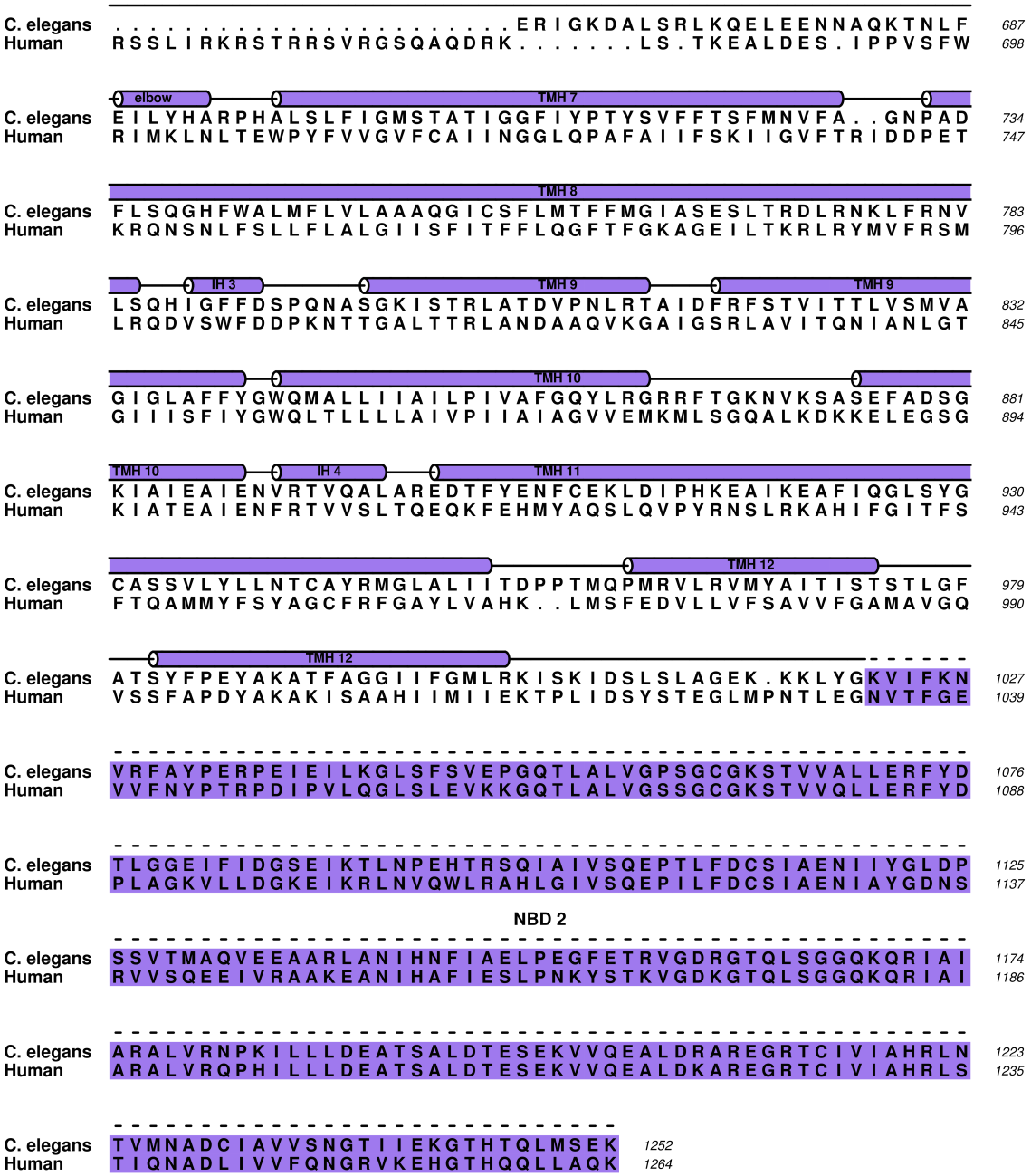
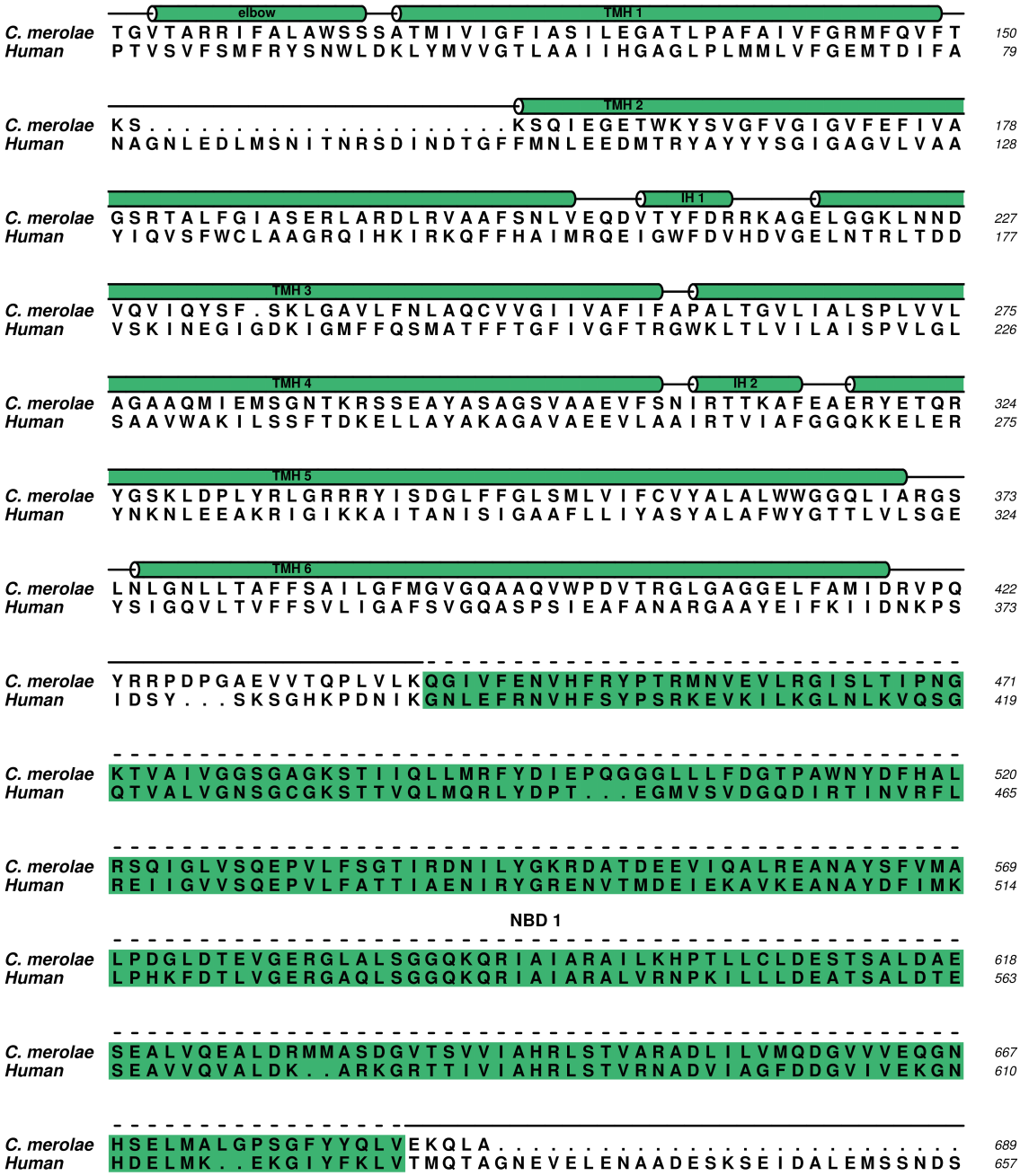


Figure A.1: The sequence alignment between *C. elegans* and human P-glycoprotein


elbow
TMH 1
C. merolae TGVTARRIFALAWSSSATMIVIGFIASILEGATLPAFAIVFGRMFQVFT 150
Human PTVSVFSMFRYSNWLDKLYMVVGTLAAIIHGAGLPLMMLVFGEMTDIFA 79
TMH 2
C. merolae KS.....KSQIEGETWKYSVGFVGIGVFEFIVA 178
Human NAGNLEDLMSNITNRSDINDTGFFMNLEEDMTRYAYYYSGIGAGVLVAA 128
IH 1
C. merolae GSRTALFGIASERLARDLRVAAFSNLVEQDVTYFDRRKAGELGGKLNND 227
Human YIQVSFWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHDVGELNTRLTDD 177
TMH 3
C. merolae VQVIQYSF.SKLGAVLFNLAQCVVGIIVAFIFAPALTGVLIALSPLVVL 275
Human VSKINEGIGDKIGMFFQSMATFFTGFIVGFTRGWKLTLVILAISPVLGL 226
TMH 4
IH 2
C. merolae AGAAQMIEMSGNTKRSSEAYASAGSVAAEVFSNIRTTKAFEAERYETQR 324
Human SAAVWAKILSSFTDKELLAYAKAGAVAEEVLAAIRTVIAFGGQKKELER 275
TMH 5
C. merolae YGSKLDPLYRLGRRRYISDGLFFGLSMLVIFCVYALALWWGGQLIARGS 373
Human YNKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGE 324
TMH 6
C. merolae LNLGNLLTAFFSAILGFMGVGQAAQVWPDVTRGLGAGGELFAMIDRVPQ 422
Human YSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPS 373
C. merolae YRRPDPGAEVVTQPLVLKQGIVFENVHFRYPTRMNVEVLRGISLTIPNG 471
Human IDSY...SKSGHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSG 419
C. merolae KTVAIVGGSGAGKSTIIQLLMRFYDIEPQGGGLLLFDGTPAWNYDFHAL 520
Human QTVALVGNSGCGKSTTVQLMQRLYDPT...EGMVSVDGQDIRTINVRFL 465
C. merolae RSQIGLVSQEPVLFSGTIRDNILYGKRDATDEEVIQALREANAYSFVMA 569
Human REIIGVVSQEPVLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMK 514
NBD 1
C. merolae LPDGLDTEVGERGLALSGGQKQRIAIARAILKHPTLLCLDESTSALDAE 618
Human LPHKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTE 563
C. merolae SEALVQEALDRMMASDGVTSVVIAHRLSTVARADLILVMQDGVVVEQGN 667
Human SEAVVQVALDK..ARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGN 610
C. merolae HSELMALGPSGFYYQLVEKQLA..... 689
Human HDELMK..EKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDS 657

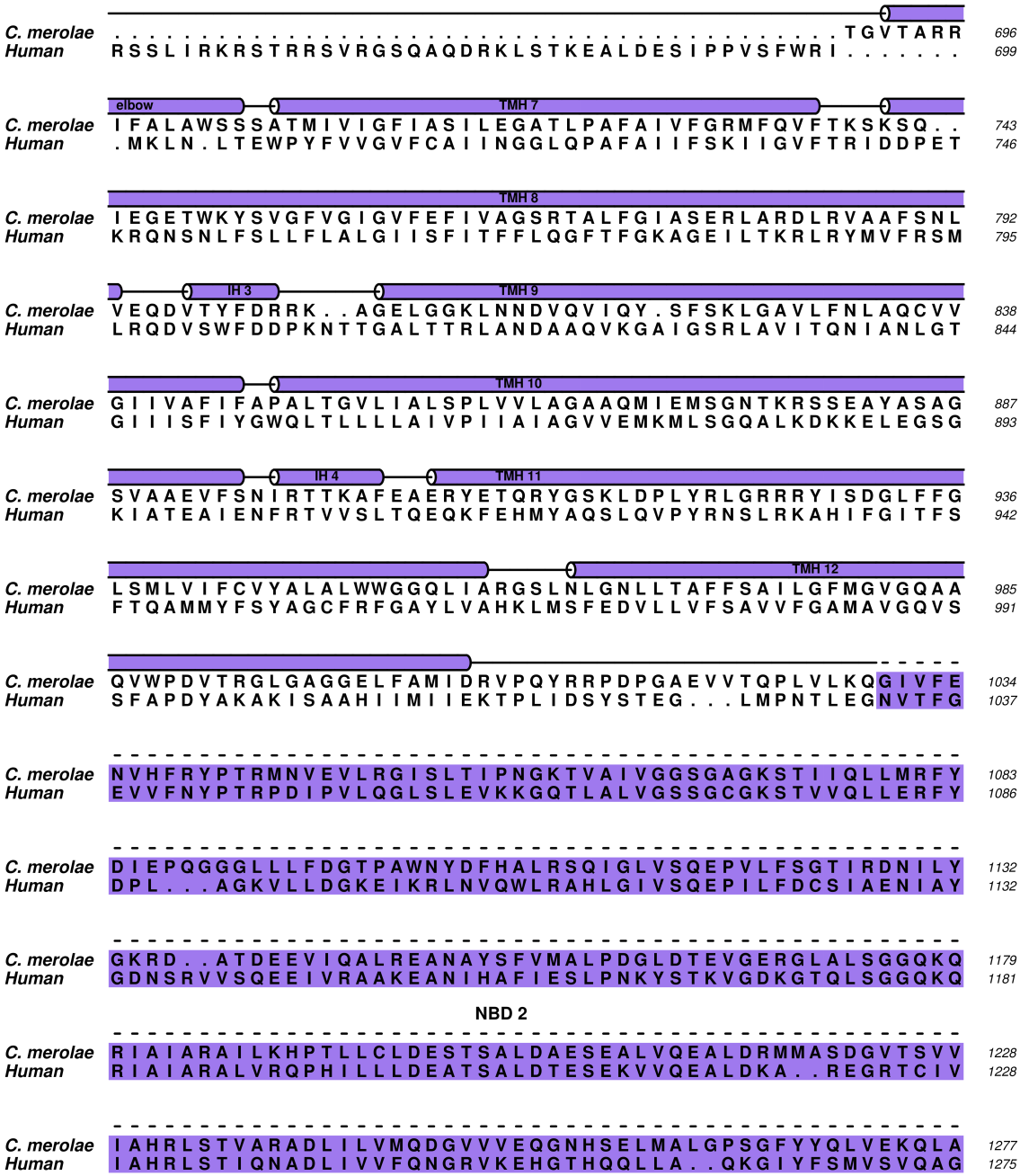
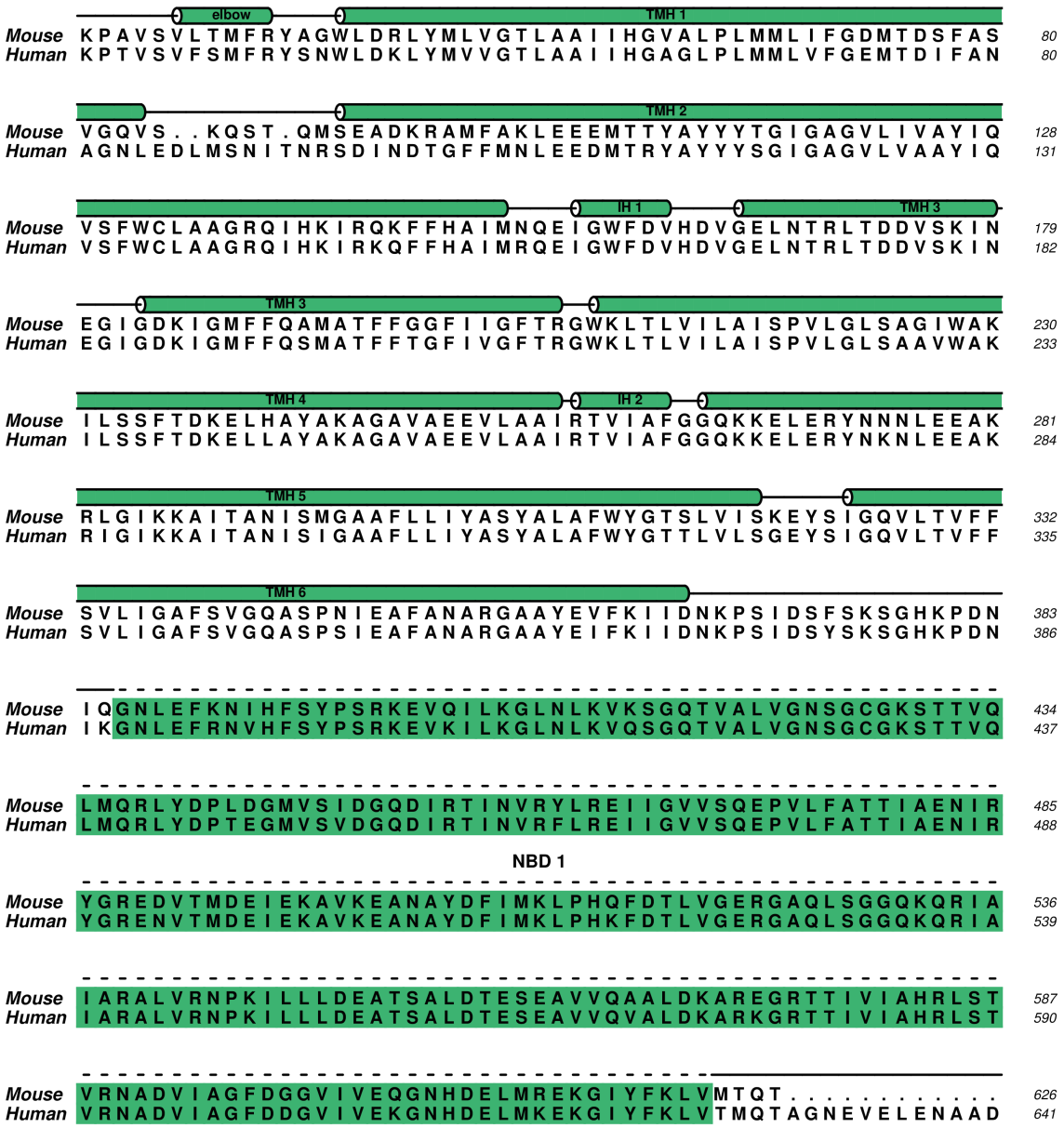


Figure A.2: The sequence alignment between *C. merolae* and human P-glycoprotein


elbow
TMH 1
Mouse K P A V S V L T M F R Y A G W L D R L Y M L V G T L A A I I H G V A L P L M M L I F G D M T D S F A S 80
Human K P T V S V F S M F R Y S N W L D K L Y M V V G T L A A I I H G A G L P L M M L V F G E M T D I F A N 80
TMH 2
Mouse V G Q V S . . K Q S T . Q M S E A D K R A M F A K L E E E M T T Y A Y Y Y T G I G A G V L I V A Y I Q 128
Human A G N L E D L M S N I T N R S D I N D T G F F M N L E E D M T R Y A Y Y Y S G I G A G V L V A A Y I Q 131
IH 1
TMH 3
Mouse V S F W C L A A G R Q I H K I R Q K F F H A I M N Q E I G W F D V H D V G E L N T R L T D D V S K I N 179
Human V S F W C L A A G R Q I H K I R K Q F F H A I M R Q E I G W F D V H D V G E L N T R L T D D V S K I N 182
TMH 3
Mouse E G I G D K I G M F F Q A M A T F F G G F I I G F T R G W K L T L V I L A I S P V L G L S A G I W A K 230
Human E G I G D K I G M F F Q S M A T F F T G F I V G F T R G W K L T L V I L A I S P V L G L S A A V W A K 233
TMH 4
IH 2
Mouse I L S S F T D K E L H A Y A K A G A V A E E V L A A I R T V I A F G G Q K K E L E R Y N N N L E E A K 281
Human I L S S F T D K E L L A Y A K A G A V A E E V L A A I R T V I A F G G Q K K E L E R Y N K N L E E A K 284
TMH 5
Mouse R L G I K K A I T A N I S M G A A F L L I Y A S Y A L A F W Y G T S L V I S K E Y S I G Q V L T V F F 332
Human R I G I K K A I T A N I S I G A A F L L I Y A S Y A L A F W Y G T T L V L S G E Y S I G Q V L T V F F 335
TMH 6
Mouse S V L I G A F S V G Q A S P N I E A F A N A R G A A Y E V F K I I D N K P S I D S F S K S G H K P D N 383
Human S V L I G A F S V G Q A S P S I E A F A N A R G A A Y E I F K I I D N K P S I D S Y S K S G H K P D N 386
Mouse I Q G N L E F K N I H F S Y P S R K E V Q I L K G L N L K V K S G Q T V A L V G N S G C G K S T T V Q 434
Human I K G N L E F R N V H F S Y P S R K E V K I L K G L N L K V Q S G Q T V A L V G N S G C G K S T T V Q 437
Mouse L M Q R L Y D P L D G M V S I D G Q D I R T I N V R Y L R E I I G V V S Q E P V L F A T T I A E N I R 485
Human L M Q R L Y D P T E G M V S V D G Q D I R T I N V R F L R E I I G V V S Q E P V L F A T T I A E N I R 488
NBD 1
Mouse Y G R E D V T M D E I E K A V K E A N A Y D F I M K L P H Q F D T L V G E R G A Q L S G G Q K Q R I A 536
Human Y G R E N V T M D E I E K A V K E A N A Y D F I M K L P H K F D T L V G E R G A Q L S G G Q K Q R I A 539
Mouse I A R A L V R N P K I L L L D E A T S A L D T E S E A V V Q A A L D K A R E G R T T I V I A H R L S T 587
Human I A R A L V R N P K I L L L D E A T S A L D T E S E A V V Q V A L D K A R K G R T T I V I A H R L S T 590
Mouse V R N A D V I A G F D G G V I V E Q G N H D E L M R E K G I Y F K L V M T Q T 626
Human V R N A D V I A G F D D G V I V E K G N H D E L M K E K G I Y F K L V T M Q T A G N E V E L E N A A D 641

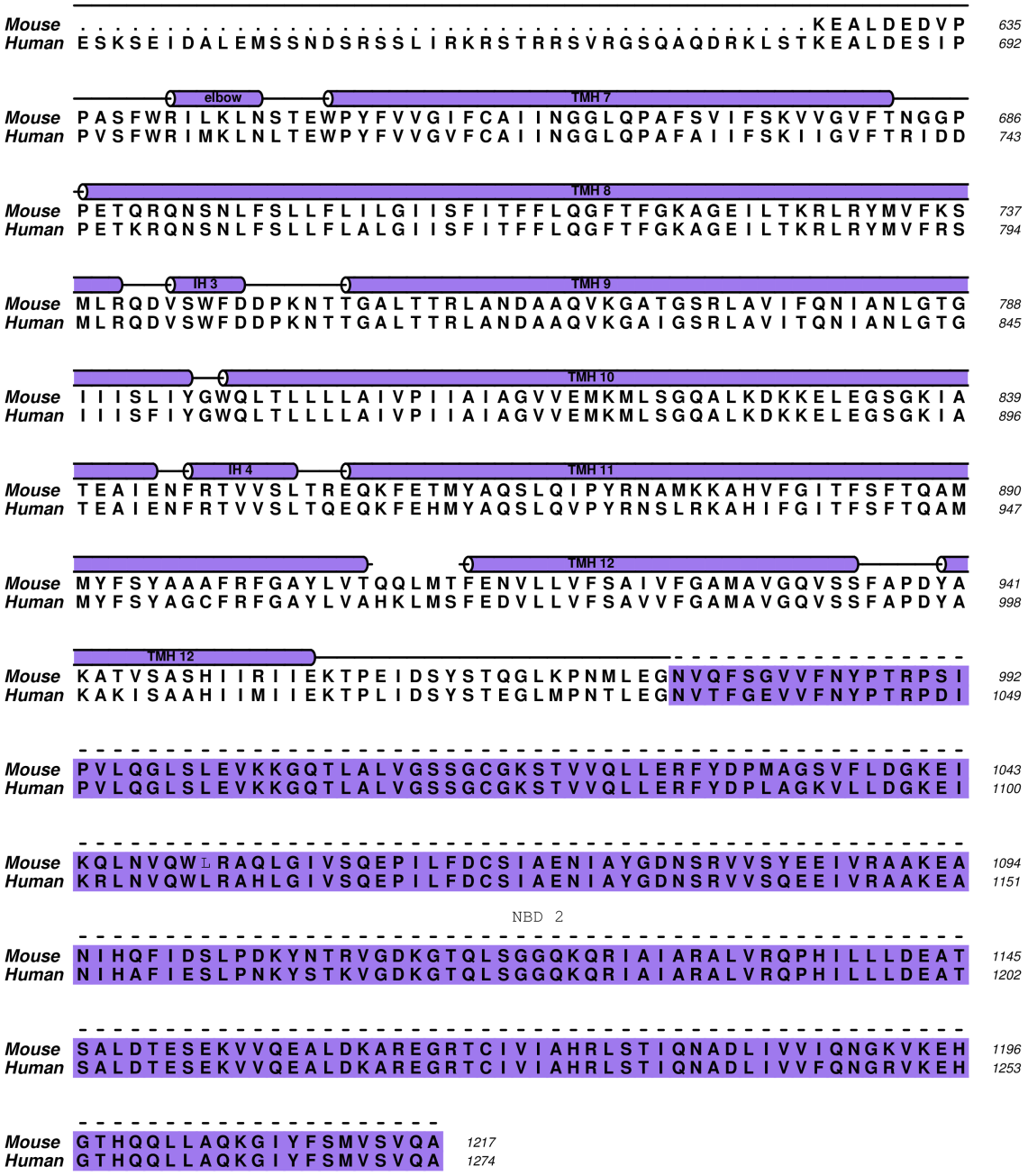


Figure A.3: The sequence alignment between mouse and human P-glycoprotein

B

NBD dynamics in the atomistic simulations of P-gp homology models and their templates

The dynamics of P-glycoprotein were investigated using atomistic molecular dynamics simulations in model lipid bilayer. The dynamics between domains of P-gp were exhibited in the RMSD and RMSF of protein Cα atoms, however, deeper insight can be gleaned from measurements of ATP binding site distances on NBDs. For Site 1,

we monitored the distance between the conserved lysine residue of the Walker A motif on NBD1 (Cα atom of K433) and the serine of the Signature motif of the other NBD2 (Cα atom of S1177). Similarly Site 2 was monitored via the distance between the Walker A lysine on NBD2 (K1076) and the Signature motif serine on NBD1 (S532).

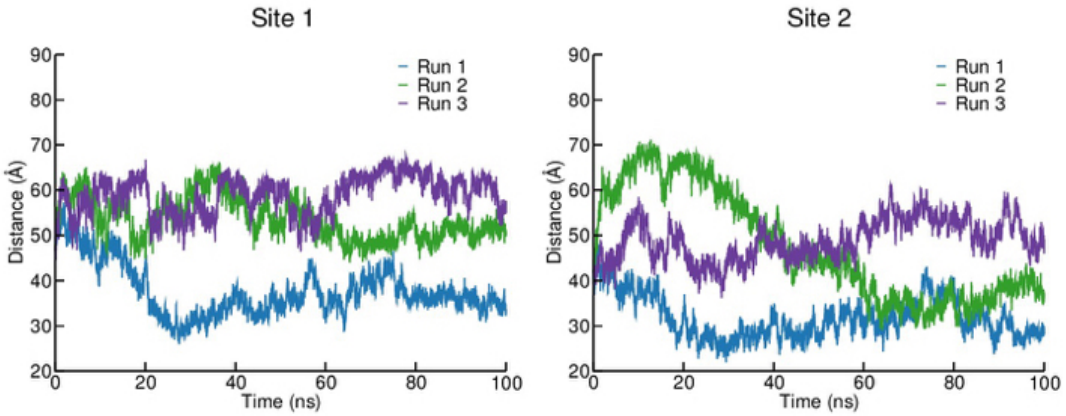


Figure B.1: The distance between NBDs of the homology model from the *C. elegans* crystal structure. The distance between NBDs decreased during the simulations.

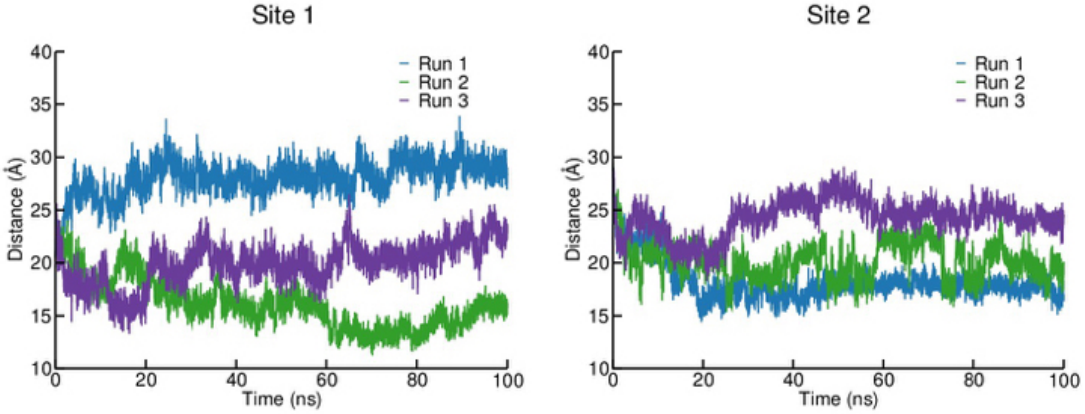


Figure B.2: The distance between NBDs of the homology model from the refined mouse crystal structure. This model exhibited the smallest distance between nucleotide binding sites, explaining the low RMSD.

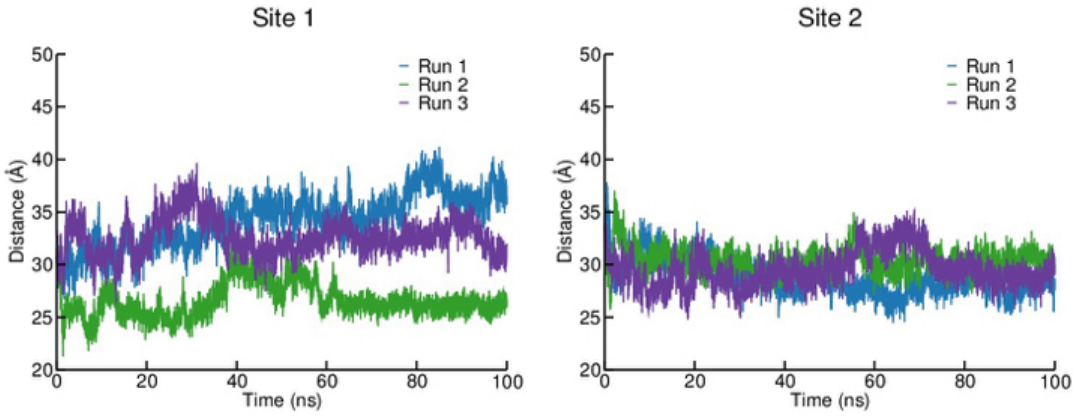


Figure B.3: The distance between NBDs of the homology model from the *C. merolae* crystal structure. The NBD distances showed no significant deviations from the start value.

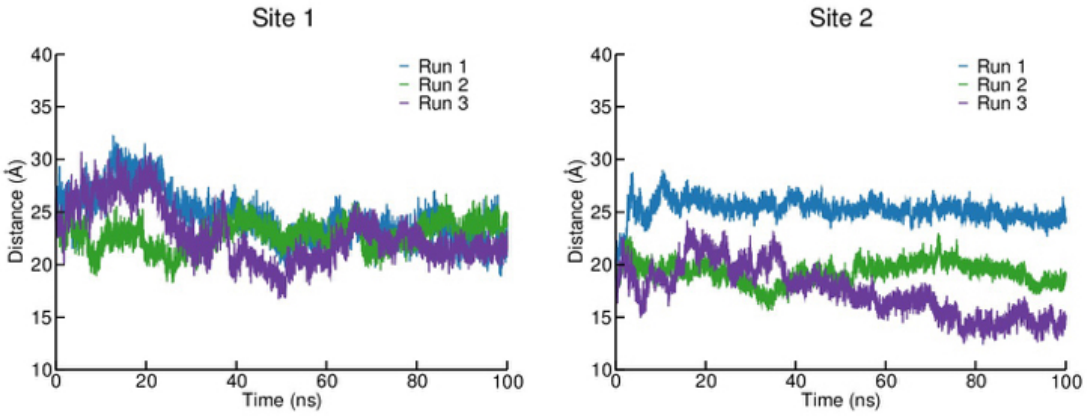


Figure B.4: The distance between NBDs of the homology model from the refined mouse crystal structure. Addition of ATP and Mg^{2+} lead to decrease of flexibility in distance of Site 1 - a behaviour directly opposite to that of apo P-gp. The asymmetric movements of NBDs have been previously observed [119].

C

Convergence of PMF profiles

The strength of cholesterol binding at the potential binding sites was determined from the PMF profiles for each binding site. In order to ensure adequate sampling, the distribution and overlap of the umbrella sampling windows was plotted as histograms. The spacing used in these simulations was sufficientFigure C.1. The convergence of PMF profiles was monitored through following the evolution of the profiles with time. Since the 0-100 ns segment showed the highest divergence, it was excluded from final analysisFigure C.2.

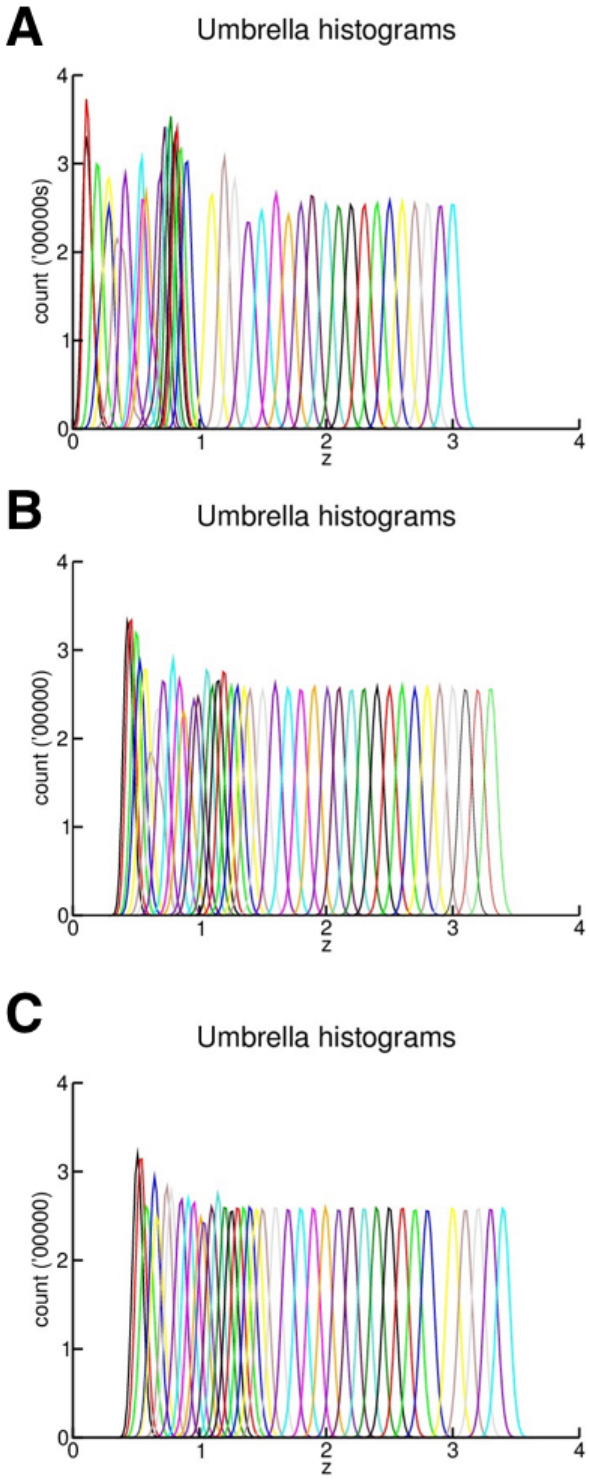


Figure C.1: Histograms of umbrella sampling windows for cholesterol binding sites at (A) TMH 10/12, (B) TMH 7/8 and (C) TMH 9/12.

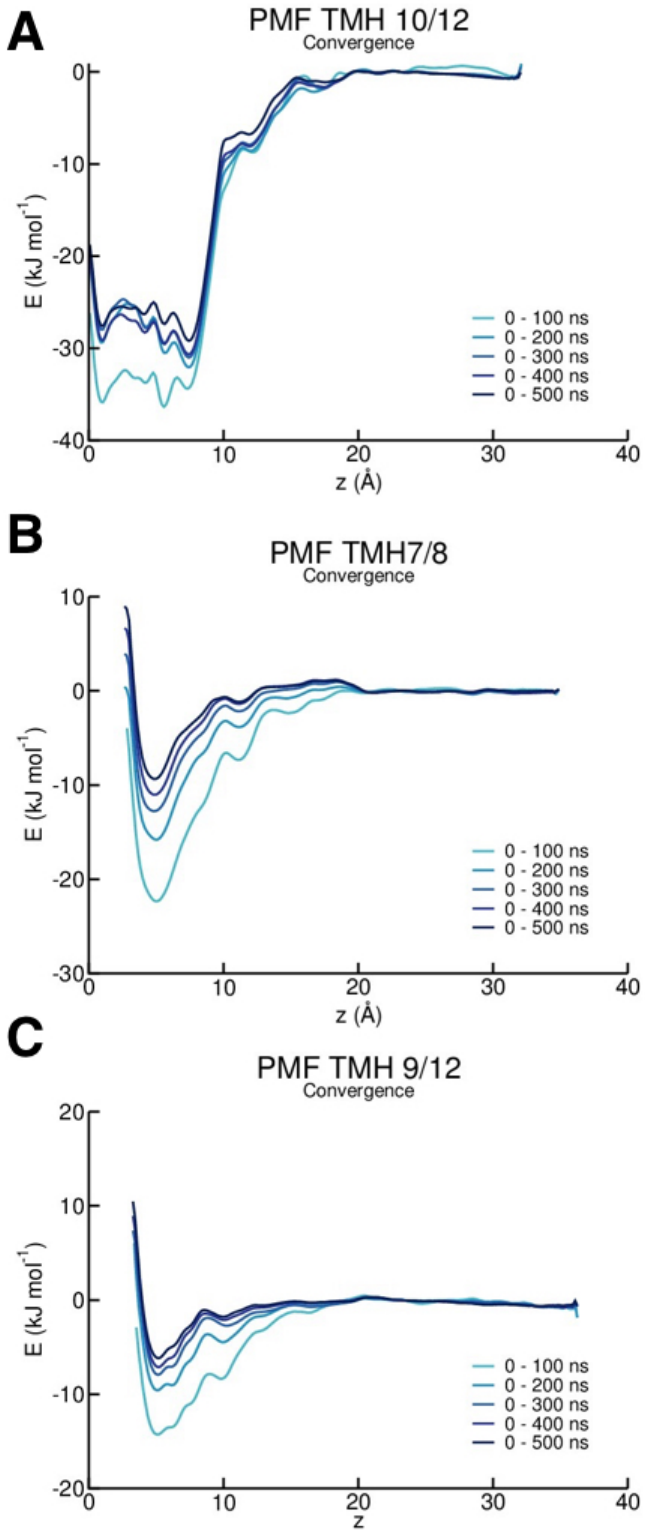


Figure C.2: Evolution of PMF profiles with time for cholesterol binding sites at (A) TMH 10/12, (B) TMH 7/8 and (C) TMH 9/12. In all profiles it could be observed that the 0-100 ns segment diverges the most, therefore, it was excluded from the final, converged PMF profile.

Bibliography

[1] J. Tanwar, S. Das, Z. Fatima, and S. Hameed, "Multidrug resistance: An emerging crisis," *Interdisciplinary Perspectives on Infectious Diseases*, vol. 2014, p. 541340, 2014.

[2] M. M. Gottesman, "Mechanisms of cancer drug resistance," *Annual Review of Medicine*, vol. 53, pp. 615–627, 2002.

[3] P. Carmeliet and R. K. Jain, "Angiogenesis in cancer and other diseases.," *Nature*, vol. 407, pp. 249–57, 2000.

[4] O. Trédan, C. M. Galmarini, K. Patel, and I. F. Tannock, "Drug resistance and the solid tumor microenvironment.," *Journal of the National Cancer Institute*, vol. 99, pp. 1441–54, 2007.

[5] M. H. Saier, C. V. Tran, and R. D. Barabote, "TCDB: the Transporter Classification Database for membrane transport protein analyses and information.," *Nucleic Acids Research*, vol. 34, pp. D181–6, 2006.

[6] G. Szakács, J. K. Paterson, J. A. Ludwig, C. Booth-Genthe, and M. M. Gottesman, "Targeting multidrug resistance in cancer.," *Nature Reviews Drug Discovery*, vol. 5, pp. 219–34, 2006.

[7] A. H. Schinkel and J. W. Jonker, "Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview.," *Advanced Drug Delivery Reviews*, vol. 55, pp. 3–29, 2003.

[8] D. C. Rees, E. Johnson, and O. Lewinson, "ABC transporters: the power to change.," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 218–27, 2009.

[9] M. Dean, A. Rzhetsky, and R. Allikmets, "The human ATP-binding cassette (ABC) transporter superfamily.," *Genome Research*, vol. 11, pp. 1156–66, 2001.

[10] E. M. Leslie, R. G. Deeley, and S. P. C. Cole, "Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense.," *Toxicology and Applied Pharmacology*, vol. 204, pp. 216–37, 2005.

[11] K. Dano, "Active outward transport of daunomycin in resistant Ehrlich ascites tumor cells.," *Biochimica et Biophysica Acta*, vol. 323, pp. 466–83, 1973.

[12] R. L. Juliano and V. Ling, "A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants.," *Biochimica et Biophysica Acta*, vol. 455, pp. 152–62, 1976.

[13] J. H. Gerlach, J. A. Endicott, P. F. Juranka, G. Henderson, F. Sarangi, K. L. Deuchars, and V. Ling, "Homology between P-glycoprotein and a bacterial haemolysin transport protein suggests a model for multidrug resistance.," *Nature*, vol. 324, pp. 485–9.

[14] F. Sharom, "The P-Glycoprotein Efflux Pump: How Does it Transport Drugs?," *Journal of Membrane Biology*, vol. 160, pp. 161–175, 1997.

[15] F. Thiebaut, T. Tsuruo, H. Hamada, M. M. Gottesman, I. Pastan, and M. C. Willingham, "Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues.," *Proceedings of the National Academy of Sciences*, vol. 84, pp. 7735–8, 1987.

[16] C. Cordon-Cardo, J. P. O'Brien, D. Casals, L. Rittman-Grauer, J. L. Biedler, M. R. Melamed, and J. R. Bertino, "Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites.," *Proceedings of the National Academy of Sciences*, vol. 86, pp. 695–8, 1989.

[17] B. Sarkadi, L. Homolya, G. Szakács, and A. Váradi, "Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system.," *Physiological Reviews*, vol. 86, pp. 1179–236, 2006.

[18] F. J. Sharom, *Multidrug Resistance Protein: P-Glycoprotein*, pp. 223–262. John Wiley Sons, Inc., 2006.

[19] F. E. O'Brien, G. Clarke, P. Fitzgerald, T. G. Dinan, B. T. Griffin, and J. F. Cryan, "Inhibition of P-glycoprotein enhances transport of imipramine across the blood-brain barrier: microdialysis studies in conscious freely moving rats.," *British Journal of Pharmacology*, vol. 166, pp. 1333–43, 2012.

[20] F. E. O'Brien, T. G. Dinan, B. T. Griffin, and J. F. Cryan, "Interactions between antidepressants and P-glycoprotein at the blood-brain barrier: clinical significance of in vitro and in vivo findings.," *British Journal of Pharmacology*, vol. 165, pp. 289–312, 2012.

[21] D. W. Boulton, C. L. DeVane, H. L. Liston, and J. S. Markowitz, "In vitro P-glycoprotein affinity for atypical and conventional antipsychotics.," *Life Sciences*, vol. 71, pp. 163–9, 2002.

[22] J. Kane, G. Honigfeld, J. Singer, and H. Meltzer, "Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine.," *Archives of General Psychiatry*, vol. 45, pp. 789–96, 1988.

[23] R. K. Solanki, P. Singh, and M. K. Swami, "Clozapine: Current perspective.," *Indian Journal of Psychiatry*, vol. 49, pp. 271–6, 2007.

[24] C. Chiu, M. C. Miller, R. Monahan, D. P. Osgood, E. G. Stopa, and G. D. Silverberg, "P-glycoprotein expression and amyloid accumulation in human aging and Alzheimer's disease: preliminary observations.," *Neurobiology of Aging*, vol. 36, pp. 2475–82, 2015.

[25] W. F. Ng, F. Sarangi, R. L. Zastawny, L. Veinot-Drebot, and V. Ling, "Identification of members of the P-glycoprotein multigene family.," *Molecular and Cellular Biology*, vol. 9, pp. 1224–32, 1989.

[26] S. Ruetz and P. Gros, "Phosphatidylcholine translocase: a physiological role for the mdr2 gene.," *Cell*, vol. 77, pp. 1071–81, 1994.

[27] Y. Romsicki and F. J. Sharom, "Phospholipid flippase activity of the reconstituted P-glycoprotein multidrug transporter.," *Biochemistry*, vol. 40, pp. 6937–47, 2001.

[28] P. D. W. Eckford and F. J. Sharom, "The reconstituted P-glycoprotein multidrug transporter is a flippase for glucosylceramide and other simple glycosphingolipids.," *The Biochemical Journal*, vol. 389, pp. 517–26, 2005.

[29] F. J. Sharom, "Complex Interplay between the P-Glycoprotein Multidrug Efflux Pump and the Membrane: Its Role in Modulating Protein Function.," *Frontiers in Oncology*, vol. 4, p. 41, 2014.

[30] F. J. Sharom, "Flipping and flopping–lipids on the move.," *IUBMB Life*, vol. 63, pp. 736–46, 2011.

[31] C. M. Frijters, C. J. Tuijn, R. Ottenhoff, B. N. Zegers, A. K. Groen, and R. P. Elferink, "The role of different P-glycoproteins in hepatobiliary secretion of fluorescently labeled short-chain phospholipids.," *Journal of Lipid Research*, vol. 40, pp. 1950–8, 1999.

[32] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch, and G. Chang, "Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding.," *Science*, vol. 323, pp. 1718–22, 2009.

[33] C. A. Hrycyna, L. E. Airan, U. A. Germann, S. V. Ambudkar, I. Pastan, and M. M. Gottesman, "Structural flexibility of the linker region of human P-glycoprotein permits ATP hydrolysis and drug transport.," *Biochemistry*, vol. 37, pp. 13660–73, 1998.

[34] T. Sato, A. Kodan, Y. Kimura, K. Ueda, T. Nakatsu, and H. Kato, "Functional role of the linker region in purified human P-glycoprotein.," *The FEBS Journal*, vol. 276, pp. 3504–16, 2009.

[35] R. J. P. Dawson and K. P. Locher, "Structure of a bacterial multidrug ABC transporter.," *Nature*, vol. 443, pp. 180–5, 2006.

[36] T. W. Loo, M. C. Bartlett, and D. M. Clarke, "Human P-glycoprotein contains a greasy ball-and-socket joint at the second transmission interface.," *The Journal of Biological Chemistry*, vol. 288, pp. 20326–33, 2013.

[37] G. F. Ames and H. Lecar, "ATP-dependent bacterial transporters and cystic fibrosis: analogy between channels and transporters.," *FASEB Journal*, vol. 6, pp. 2660–6, 1992.

[38] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.," *The EMBO Journal*, vol. 1, pp. 945–51, 1982.

[39] O. Jardetzky, "Simple allosteric model for membrane pumps.," *Nature*, vol. 211, pp. 969–70, 1966.

[40] R. J. P. Dawson and K. P. Locher, "Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP.," *FEBS Letters*, vol. 581, pp. 935–8, 2007.

[41] A. Ward, C. L. Reyes, J. Yu, C. B. Roth, and G. Chang, "Flexibility in the ABC transporter MsbA: Alternating access with a twist.," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 19005–10, 2007.

[42] M. Hohl, C. Briand, M. G. Grütter, and M. A. Seeger, "Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation," *Nature Structural & Molecular Biology*, vol. 19, pp. 395–402, 2012.

[43] M. S. Jin, M. L. Oldham, Q. Zhang, and J. Chen, "Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans.," *Nature*, vol. 490, pp. 566–9, 2012.

[44] A. B. Ward, P. Szewczyk, V. Grimard, C.-W. Lee, L. Martinez, R. Doshi, A. Caya, M. Villaluz, E. Pardon, C. Cregger, D. J. Swartz, P. G. Falson, I. L. Urbatsch, C. Govaerts, J. Steyaert, and G. Chang, "Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain.," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 13386–91, 2013.

[45] C. Shintre, A. C. W. Pike, Q. Li, J.-I. Kim, A. J. Barr, S. Goubin, L. Shrestha, J. Yang, G. Berridge, J. Ross, P. J. Stansfeld, M. S. P. Sansom, A. M. Edwards, C. Bountra, B. D. Marsden, F. von Delft, A. N. Bullock, O. Gileadi, N. Burgess-Brown, and E. P. Carpenter, "Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states.," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 9710–5, 2013.

[46] A. Kodan, T. Yamaguchi, T. Nakatsu, K. Sakiyama, C. J. Hipolito, A. Fujioka, R. Hirokane, K. Ikeguchi, B. Watanabe, J. Hiratake, Y. Kimura,

H. Suga, K. Ueda, and H. Kato *Proceedings of the National Academy of Sciences*, pp. 4049–54.

[47] J. Li, K. F. Jaimes, and S. G. Aller, "Refined structures of mouse P-glycoprotein.," *Protein Science*, vol. 23, pp. 34–46, 2014.

[48] H. G. Choudhury, Z. Tong, I. Mathavan, Y. Li, S. Iwata, S. Zirah, S. Rebuffat, H. W. van Veen, and K. Beis, "Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 9145–9150, 2014.

[49] M. Hohl, L. M. Hurlimann, S. Bohm, J. Schoppe, M. G. Grutter, E. Bordignon, and M. A. Seeger, "Structural basis for allosteric cross-talk between the asymmetric nucleotide binding sites of a heterodimeric ABC exporter," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 11025–11030, 2014.

[50] T. W. Loo and D. M. Clarke, "Drug rescue distinguishes between different structural models of human P-glycoprotein.," *Biochemistry*, vol. 52, pp. 7167–9, 2013.

[51] A. Moeller, S. Lee, H. Tao, J. Speir, G. Chang, I. Urbatsch, C. Potter, B. Carragher, and Q. Zhang, "Distinct Conformational Spectrum of Homologous Multidrug ABC Transporters," *Structure*, vol. 23, pp. 450–60, 2015.

[52] T. W. Loo and D. M. Clarke, "P-glycoprotein ATPase activity requires lipids to activate a switch at the first transmission interface.," *Biochemical and Biophysical Research Communications*, vol. 472, pp. 379–383, 2016.

[53] M. E. Zoghbi, R. S. Cooper, and G. A. Altenberg, "The Lipid Bilayer Modulates the Structure and Function of an ATP-binding Cassette Exporter," *Journal of Biological Chemistry*, vol. 291, pp. 4453–4461, 2016.

[54] M. M. Gottesman, S. V. Ambudkar, and D. Xia, "Structure of a multidrug transporter.," *Nature Biotechnology*, vol. 27, pp. 546–7, 2009.

[55] T. W. Loo, M. C. Bartlett, and D. M. Clarke, "Identification of residues in the drug translocation pathway of the human multidrug resistance P-glycoprotein by arginine mutagenesis.," *The Journal of Biological Chemistry*, vol. 284, pp. 24074–87, 2009.

[56] T. W. Loo and D. M. Clarke, "Identification of residues within the drug-binding domain of the human multidrug resistance P-glycoprotein by cysteine-scanning mutagenesis and reaction with dibromobimane.," *The Journal of Biological Chemistry*, vol. 275, pp. 39272–8, 2000.

[57] T. W. Loo and D. M. Clarke, "Defining the drug-binding site in the human multidrug resistance P-glycoprotein using a methanethiosulfonate analog of verapamil, MTS-verapamil.," *The Journal of Biological Chemistry*, vol. 276, pp. 14972–9, 2001.

[58] D. Mandal, K. Moitra, D. Ghosh, D. Xia, and S. Dey, "Evidence for modulatory sites at the lipid-protein interface of the human multidrug transporter P-glycoprotein.," *Biochemistry*, vol. 51, pp. 2852–66, 2012.

[59] E. E. Chufan, K. Kapoor, H.-M. Sim, S. Singh, T. T. Talele, S. R. Durell, and S. V. Ambudkar, "Multiple Transport-Active Binding Sites Are Available for a Single Substrate on Human P-Glycoprotein (ABCB1).," *PloS One*, vol. 8, p. e82463, 2013.

[60] C. Martin, G. Berridge, C. F. Higgins, P. Mistry, P. Charlton, and R. Callaghan, "Communication between Multiple Drug Binding Sites on P-glycoprotein," *Molecular Pharmacology*, vol. 58, pp. 624–632, 2000.

[61] S.-F. Zhou, "Structure, function and regulation of P-glycoprotein and its clinical relevance in drug disposition.," *Xenobiotica*, vol. 38, pp. 802–32, 2008.

[62] A. Seelig, "A general pattern for substrate recognition by P-glycoprotein," *European Journal of Biochemistry*, vol. 251, pp. 252–261, 1998.

[63] J. M. Zamora, H. L. Pearce, and W. T. Beck, "Physical-chemical properties shared by compounds that modulate multidrug resistance in human leukemic cells.," *Molecular Pharmacology*, vol. 33, pp. 454–62, 1988.

[64] K. Ueda, N. Okamura, M. Hirai, Y. Tanigawara, T. Saeki, N. Kioka, T. Komano, and R. Hori, "Human P-glycoprotein transports cortisol, aldosterone, and dexamethasone, but not progesterone.," *The Journal of Biological Chemistry*, vol. 267, pp. 24248–52, 1992.

[65] A. H. Schinkel, E. Wagenaar, C. A. Mol, and L. van Deemter, "P-glycoprotein in the blood-brain barrier of mice influences the brain penetration and pharmacological activity of many drugs.," *The Journal of Clinical Investigation*, vol. 97, pp. 2517–24, 1996.

[66] L. J. Bain, J. B. McLachlan, and G. A. LeBlanc, "Structure-activity relationships for xenobiotic transport substrates and inhibitory ligands of P-glycoprotein," *Environmental Health Perspectives*, vol. 105, pp. 812–818, 1997.

[67] R. B. Wang, C. L. Kuo, L. L. Lien, and E. J. Lien, "Structure-activity relationship: analyses of p-glycoprotein substrates and inhibitors," *Journal of Clinical Pharmacy and Therapeutics*, vol. 28, pp. 203–228, 2003.

[68] A. Seelig and E. Landwojtowicz, "Structure-activity relationship of P-glycoprotein substrates and modifiers.," *European Journal of Pharmaceutical Sciences*, vol. 12, pp. 31–40, 2000.

[69] F. Montanari and G. F. Ecker, "Prediction of drug-ABC-transporter interaction–Recent advances and future challenges.," *Advanced Drug Delivery Reviews*, vol. 86, pp. 17–26, 2015.

[70] J. Robert and C. Jarry, "Multidrug resistance reversal agents.," *Journal of Medicinal Chemistry*, vol. 46, pp. 4805–17, 2003.

[71] T. Saeki, K. Ueda, Y. Tanigawara, R. Hori, and T. Komano, "Human P-glycoprotein transports cyclosporin A and FK506.," *Journal of Biological Chemistry*, vol. 268, pp. 6077–80, 1993.

[72] A. H. Dantzig, K. L. Law, J. Cao, and J. J. Starling, "Reversal of multidrug resistance by the P-glycoprotein modulator, LY335979, from the bench to the clinic.," *Current Medicinal Chemistry*, vol. 8, pp. 39–50, 2001.

[73] T. W. Loo, "Blockage of Drug Resistance In Vitro by Disulfiram, a Drug Used to Treat Alcoholism," *Journal of the National Cancer Institute*, vol. 92, pp. 898–902, 2000.

[74] D. M. Woodcock, M. E. Linsenmeyer, G. Chojnowski, A. B. Kriegler, V. Nink, L. K. Webster, and W. H. Sawyer, "Reversal of multidrug resistance by surfactants.," *British Journal of Cancer*, vol. 66, pp. 62–8, 1992.

[75] T. Tsuruo, H. Iida, S. Tsukagoshi, and Y. Sakurai, "Overcoming of vincristine resistance in P388 leukemia in vivo and in vitro through enhanced cytotoxicity of vincristine and vinblastine by verapamil.," *Cancer Research*, vol. 41, pp. 1967–72, 1981.

[76] R. Krishna and L. D. Mayer, "Multidrug resistance (MDR) in cancer. Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs.," *European Journal of Pharmaceutical Sciences*, vol. 11, pp. 265–83, 2000.

[77] T. J. Lampidis, A. Krishan, L. Planas, and H. Tapiero, "Reversal of intrinsic resistance to adriamycin in normal cells by verapamil.," *Cancer Drug Delivery*, vol. 3, pp. 251–9, 1986.

[78] D. Boesch, K. Muller, A. Pourtier-Manzanedo, and F. Loor, "Restoration of daunomycin retention in multidrug-resistant P388 cells by submicromolar concentrations of SDZ PSC 833, a nonimmunosuppressive cyclosporin derivative.," *Experimental Cell Research*, vol. 196, pp. 26–32, 1991.

[79] R. Krishna, G. de Jong, and L. D. Mayer, "Pulsed exposure of SDZ PSC 833 to multidrug resistant P388/ADR and MCF7/ADR cells in the absence of

anticancer drugs can fully restore sensitivity to doxorubicin.," *Anticancer Research*, vol. 17, pp. 3329–34.

[80] M. A. Alaoui-Jamali, R. L. Schecter, Y. M. Rustum, M. G. Centurioni, S. Lehnert, and G. Batist, "In vivo reversal of doxorubicin resistance by a new tiapamil analog Ro11-2933.," *Journal of Pharmacology and Experimental Therapeutics*, vol. 264, pp. 1299–304, 1993.

[81] A. H. Dantzig, R. L. Shepard, J. Cao, K. L. Law, W. J. Ehlhardt, T. M. Baughman, T. F. Bumol, and J. J. Starling, "Reversal of P-glycoprotein-mediated multidrug resistance by a potent cyclopropyldibenzosuberane modulator, LY335979.," *Cancer Research*, vol. 56, pp. 4171–9, 1996.

[82] F. Hyafil, C. Vergely, P. Du Vignaud, and T. Grand-Perret, "In vitro and in vivo reversal of multidrug resistance by GF120918, an acridonecarboxamide derivative.," *Cancer Research*, vol. 53, pp. 4595–602, 1993.

[83] S. Mullin, N. Mani, and T. H. Grossman, "Inhibition of antibiotic efflux in bacteria by the novel multidrug resistance inhibitors biricodar (VX-710) and timcodar (VX-853).," *Antimicrobial Agents and Chemotherapy*, vol. 48, pp. 4171–6, 2004.

[84] P. Kannan, S. Telu, S. Shukla, S. V. Ambudkar, V. W. Pike, C. Halldin, M. M. Gottesman, R. B. Innis, and M. D. Hall, "The "specific" P-glycoprotein inhibitor Tariquidar is also a substrate and an inhibitor for breast cancer resistance protein (BCRP/ABCG2).," *ACS Chemical Neuroscience*, vol. 2, pp. 82–9, 2011.

[85] A. Tamaki, C. Ierano, G. Szakacs, R. W. Robey, and S. E. Bates, "The controversial role of ABC transporters in clinical oncology.," *Essays in Biochemistry*, vol. 50, pp. 209–32, 2011.

[86] H. M. Abdallah, A. M. Al-Abd, R. S. El-Dine, and A. M. El-Halawany, "P-glycoprotein inhibitors of natural origin as potential tumor chemo-sensitizers: A review.," *Journal of Advanced Research*, vol. 6, pp. 45–62, 2015.

[87] K. Michalak and O. Wesolowska, "Polyphenols counteract tumor cell chemoresistance conferred by multidrug resistance proteins.," *Anti-cancer Agents in Medicinal Chemistry*, vol. 12, pp. 880–90, 2012.

[88] P. M. Jones and A. M. George, "Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle.," *Proteins*, vol. 75, pp. 387–96, 2009.

[89] C. F. Higgins and K. J. Linton, "The ATP switch model for ABC transporters.," *Nature Structural & Molecular Biology*, vol. 11, pp. 918–26, 2004.

[90] A. M. George and P. M. Jones, "Perspectives on the structure-function of ABC transporters: the Switch and Constant Contact models.," *Progress in Biophysics and Molecular Biology*, vol. 109, pp. 95–107, 2012.

[91] O. Bársony, G. Szalóki, D. Türk, S. Tarapcsák, Z. Gutay-Tóth, Z. Bacsó, I. J. Holb, L. Székvölgyi, G. Szabó, L. Csanády, G. Szakács, and K. Goda, "A single active catalytic site is sufficient to promote transport in P-glycoprotein.," *Scientific Reports*, vol. 6, p. 24810, 2016.

[92] T. W. Loo, M. C. Bartlett, and D. M. Clarke, "Human P-glycoprotein is active when the two halves are clamped together in the closed conformation.," *Biochemical and Biophysical Research Communications*, vol. 395, pp. 436–40, 2010.

[93] C. F. Higgins and M. M. Gottesman, "Is the multidrug transporter a flippase?," *Trends in Biochemical Sciences*, vol. 17, pp. 18–21, 1992.

[94] F. J. Sharom, "The p-glycoprotein multidrug transporter," *Essays In Biochemistry*, vol. 50, pp. 161–178, 2011.

[95] A. B. Shapiro and V. Ling, "Extraction of Hoechst 33342 from the cytoplasmic leaflet of the plasma membrane by P-glycoprotein.," *European Journal of Biochemistry*, vol. 250, pp. 122–9, 1997.

[96] L. Domicevica and P. C. Biggin, "Homology modelling of human P-glycoprotein," *Biochemical Society Transactions*, vol. 43, pp. 952–958, 2015.

[97] J. Ferté, "Analysis of the tangled relationships between P-glycoprotein-mediated multidrug resistance and the lipid phase of the cell membrane," *European Journal of Biochemistry*, vol. 267, pp. 277–294, 2000.

[98] F. J. Sharom, X. Yu, J. W. Chu, and C. A. Doige, "Characterization of the ATPase activity of P-glycoprotein from multidrug-resistant Chinese hamster ovary cells.," *The Biochemical Journal*, vol. 308 (Pt 2, pp. 381–90, 1995.

[99] J. Marcoux, S. C. Wang, A. Politis, E. Reading, J. Ma, P. C. Biggin, M. Zhou, H. Tao, Q. Zhang, G. Chang, N. Morgner, and C. V. Robinson, "Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump.," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 9704–9, 2013.

[100] Y. Song, A. K. Kenworthy, and C. R. Sanders, "Cholesterol as a co-solvent and a ligand for membrane proteins.," *Protein Science*, vol. 23, pp. 1–22, 2014.

[101] A. Rothnie, D. Theron, L. Soceneantu, C. Martin, M. Traikia, G. Berridge, C. F. Higgins, P. F. Devaux, and R. Callaghan, "The importance of cholesterol in maintenance of P-glycoprotein activity and its membrane perturbing influence," *European Biophysics Journal*, vol. 30, pp. 430–442, 2001.

[102] P. D. W. Eckford and F. J. Sharom, "Interaction of the P-glycoprotein multidrug efflux pump with cholesterol: effects on ATPase activity, drug binding and transport.," *Biochemistry*, vol. 47, pp. 13686–98, 2008.

[103] Y. Kimura, N. Kioka, H. Kato, M. Matsuo, and K. Ueda, "Modulation of drug-stimulated ATPase activity of human MDR1/P-glycoprotein by cholesterol.," *The Biochemical Journal*, vol. 401, pp. 597–605, 2007.

[104] Y. Kimura, A. Kodan, M. Matsuo, and K. Ueda, "Cholesterol fill-in model: mechanism for substrate recognition by ABC proteins.," *Journal of Bioenergetics and Biomembranes*, vol. 39, pp. 447–52, 2007.

[105] A. T. Clay, P. Lu, and F. J. Sharom, "Interaction of the P-Glycoprotein Multidrug Transporter with Sterols.," *Biochemistry*, vol. 54, pp. 6586–97, 2015.

[106] A. T. Clay and F. J. Sharom, "Lipid bilayer properties control membrane partitioning, binding, and transport of p-glycoprotein substrates.," *Biochemistry*, vol. 52, pp. 343–54, 2013.

[107] L. Chen, Y. Li, H. Yu, L. Zhang, and T. Hou, "Computational models for predicting substrates or inhibitors of P-glycoprotein.," *Drug Discovery Today*, vol. 17, pp. 343–51, 2012.

[108] I. K. Pajeva, C. Globisch, and M. Wiese, "Comparison of the inward- and outward-open homology models and ligand binding of human P-glycoprotein.," *The FEBS Journal*, vol. 276, pp. 7016–26, 2009.

[109] A. Palmeira, M. H. Vasconcelos, A. Paiva, M. X. Fernandes, M. Pinto, and E. Sousa, "Dual inhibitors of P-glycoprotein and tumor cell growth: (Re)discovering thioxanthones," *Biochemical Pharmacology*, vol. 83, pp. 57–68, 2012.

[110] H. Yamaguchi, Y. Kidachi, K. Kamiie, T. Noshita, and H. Umetsu, "Homology modeling and structural analysis of human P-glycoprotein.," *Bioinformation*, vol. 8, pp. 1066–74, 2012.

[111] Y. Tajima, H. Nakagawa, A. Tamura, O. Kadioglu, K. Satake, Y. Mitani, H. Murase, L. O. Regasini, V. d. S. Bolzani, T. Ishikawa, G. Fricker, and T. Efferth, "Nitensidine A, a guanidine alkaloid from Pterogyne nitens, is a novel substrate for human ABC transporter ABCB1.," *Phytomedicine*, vol. 21, pp. 323–32, 2014.

[112] D. V. Singh, M. M. Godbole, and K. Misra, "A plausible explanation for enhanced bioavailability of P-gp substrates in presence of piperine: simulation for next generation of P-gp inhibitors.," *Journal of Molecular Modeling*, vol. 19, pp. 227–38, 2013.

[113] R. Prajapati, U. Singh, A. Patil, K. S. Khomane, P. Bagul, A. K. Bansal, and A. T. Sangamwar, "In silico model for P-glycoprotein substrate prediction: Insights from molecular dynamics and in vitro studies," *Journal of Computer-Aided Molecular Design*, vol. 27, pp. 347–363, 2013.

[114] X. Xu, R. Li, M. Ma, X. Wang, Y. Wang, and H. Zou, "Multidrug resistance protein P-glycoprotein does not recognize nanoparticle C60: experiment and modeling," *Soft Matter*, vol. 8, p. 2915, 2012.

[115] J. G. Wise, "Catalytic transitions in the human MDR1 P-glycoprotein drug binding sites.," *Biochemistry*, vol. 51, pp. 5125–41, 2012.

[116] F. K. Brewer, C. A. Follit, P. D. Vogel, and J. G. Wise, "In Silico Screening for Inhibitors of P-Glycoprotein That Target the Nucleotide Binding Domains," *Molecular Pharmacology*, vol. 86, pp. 716–726, 2014.

[117] I. K. Pajeva, M. Hanl, and M. Wiese, "Protein contacts and ligand binding in the inward-facing model of human P-glycoprotein.," *ChemMedChem*, vol. 8, pp. 748–62, 2013.

[118] M. Zeino, M. E. M. Saeed, O. Kadioglu, and T. Efferth, "The ability of molecular docking to unravel the controversy and challenges related to P-glycoprotein-a well-known, yet poorly understood drug transporter.," *Investigational New Drugs*, vol. 32, pp. 618–25, 2014.

[119] L. Pan and S. G. Aller, "Equilibrated atomic models of outward-facing P-glycoprotein and effect of ATP binding on structural dynamics.," *Scientific Reports*, vol. 5, p. 7880, 2015.

[120] J. Zhang, T. Sun, L. Liang, T. Wu, and Q. Wang, "Drug promiscuity of P-glycoprotein and its mechanism of interaction with paclitaxel and doxorubicin," *Soft Matter*, vol. 10, pp. 438–445, 2014.

[121] J. Zhang, D. Li, T. Sun, L. Liang, and Q. Wang, "Interaction of P-glycoprotein with anti-tumor drugs: the site, gate and pathway," *Soft Matter*, vol. 11, pp. 6633–6641, 2015.

[122] N. Subramanian, K. Condic-Jurkic, A. E. Mark, and M. L. O'Mara, "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*, vol. 55, pp. 1202–17, 2015.

[123] R. J. Ferreira, M.-J. U. Ferreira, and D. J. V. A. dos Santos, "Do Drugs Have Access to the P-Glycoprotein Drug-Binding Pocket through Gates?," *Journal of Chemical Theory and Computation*, vol. 11, pp. 4525–4529, 2015.

[124] P.-C. Wen, B. Verhalen, S. Wilkens, H. S. Mchaourab, and E. Tajkhorshid, "On the Origin of Large Flexibility of P-glycoprotein in the Inward-facing State," *Journal of Biological Chemistry*, vol. 288, pp. 19211–19220, 2013.

[125] S. K. Tripathi, S. K. Singh, P. Singh, P. Chellaperumal, K. K. Reddy, and C. Selvaraj, "Exploring the selectivity of a ligand complex with CDK2/CDK1: a molecular dynamics simulation approach," *Journal of Molecular Recognition*, vol. 25, pp. 504–512, 2012.

[126] J. G. Kirkwood, "Statistical Mechanics of Fluid Mixtures," *The Journal of Chemical Physics*, vol. 3, p. 300, 1935.

[127] R. W. Zwanzig, J. G. Kirkwood, I. Oppenheim, and B. J. Alder, "Statistical Mechanical Theory of Transport Processes. VII. The Coefficient of Thermal Conductivity of Monatomic Liquids," *The Journal of Chemical Physics*, vol. 22, p. 783, 1954.

[128] T. Hansson, J. Marelius, and J. Aqvist, "Ligand binding affinity prediction by linear interaction energy methods.," *Journal of Computer-Aided Molecular Design*, vol. 12, pp. 27–35, 1998.

[129] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations.," *Journal of Molecular Graphics & Modelling*, vol. 25, pp. 247–60, 2006.

[130] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models," *Accounts of Chemical Research*, vol. 33, pp. 889–897, 2000.

[131] R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, and D. E. Shaw, "Pathway and mechanism of drug binding to G-protein-coupled receptors.," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 13118–23, 2011.

[132] H. Grubmüller, B. Heymann, and P. Tavan, "Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force," *Science*, vol. 271, pp. 997–999, 1996.

[133] V. Limongelli, M. Bonomi, and M. Parrinello, "Funnel metadynamics as accurate binding free-energy method.," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 6358–63, 2013.

[134] S. K. Lüdemann, V. Lounnas, and R. C. Wade, "How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms.," *Journal of Molecular Biology*, vol. 303, pp. 797–811, 2000.

[135] S. Kalyaanamoorthy and Y.-P. P. Chen, "Modelling and enhanced molecular dynamics to steer structure-based drug discovery.," *Progress in Biophysics and Molecular Biology*, vol. 114, pp. 123–36, 2014.

[136] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," *Computer Physics Communications*, vol. 185, pp. 604–613, 2014.

[137] J. S. Patel, A. Berteotti, S. Ronsisvalle, W. Rocchia, and A. Cavalli, "Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5.," *Journal of Chemical Information and Modeling*, vol. 54, pp. 470–80, 2014.

[138] R. Yu, Q. Kaas, and D. J. Craik, "Delineation of the unbinding pathway of α-conotoxin ImI from the α7 nicotinic acetylcholine receptor.," *Journal of Physical Chemistry B*, vol. 116, pp. 6097–105, 2012.

[139] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems.," *Angewandte Chemie (International ed. in English)*, vol. 48, pp. 1198–229, 2009.

[140] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," *Journal of Chemical Theory and Computation*, vol. 4, pp. 435–447, 2008.

[141] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, vol. 117, pp. 5179–5197, 1995.

[142] A. R. Leach, *Molecular modelling: principles and application*. Prentice Hall, 2nd ed., 2001.

[143] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides ," *The Journal of Physical Chemistry B*, vol. 105, pp. 6474–6487, 2001.

[144] L. D. Schuler, X. Daura, and W. F. van Gunsteren, "An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase," *Journal of Computational Chemistry*, vol. 22, pp. 1205–1218, 2001.

[145] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and

M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins ," *The Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, 1998.

[146] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Structure, Function and Bioinformatics*, vol. 78, pp. 1950–1958, 2010.

[147] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic Validation of Protein Force Fields against Experimental Data," *PLoS ONE*, vol. 7, p. e32131, 2012.

[148] J. P. M. Jämbeck and A. P. Lyubartsev, "Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids.," *The Journal of Physical Chemistry B*, vol. 116, pp. 3164–79, 2012.

[149] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, "The MARTINI Coarse-Grained Force Field: Extension to Proteins," *Journal of Chemical Theory and Computation*, vol. 4, pp. 819–834, 2008.

[150] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: coarse grained model for biomolecular simulations.," *The Journal of Physical Chemistry B*, vol. 111, pp. 7812–24, 2007.

[151] S. J. Marrink, A. H. de Vries, and A. E. Mark, "Coarse Grained Model for Semiquantitative Lipid Simulations," *The Journal of Physical Chemistry B*, vol. 108, pp. 750–760, 2004.

[152] P. J. Stansfeld and M. S. Sansom, "From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations," *Journal of Chemical Theory and Computation*, vol. 7, pp. 1157–1166, 2011.

[153] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, "Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models," *Journal of Chemical Theory and Computation*, vol. 10, pp. 676–690, 2014.

[154] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, pp. 327–341, 1977.

[155] H. C. Andersen, "Rattle: A velocity version of the shake algorithm for molecular dynamics calculations," *Journal of Computational Physics*, vol. 52, pp. 24–34, 1983.

[156] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, pp. 1463–1472, 1997.

[157] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, p. 10089, 1993.

[158] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, p. 3684, 1984.

[159] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of Chemical Physics*, vol. 126, p. 014101, 2007.

[160] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, p. 7182, 1981.

[161] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 826–843, 2011.

[162] E. Darve and A. Pohorille, "Calculating free energies using average force," *The Journal of Chemical Physics*, vol. 115, p. 9169, 2001.

[163] G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics*, vol. 23, pp. 187–199, 1977.

[164] B. Roux, "The calculation of the potential of mean force using computer simulations," *Computer Physics Communications*, vol. 91, pp. 275–282, 1995.

[165] T. Paramo, S. M. Tomasio, K. L. Irvine, C. E. Bryant, and P. J. Bond, "Energetics of Endotoxin Recognition in the Toll-Like Receptor 4 Innate Immune Response.," *Scientific Reports*, vol. 5, p. 17997, 2015.

[166] F. B. Naughton, A. C. Kalli, and M. S. P. Sansom, "Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of GRP1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers," *The Journal of Physical Chemistry Letters*, vol. 7, pp. 1219–1224, 2016.

[167] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *Journal of Computational Chemistry*, vol. 13, pp. 1011–1021, 1992.

[168] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints.," *Journal of Molecular Biology*, vol. 234, pp. 779–815, 1993.

[169] P. Benkert, M. Biasini, and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models.," *Bioinformatics (Oxford, England)*, vol. 27, pp. 343–50, 2011.

[170] R. J. Ferreira, M.-J. U. Ferreira, and D. J. V. A. dos Santos, "Reversing cancer multidrug resistance: insights into the efflux by ABC transports from in silico studies," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 5, pp. 27–55, 2015.

[171] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Clustal W and Clustal X version 2.0.," *Bioinformatics (Oxford, England)*, vol. 23, pp. 2947–8, 2007.

[172] G. Vriend, "WHAT IF: a molecular modeling and drug design program.," *Journal of Molecular Graphics*, vol. 8, pp. 52–6, 29, 1990.

[173] M.-Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures.," *Protein Science*, vol. 15, pp. 2507–24, 2006.

[174] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics.," *Journal of Molecular Graphics*, vol. 14, pp. 33–8, 27–8, 1996.

[175] J. Ma and P. C. Biggin, "Substrate versus inhibitor dynamics of P-glycoprotein.," *Proteins*, vol. 81, pp. 1653–68, 2013.

[176] W. DeLano, "The PyMOL Molecular Graphics System (2002) on World Wide Web," 2002.

[177] K. L. Meagher, L. T. Redman, and H. A. Carlson, "Development of polyphosphate parameters for use with the AMBER force field.," *Journal of Computational Chemistry*, vol. 24, pp. 1016–25, 2003.

[178] M. L. O'Mara and A. E. Mark, "Structural Characterization of Two Metastable ATP-Bound States of P-Glycoprotein.," *PloS One*, vol. 9, p. e91916, 2014.

[179] G. A. Altenberg, C. G. Vanoye, J. K. Horton, and L. Reuss, "Unidirectional fluxes of rhodamine 123 in multidrug-resistant cells: evidence against direct drug extrusion from the plasma membrane.," *Proceedings of the National Academy of Sciences*, vol. 91, pp. 4654–4657, 1994.

[180] D. Branduardi, F. L. Gervasio, and M. Parrinello, "From A to B in free energy space.," *The Journal of Chemical Physics*, vol. 126, p. 054103, 2007.

[181] J. Ma, L. Domicevica, J. R. Schnell, and P. C. Biggin, "Position and orientational preferences of drug-like compounds in lipid membranes: a computational and NMR approach.," *Physical Chemistry Chemical Physics : PCCP*, vol. 17, pp. 19766–76, 2015.

[182] E. Jefferys, Z. A. Sands, J. Shi, M. S. P. Sansom, and P. W. Fowler, "Alchembed: A Computational Method for Incorporating Multiple Proteins into Complex Lipid Geometries.," *Journal of Chemical Theory and Computation*, vol. 11, pp. 2743–2754, 2015.

[183] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field.," *Journal of Computational Chemistry*, vol. 25, pp. 1157–74, 2004.

[184] A. W. Sousa da Silva and W. F. Vranken, "ACPYPE - AnteChamber PYthon Parser interfacE.," *BMC Research Notes*, vol. 5, p. 367, 2012.

[185] L. Domicevica, T. Paramo, T. Newport, and P. C. Biggin, "Lintools - dynamic representation of protein-ligand interactions," 2016.

[186] T. W. Loo, M. C. Bartlett, and D. M. Clarke, "Methanethiosulfonate derivatives of rhodamine and verapamil activate human P-glycoprotein at different sites.," *The Journal of Biological Chemistry*, vol. 278, pp. 50136–41, 2003.

[187] T. W. Loo and D. M. Clarke, "Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein.," *The Journal of Biological Chemistry*, vol. 277, pp. 44332–8, 2002.

[188] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.," *Journal of Computational Chemistry*, vol. 31, pp. 455–61, 2010.

[189] G. Jones, P. Willett, and R. C. Glen, "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation," *Journal of Molecular Biology*, vol. 245, pp. 43–53, 1995.

[190] O. Korb, T. Stützle, and T. E. Exner, "Empirical scoring functions for advanced protein-ligand docking with PLANTS.," *Journal of Chemical Information and Modeling*, vol. 49, pp. 84–96, 2009.

[191] D. R. Houston and M. D. Walkinshaw, "Consensus docking: improving the reliability of docking in a virtual screening context.," *Journal of Chemical Information and Modeling*, vol. 53, pp. 384–90, 2013.

[192] T. Tuccinardi, G. Poli, V. Romboli, A. Giordano, and A. Martinelli, "Extensive consensus docking evaluation for ligand pose prediction and virtual

screening studies.," *Journal of Chemical Information and Modeling*, vol. 54, pp. 2980–6, 2014.

[193] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique," *Journal of Chemical Theory and Computation*, vol. 1, pp. 1304–1311, 2005.

[194] A. Madadkar-Sobhani and V. Guallar, "PELE web server: atomistic study of biomolecular systems at your fingertips.," *Nucleic Acids Research*, vol. 41, pp. W322–8, 2013.

[195] T. W. Loo and D. M. Clarke, "Identification of the distance between the homologous halves of P-glycoprotein that triggers the high/low ATPase activity switch.," *The Journal of Biological Chemistry*, vol. 289, pp. 8484–92, 2014.

[196] S. D. Campbell, K. J. Regina, and E. D. Kharasch, "Significance of Lipid Composition in a Blood-Brain Barrier-Mimetic PAMPA Assay," *Journal of Biomolecular Screening*, vol. 19, pp. 437–444, 2013.

[197] G. van Meer and A. I. P. M. de Kroon, "Lipid map of the mammalian cell.," *Journal of Cell Science*, vol. 124, pp. 5–8, 2011.

[198] H. Koldsø, D. Shorthouse, J. Hélie, and M. S. P. Sansom, "Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers.," *PLoS Computational Biology*, vol. 10, p. e1003911, 2014.

[199] H. Koldsø and M. S. P. Sansom, "Organization and Dynamics of Receptor Proteins in a Plasma Membrane.," *Journal of the American Chemical Society*, vol. 137, pp. 14694–704, 2015.

[200] T. Reddy, D. Shorthouse, D. L. Parton, E. Jefferys, P. W. Fowler, M. Chavent, M. Baaden, and M. S. P. Sansom, "Nothing to sneeze at: a dynamic and integrative computational model of an influenza A virion.," *Structure*, vol. 23, pp. 584–97, 2015.

[201] G. R. Kneller, K. Baczynski, and M. Pasenkiewicz-Gierula, "Communication: consistent picture of lateral subdiffusion in lipid bilayers: molecular dynamics simulation and exact results.," *The Journal of Chemical Physics*, vol. 135, p. 141105, 2011.

[202] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, vol. 9, pp. 90–95, 2007.

[203] P. Schwille, U. Haupts, S. Maiti, and W. W. Webb, "Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation," *Biophysical Journal*, vol. 77, pp. 2251–2265, 1999.

[204] H. I. Ingólfsson, M. N. Melo, F. J. van Eerden, C. Arnarez, C. A. Lopez, T. A. Wassenaar, X. Periole, A. H. de Vries, D. P. Tieleman, and S. J. Marrink, "Lipid organization of the plasma membrane.," *Journal of the American Chemical Society*, vol. 136, pp. 14554–9, 2014.

[205] J.-H. Jeon, H. M.-S. Monne, M. Javanainen, and R. Metzler, "Anomalous diffusion of phospholipids and cholesterols in a lipid bilayer and its origins.," *Physical Review Letters*, vol. 109, p. 188103, 2012.

[206] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink, "Improved Parameters for the Martini Coarse-Grained Protein Force Field.," *Journal of Chemical Theory and Computation*, vol. 9, pp. 687–97, 2013.

[207] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations.," *Journal of Computational Chemistry*, vol. 32, pp. 2319–27, 2011.

[208] H. Koldsø and M. S. P. Sansom, "Local lipid reorganization by a transmembrane protein domain," *Journal of Physical Chemistry Letters*, vol. 3, pp. 3498–3502, 2012.

[209] C. Bechara, A. Nöll, N. Morgner, M. T. Degiacomi, R. Tampé, and C. V. Robinson, "A subset of annular lipids is linked to the flippase activity of an ABC transporter.," *Nature Chemistry*, vol. 7, pp. 255–62, 2015.

[210] D. Loe and F. Sharom, "Interaction of multidrug-resistant Chinese hamster ovary cells with amphiphiles," *British Journal of Cancer*, vol. 68, pp. 342–351, 1993.

[211] H. Palsdottir and C. Hunte, "Lipids in membrane protein structures.," *Biochimica et Biophysica Acta*, vol. 1666, pp. 2–18, 2004.

[212] A. G. Abulrob and M. Gumbleton, "Transport of phosphatidylcholine in MDR3-negative epithelial cell lines via drug-induced MDR1 P-glycoprotein.," *Biochemical and Biophysical Research Communications*, vol. 262, pp. 121–6, 1999.

[213] I. Bosch, K. Dunussi-Joannopoulos, R. L. Wu, S. T. Furlong, and J. Croop, "Phosphatidylcholine and phosphatidylethanolamine behave as substrates of the human MDR1 P-glycoprotein.," *Biochemistry*, vol. 36, pp. 5685–94, 1997.

[214] A. van Helvoort, M. Giudici, M. Thielemans, and G. van Meer, "Transport of sphingomyelin to the cell surface is inhibited by brefeldin A and in mitosis, where C6-NBD-sphingomyelin is translocated across the plasma membrane by a multidrug transporter activity," *Journal of Cell Science*, vol. 110, pp. 75–83, 1997.

[215] A. Pohl, H. Lage, P. Müller, T. Pomorski, and A. Herrmann, "Transport of phosphatidylserine via MDR1 (multidrug resistance 1)P-glycoprotein in a human gastric carcinoma cell line.," *The Biochemical Journal*, vol. 365, pp. 259–68, 2002.

[216] C. J. Baier, J. Fantini, and F. J. Barrantes, "Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor.," *Scientific Reports*, vol. 1, p. 69, 2011.

[217] D. A. Antunes, D. Devaurs, and L. E. Kavraki, "Understanding the challenges of protein flexibility in drug design," *Expert Opinion on Drug Discovery*, vol. 10, pp. 1301–1313, 2015.

[218] R. A. Copeland, D. L. Pompliano, and T. D. Meek, "Drug-target residence time and its implications for lead optimization," *Nature Reviews Drug Discovery*, vol. 5, pp. 730–739, 2006.

[219] S. Núñez, J. Venhorst, and C. G. Kruse, "Target-drug interactions: first principles and their application to drug discovery," *Drug Discovery Today*, vol. 17, pp. 10–22, 2012.

[220] J. Mortier, C. Rakers, M. Bermudez, M. S. Murgueitio, S. Riniker, and G. Wolber, "The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes.," *Drug Discovery Today*, vol. 20, pp. 686–702, 2015.

[221] F. Samsudin, J. Parker, M. Sansom, S. Newstead, and P. Fowler, "Accurate Prediction of Ligand Affinities for a Proton-Dependent Oligopeptide Transporter," *Cell Chemical Biology*, vol. 23, no. 2, pp. 299–309, 2016.

[222] A. J. Clark, P. Tiwary, K. Borrelli, S. Feng, E. B. Miller, R. Abel, R. A. Friesner, and B. J. Berne, "Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations," *Journal of Chemical Theory and Computation*, vol. 12, pp. 2990–2998, 2016.

[223] A. C. Wallace, R. A. Laskowski, and J. M. Thornton, "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.," *Protein Engineering*, vol. 8, pp. 127–34, 1995.

[224] R. A. Laskowski and M. B. Swindells, "LigPlot+: multiple ligand-protein interaction diagrams for drug discovery.," *Journal of Chemical Information and Modeling*, vol. 51, pp. 2778–86, 2011.

[225] K. Stierand, P. C. Maass, and M. Rarey, "Molecular complexes at a glance: automated generation of two-dimensional complex diagrams.," *Bioinformatics (Oxford, England)*, vol. 22, pp. 1710–6, 2006.

[226] A. M. Clark and P. Labute, "2D depiction of protein-ligand complexes," *Journal of Chemical Information and Modeling*, vol. 47, pp. 1933–1944, 2007.

[227] S. Caboche, "LeView: automatic and interactive generation of 2D diagrams for biomacromolecule/ligand interactions.," *Journal of Cheminformatics*, vol. 5, p. 40, 2013.

[228] Python Software Foundation, "Python language reference, version 2.7.11."

[229] R. McGibbon, K. Beauchamp, M. Harrigan, C. Klein, J. Swails, C. Hernández, C. Schwantes, L.-P. Wang, T. Lane, and V. Pande, "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories," *Biophysical Journal*, vol. 109, pp. 1528–1532, 2015.

[230] G. Landrum, "RDKit: open source chemoinformatics."

[231] S. Gillies, A. Bierbaum, K. Lautaportti, and O. Tonnhoffer, "Shapely: manipulation and analysis of geometrical objects."

[232] K. H. Wang, A. Penmatsa, and E. Gouaux, "Neurotransmitter and psychostimulant recognition by the dopamine transporter.," *Nature*, vol. 521, pp. 322–7, 2015.

[233] Travis CI GmbH, "Travis CI."

[234] Lemur Heavy Industries, "Coveralls."

[235] D. Merkel, "Docker: lightweight Linux containers for consistent development and deployment," *Linux Journal*, vol. 2014, p. 2, 2014.

[236] P. Filippakopoulos and S. Knapp, "Targeting bromodomains: epigenetic readers of lysine acetylation," *Nature Reviews Drug Discovery*, vol. 13, pp. 337–356, 2014.

[237] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin, "Accurate calculation of the absolute free energy of binding for drug molecules," *Chemical Science*, vol. 7, pp. 207–218, 2016.

[238] P. Filippakopoulos, S. Picaud, O. Fedorov, M. Keller, M. Wrobel, O. Morgenstern, F. Bracher, and S. Knapp, "Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family," *Bioorganic & Medicinal Chemistry*, vol. 20, pp. 1878–1886, 2012.

[239] D. S. Hewings, O. Fedorov, P. Filippakopoulos, S. Martin, S. Picaud, A. Tumber, C. Wells, M. M. Olcina, K. Freeman, A. Gill, A. J. Ritchie, D. W. Sheppard, A. J. Russell, E. M. Hammond, S. Knapp, P. E. Brennan, and S. J. Conway, "Optimization of 3,5-Dimethylisoxazole Derivatives as Potent Bromodomain Ligands," *Journal of Medicinal Chemistry*, vol. 56, pp. 3217–3227, 2013.

[240] S. Picaud, C. Wells, I. Felletar, D. Brotherton, S. Martin, P. Savitsky, B. Diez-Dacal, M. Philpott, C. Bountra, H. Lingard, O. Fedorov, S. Muller, P. E. Brennan, S. Knapp, and P. Filippakopoulos, "RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 19754–19759, 2013.

[241] P. Ciceri, S. Müller, A. O'Mahony, O. Fedorov, P. Filippakopoulos, J. P. Hunt, E. A. Lasater, G. Pallares, S. Picaud, C. Wells, S. Martin, L. M. Wodicka, N. P. Shah, D. K. Treiber, and S. Knapp, "Dual kinase-bromodomain inhibitors for rationally designed polypharmacology," *Nature Chemical Biology*, vol. 10, pp. 305–312, 2014.

[242] E. E. Chufan, H. M. Sim, and S. V. Ambudkar, "Molecular Basis of the Polyspecificity of P-Glycoprotein (ABCB1). Recent Biochemical and Structural Studies.," *Advances in Cancer Research*, vol. 125, pp. 71–96, 2015.