

# Characterization of drug translocation pathway in multidrug resistance proteins: P-glycoprotein and ABCG2 using structural approaches

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health.

2021

**Talha Shafi**

School of Biological Sciences

# List of Content

List of Content .....	2
List of Figures.....	8
List of Tables .....	11
List of Abbreviations.....	12
Abstract .....	14
DECLARATION .....	15
COPYRIGHT STATEMENT .....	15
Acknowledgements .....	16
Chapter 1: Introduction .....	17
1.1    ABC transporter Family.....	17
1.1.1    Multidrug resistance-related protein .....	18
1.1.1.1    P-glycoprotein.....	19
1.1.1.1.1    Structure of P-glycoprotein.....	19
1.1.1.1.2    Drug binding cavity .....	24
1.1.1.2    ABCG2 .....	29
1.1.1.2.1    Structure of ABCG2: .....	30
1.1.1.2.2    Binding cavity: .....	33
1.2    Recombinant expression of membrane proteins: .....	37
1.3    Purification of membrane protein: .....	38
1.3.1    Solubilisation of membrane:.....	39
1.3.2    Protein Isolation:.....	40

1.4	Structural techniques to study membrane proteins .....	41
1.4.1	X-ray crystallography .....	41
1.4.2	Cryo-Electron Microscopy (cryo-EM).....	42
1.4.2.1	Grid Preparation .....	43
1.4.2.2	Single-particle analysis .....	44
1.4.3	Small Angle X-ray Scattering .....	45
	Aims and objectives:.....	49
	Material and Methods .....	50
2.1	Materials and Equipments .....	50
2.1.1	Chemicals .....	50
2.1.2	Equipment and Consumable.....	51
2.2	Buffer recipes.....	52
2.3	Methodology.....	54
2.3.1	Expression and purifications:.....	54
2.3.1.1	Constructs:.....	54
2.3.1.2	Expression of mP-gp .....	54
2.3.1.3	Expression of human P-gp .....	55
2.3.1.4	Microsome preparation.....	57
2.3.1.5	Immobilized metal affinity chromatography:.....	58
2.3.1.6	Size Exclusion Chromatography (SEC): .....	59
2.3.1.7	SDS-PAGE: .....	59

2.3.1.8	Protein quantification:.....	60
2.3.2	Protein characterization: .....	61
2.3.2.1	Western blotting:.....	61
2.3.2.2	Thermal stability assay: .....	62
2.3.2.3	Glycosylation assay:.....	62
2.3.2.4	Drug binding assay:.....	63
2.3.2.5	ATP binding assay: .....	64
2.3.3	Structural studies:.....	64
2.3.3.1	Negative staining electron microscopy: .....	64
2.3.3.2	Small Angle X-ray Scattering (SAXS) analysis:.....	65
2.3.3.3	Cryo-electron microscopy (cryo-EM).....	65
2.3.3.3.1	Grid preparation.....	65
2.3.3.3.2	Data acquisition: .....	66
2.3.3.3.3	Single-particle analysis:.....	66
2.3.4	Computational studies:.....	67
2.3.4.1	Computational Studies on P-glycoprotein (ABCB1) .....	67
2.3.4.1.1	Ivacaftor-P-gp interactions studies: .....	67
2.3.4.1.2	Virtual screening: .....	68
2.3.4.2	Computational studies on ABCBG2 (BCRP) .....	68
2.3.4.2.1	Docking of kinase inhibitors with ABCG2:.....	68



2.3.4.2.2	Pharmacophore modelling: .....	69
2.3.4.2.3	Residue mutagenesis analysis:.....	69
2.3.4.2.4	Virtual screening .....	70
Chapter 3: Protein Expression and Purification:.....		71
3.1	Expression of mouse P-glycoprotein .....	71
3.2	Expression of human P-glycoprotein:.....	78
3.2.1	Time course experiment .....	78
3.2.2	Microsome preparation and purification: .....	82
3.3	Characterization of protein:.....	86
3.3.1	Post-translational modifications:.....	86
3.3.2	ATP binding assay: .....	88
3.4	Discussion and conclusions.....	89
Chapter 4: Small-Angle X-ray Scattering (SAXS) studies on P-glycoprotein catalytic states		92
4.1	Preliminary analysis: .....	93
4.2	Experiment 1: P-glycoprotein in apo and ADP trapped state:.....	94
4.3	Experiment 2: Concentration-dependent SAXS profiles.....	101
4.4	Dataset reduction: .....	107
4.5	Negative stain EM and D <sub>max</sub> analysis: .....	108
4.6	Discussion and Conclusions .....	111
Chapter 5: Structural studies on ligand-protein interaction of ivacaftor with P-glycoprotein		114

5.1	Preliminary analysis on ivacaftor binding with Mouse P-glycoprotein .....	114
5.2	Docking of ivacaftor with mouse P-glycoprotein (4KSB): .....	117
5.3	Docking of ivacaftor with human P-glycoprotein .....	121
5.4	Structural studies on ivacaftor binding:.....	122
5.5	Docking of Ivacaftor with in-house mouse P-glycoprotein .....	127
5.6	Structural studies with a molar excess of ivacaftor .....	128
5.7	Structural studies on ivacaftor binding with human P-glycoprotein:.....	136
5.8	Virtual screening: .....	137
5.8.1	Optimization of cut-off threshold .....	137
5.8.2	Library screening.....	140
5.9	Discussion and conclusions:.....	142
	Chapter 6, ABCG2: mechanistic in-sights to ligand-protein interaction.....	145
6.1	Preliminary analysis with Kinase inhibitors .....	147
6.2	Docking of Lapatinib derivatives:.....	149
6.3	Evolutionary conservation of Phe439:.....	153
6.4	Docking of lapatinib with $\Delta$ F439 .....	154
6.5	Docking of structurally different substrates: .....	156
6.6	CeTSA analysis and the ABCG2 'clamp' model: .....	158
6.7	Pharmacophore modelling: .....	158
6.7.1	Model evaluations .....	161
6.8	Virtual screening .....	163

6.8.1	Kinase class: .....	165
6.8.2	Non-kinase class.....	166
6.9	Discussion and conclusions:.....	168
	Chapter 7 General Conclusions and Future Prospects .....	171
	Future Prospect: .....	179
	References .....	181
	Appendix .....	206

Word count : 35,900

## List of Figures

<i>Figure 1.1. Structural representation of P-glycoprotein.....</i>	<i>20</i>
<i>Figure 1.2 Outward-facing conformations of ABC transporters. ....</i>	<i>24</i>
<i>Figure 1.3. Poly-specificity of inhibitor binding in human P-glycoprotein.....</i>	<i>25</i>
<i>Figure 1.4. Diagrammatic representation of QZ59 in the binding cavity of P-glycoprotein. ...</i>	<i>27</i>
<i>Figure 1.5. Venn diagram of residues involved in ligand-protein interaction of peptide-cyclic compounds with mouse P-glycoprotein. ....</i>	<i>28</i>
<i>Figure 1.6. Venn diagram of ligand-protein interaction of structurally diverse compounds with mouse P-glycoprotein. ....</i>	<i>29</i>
<i>Figure 1.7. Structural representation of ABCG2. ....</i>	<i>31</i>
<i>Figure 1.8. Diagrammatic representation of ABCG2 binding cavity. ....</i>	<i>34</i>
<i>Figure 1.9. Venn diagram of ligand-protein interaction of structurally diverse compounds with ABCG2.....</i>	<i>35</i>
<i>Figure 1.10. Poly-specificity of ligand protein interaction in ABCG2.....</i>	<i>36</i>
<i>Figure 1.11 Graphical representation of EM structures reported in the last two decades. ....</i>	<i>43</i>
<i>Figure 1.12. Diagrammatic representation of SAXS instrument. ....</i>	<i>46</i>
<i>Figure 3.1. Validation of mouse P-glycoprotein expression in Pichia pastoris.....</i>	<i>71</i>
<i>Figure 3.2. Nickel affinity purification of mouse P-glycoprotein. ....</i>	<i>74</i>
<i>Figure 3.3. SEC purification of mouse P-glycoprotein. ....</i>	<i>76</i>
<i>Figure 3.4 Schematic diagram of protein quantification using ImageJ Package. ....</i>	<i>77</i>
<i>Figure 3.5 Time course experiment of human P-glycoprotein expressed in yeast. ....</i>	<i>80</i>
<i>Figure 3.6. Graphical representation time course experiment.....</i>	<i>81</i>
<i>Figure 3.7. Ni affinity chromatography of human P-glycoprotein.. ....</i>	<i>83</i>
<i>Figure 3.8. Size exclusion chromatography of Human P-glycoprotein.....</i>	<i>85</i>
<i>Figure 3.9. Glycosylation of human and mouse P-glycoprotein.....</i>	<i>87</i>
<i>Figure 3.10. Graphical representation of ATP-TNP binding assay used for nucleotide binding. ....</i>	<i>89</i>
<i>Figure 4.1 Theoretical scattering profile of ABC transporters generated using FoXS server. ..</i>	<i>94</i>
<i>Figure 4.2 Scattering profiles of P-glycoprotein in apo and ADP trapped state. ....</i>	<i>95</i>

<i>Figure 4.3 Guinier analysis of mouse P-glycoprotein in apo and ADP trapped states. ....</i>	<i>97</i>
<i>Figure 4.4 Guinier analysis of human P-glycoprotein in apo and ADP trapped states. ....</i>	<i>98</i>
<i>Figure 4.5 Comparison of experimental scattering profile and theoretically calculated profiles. ....</i>	<i>99</i>
<i>Figure 4.6 Representative scattering profile of mouse P-glycoprotein, ....</i>	<i>102</i>
<i>Figure 4.7 Extrapolated scattering profiles of human P-gp in Apo and ADP trapped state. .</i>	<i>103</i>
<i>Figure 4.8 Radius of gyration estimate from individual frames from the mouse apo dataset. ....</i>	<i>105</i>
<i>Figure 4.9 Radius of gyration estimated from individual frames from the mouse ADP trapped dataset. ....</i>	<i>106</i>
<i>Figure 4.10 Representative graphical illustration of concentration dependent effect on Rg. ....</i>	<i>107</i>
<i>Figure 4.11 Representative negative stain images of the mouse P-glycoprotein. ....</i>	<i>109</i>
<i>Figure 4.12. Representative negative stain images of the human P-glycoprotein.. ....</i>	<i>110</i>
<i>Figure 5.1 Thermal unfolding assay for ivacaftor-mP-gp interaction. ....</i>	<i>115</i>
<i>Figure 5.2 ABCB1 affinity for ivacaftor measured by thermostability changes ....</i>	<i>116</i>
<i>Figure 5.3 Ligand-protein interaction of ivacaftor with mP-gp (4KSB). ....</i>	<i>119</i>
<i>Figure 5.4 Ligand-protein interaction analysis of consensus conformation 34. ....</i>	<i>120</i>
<i>Figure 5.5 ligand-protein interaction of ivacaftor with human P-glycoprotein. ....</i>	<i>122</i>
<i>Figure 5.6 Differentiation of five 3D classes. ....</i>	<i>123</i>
<i>Figure 5.7 3D representation of Ivacaftor binding site. ....</i>	<i>125</i>
<i>Figure 5.8 Re-arrangement of side chains in mP-gp. ....</i>	<i>126</i>
<i>Figure 5.9. Ligand protein interaction analysis of ivacaftor with in-house P-glycoprotein. ..</i>	<i>128</i>
<i>Figure 5.10 Single particle analysis of mouse P-glycoprotein with ivacaftor. ....</i>	<i>130</i>
<i>Figure 5.11 2D classification and particle selection. ....</i>	<i>131</i>
<i>Figure 5.12 2D and 3D classification of mouse P-glycoprotein from 7434 particles. ....</i>	<i>133</i>
<i>Figure 5.13 Resolution estimation from Fourier shell correlation. ....</i>	<i>133</i>
<i>Figure 5.14 Electron density map of mouse P-glycoprotein with ivacaftor fitted with atomic resolution mouse P-glycoprotein. ....</i>	<i>135</i>
<i>Figure 5.15. Single-particle analysis of human P-glycoprotein.. ....</i>	<i>136</i>
<i>Figure 5.16 Graphical representation of Optimisation of cut off threshold for ABCB1 virtual screening. ....</i>	<i>139</i>

<i>Figure 5.17 Histogram representation of major drug hits according to their targets. ....</i>	<i>141</i>
<i>Figure 6.1 Diagrammatic representation of ABCG2 binding cavity. ....</i>	<i>146</i>
<i>Figure 6.2. Ligand protein interaction of lapatinib analogues with ABCG2. ....</i>	<i>148</i>
<i>Figure 6.3. Protein ligand interactions fingerprint of lapatinib analogues. ....</i>	<i>151</i>
<i>Figure 6.4 Graphical representation of ligand-protein interaction. ....</i>	<i>152</i>
<i>Figure 6.5. Sequence analysis ABCG2. ....</i>	<i>154</i>
<i>Figure 6.6. Box and whisker plot representation of binding scores of top ten conformations of lapatinib with different Phe439 mutant models.. ....</i>	<i>155</i>
<i>Figure 6.7. Diagrammatic representation of structurally different ligands in the ABCG2 binding cavity and a depiction of the F439 interaction. ....</i>	<i>157</i>
<i>Figure 6.8 Spatial orientation of pharmacophore models. ....</i>	<i>160</i>
<i>Figure 6.9 Graphical representation of scoring distribution in substrates and non-substrate with ChemPLP and GOLD scoring functions. ....</i>	<i>164</i>
<i>Figure 6.10 Histogram representation of hits from the kinase class according to their targets.....</i>	<i>165</i>
<i>Figure 6.11. Histogram representation of major drug hits from the non-kinase class of inhibitors.....</i>	<i>167</i>
<i>Figure 6.12 Graphical representation of 'clamp' model. ....</i>	<i>169</i>
<i>Figure 7.1. Diagrammatic representations of putative ABC substrate translocations models. ....</i>	<i>172</i>
<i>Figure 7.2 Diagrammatic representation of NBD re-arrangement in drug binding state. ....</i>	<i>174</i>
<i>Figure 7.3 Diagrammatic representation of single residue clamp models.....</i>	<i>178</i>

## List of Tables

<i>Table 1.1. List of the available mouse and human P-glycoprotein structure and their states in the Protein database. ....</i>	<i>21</i>
<i>Table 1.2 List of available structures of ABCG2 analogues from the protein data bank. ....</i>	<i>32</i>
<i>Table 4.1 Summary of SAXS derived structural parameters.....</i>	<i>100</i>
<i>Table 4.2 Structural parameters obtained from extrapolated scattering profiles.....</i>	<i>104</i>
<i>Table 4.3 Structural parameter of extrapolated profiles from the reduced dataset. ....</i>	<i>108</i>
<i>Table 5.1 Shows the solution number and their corresponding scoring values of the top ten ivacaftor conformers with mouse P-glycoprotein (PDBID: 4KSB).....</i>	<i>118</i>
<i>Table 5.2 Top ten solutions of ivacaftor with human P-glycoprotein (7A69).....</i>	<i>121</i>
<i>Table 5.3 Distances between two NBD domains measure between C<math>\alpha</math> of Thr625 and C<math>\alpha</math> of Met1266 .....</i>	<i>124</i>
<i>Table 5.4. Shows the solution number and their corresponding scoring values of the top ten ivacaftor conformers with the in-house mouse P-glycoprotein model. ....</i>	<i>127</i>
<i>Table 6.1 Features present in pharmacophore models.....</i>	<i>159</i>
<i>Table 6.2 Prediction efficiency of pharmacophore models against the known dataset. ....</i>	<i>162</i>

## List of Abbreviations

$\mu\text{M}$	MicroMolar
2D	Two dimensional
3D	Three dimensional
Å	Angstrom
ABC	ATP binding cassette
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BSA	Bovine Serum Albumin
CeTSA	Cellular Thermal Shift Assay
CFTR	Cystic fibrosis transmembrane conductance regulator
CTF	Contrast Transfer Function
CV	Column volume
DDM	n-Dodecyl $\beta$ -D-maltoside
$e^-$	Electrons
EM	Electron Microscopy
g	Gravitational force
GDa	Giga dalton
GFP	Green Fluorescent Protein
His	histidine
hP-gp	Human P-glycoprotein
IMAC	Immobilized Metal Affinity Chromatography
$K_d$	Dissociation constant
kDa	kilo Dalton
M	Molar
MDa	Mega Dalton
MDFF	Molecular Dynamic Flexible Field
MDR	Multidrug Resistance
mM	milliMolar



mP-gp	Mouse P-glycoprotein
NBD	Nucleotide Binding Domain
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PLIF	Protein Ligand Interaction Fingerprint
P <sub>i</sub>	Inorganic Phosphate
Rpm	Rotation per minute
SDS	Sodium Dodecyl Sulphate
SLS	Static Light Scattering
T <sub>m</sub>	Melting temperature
TMD	Transmembrane Domain
UV	Ultra violet
v/v	volume by volume
w/v	weight by volume
YNB	Yeast Nitrogenous Base

## Abstract

P-glycoprotein (ABCB1) and ABCG2 are transmembrane, ATP dependent exporters that transport structurally diverse compounds across the membrane. They are present at the blood tissue barriers and control substrate movement with roles in xenobiotic protections. Additionally, overexpression of P-glycoprotein and ABCG2 in tumour cells is associated with efflux of drugs out of cells, a phenomenon called multidrug resistance. Although both proteins have been studied for decades, the drug translocation pathways still require understanding. Towards this goal, a mixed approach of computational docking and thermostability assays was used. Further, cryo-EM was used to study the interaction of ivacaftor with mouse P-glycoprotein. Our results suggested that ivacaftor bind to mouse P-glycoprotein with a  $K_d$  of  $1\mu\text{M}$ . further, docking results suggested the binding of ivacaftor to a novel binding site between TM9, 10 and 11 near the entry portal. This site was also observed in cryo-EM data of mouse P-glycoprotein with ivacaftor. Small-angle X-ray scattering analysis showed aggregation as a confounding factor in the scattering profiles. however, mouse P-glycoprotein showed improved protein properties in the presence of substrates. Docking of lapatinib analogues with ABCG2 showed binding of the kinase inhibitors between the paired and evolutionary conserved residue Phe439. Mutation of this residue in collaborating laboratory to a lipophilic residue reduces the affinity of the molecule for ABCG2. However, mutation to an aromatic residue did not affect the binding which suggests the key role of  $\pi$ - $\pi$  interaction in ligand-protein interaction. This observation was supported with pharmacophore models where the presence of aromatic moiety was common. Virtual screening of drug libraries showed kinase inhibitors as a major class of interactors for both P-glycoprotein and ABCG2. Ultimately, this project sheds light on the ligand-protein interaction of drugs with P-glycoprotein and ABCG2. Also, we proposed a novel 'single residue clamp' model as a mechanism of drug translocation in ABCG2.

## DECLARATION

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institutes of learning.

## COPYRIGHT STATEMENT

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example, graphs and tables (“reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialization of this thesis, the Copyright and any Intellectual Property and/or reproduction described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in The University’s policy on presentation of Theses.

## Acknowledgements

I want to express gratitude to my supervisor Prof. Robert C. Ford for his constant support and guidance. Indeed his dedication was a source of inspiration for me throughout this tenure.

I also want to thank The University of Manchester for providing me with the platform and facilities to conduct this project, especially, my co-supervisor Dr Steve Prince for his valuable suggestions and discussions. which helped me in understanding the fundamentals of Small-angle X-ray scattering. My grateful thank extends to Richard Collins and Aleksandr Mironov who assisted me in electron microscopy and data collection. I also want to acknowledge Diamond Light Source (DLS) Ltd (Harwell, UK) for providing the opportunity to work on the B21 beamline.

I would like to thank my colleagues and friends at the Ford lab, Nopnithi Thonghin, Alessandro Barbieri, Jack Clews, Eleanor Martin, Xin Meng and Iqra Younus for providing the best working atmosphere. Especially, Nopnithi Thonghin and Alessandro Barbieri for sharing their valuable experiences in protein purification and cryo-EM analysis.

I also want to express my gratitude to John D. Schuetz's and his Lab, at St. Jude Children Hospital Memphis, for sharing data and their insights on ABCG2. Especially Tomoka Gose for performing the wet lab work on ABCG2.

Furthermore, I want to acknowledge the Punjab Education Endowment Fund (PEEF) for financing my study at The University of Manchester.

At last but not least I want to say thanks to my family who continuous backing and encouragement is a source of motivation.

## Chapter 1: Introduction

Membrane proteins are one of the largest protein families. About 30% of the human genome encodes membrane proteins (Wallin and von Heijne 1998). Typically, these membrane proteins are classified into two broad categories i.e. (i) peripheral membrane proteins (PMPs) and (ii) Integral membrane proteins (IMPs). PMPs are tethered to the membrane surfaces and have many functions including acting as signal-transducing proteins that receive the signal from upstream sources and transduce it to the next in a cascade. However, integral membrane proteins span the cellular membranes and thus create a communication portal between the intracellular and extracellular space. Integral membrane proteins demonstrate a prodigious structural diversity ranging from monotopic helical structures to polytopic channels and can have  $\beta$ -barrels and helical structures (Lodish H, Berk A et al. 2000). Due to this diversity in structure and function, they have been the centre of attention for molecular biologists. Furthermore, about 50% of pharmaceutical drugs target membrane proteins emphasizing their clinical importance (Terstappen and Reggiani 2001, Arinaminpathy, Khurana et al. 2009). Structural and biochemical characterization of these proteins can be vital. It can not only help us to understand the underlying molecular mechanism of action but can also aid the structure-based drug discovery process, which can eventually lead to better and effective drug development.

### 1.1 ABC transporter Family

ATP binding cassette (ABC) transporter proteins form one of the largest families of transporter proteins (Dassa and Bouige 2001, Verrier, Bird et al. 2008). Until now 48 isotypes of ABC transporters have been identified in humans and have been classified into 6 groups (Dean M, Rando R et al.). The ABCA group encodes a few of the largest ABC proteins and are usually

involved in the transport of lipids across the membrane. ABCB subfamily consists of four full transporters and seven half transporters that are involved in substrate regulation. ABCC is a subfamily of 12 genes involved in diverse spectrum functions including ion transport, toxin effluxing transporters and ion channels. ABCG consist of six half transporters that are mostly involved in the regulation of substrate movement across the barriers. ABCD is involved in the transport of amino acids, polypeptides, lipids and carbohydrates, and xenobiotics. ABCE and ABCF are non-classical ABC transporters proteins that have the characteristic NBDs but lack the TM domains making them unlikely to involve in transportation (Dean, Rzhetsky et al. 2001). All these ABC proteins take part in vital physiological functions. Mutations in these transporters are linked to several clinical conditions (Linton 2011).

Despite their role in normal physiological functions at blood tissue barriers and substrate regulation, ABC transporters are notorious for their role in drug resistance (Choi 2005). Their ability to interact with structurally diverse molecules and efflux mechanisms has become the Achilles' heel of chemotherapy. Cancer cells overexpress these transporters to efflux structurally diverse molecules, therefore, reducing the intracellular drug bioavailability (Gottesman M.M and I Pastan 1988, Sodani, Patel et al. 2012). This phenomenon of acquired resistance is referred to as multidrug resistance (MDR) in cancer. Subsequently, the proteins that are involved in MDR are referred to as MDR related proteins (Robey, Pluchino et al. 2018).

#### 1.1.1 Multidrug resistance-related protein

In recent times, a plethora of studies have established the role of certain ABC transporters in MDR in cancer (Robey, Pluchino et al. 2018). P-glycoprotein (ABCB1/MDR1), breast cancer resistance protein (BCRP/ABCG2) and multidrug resistance-related protein 1 (ABCC1/MRP1) are ABC proteins that have been extensively studied for their role in multidrug resistance.

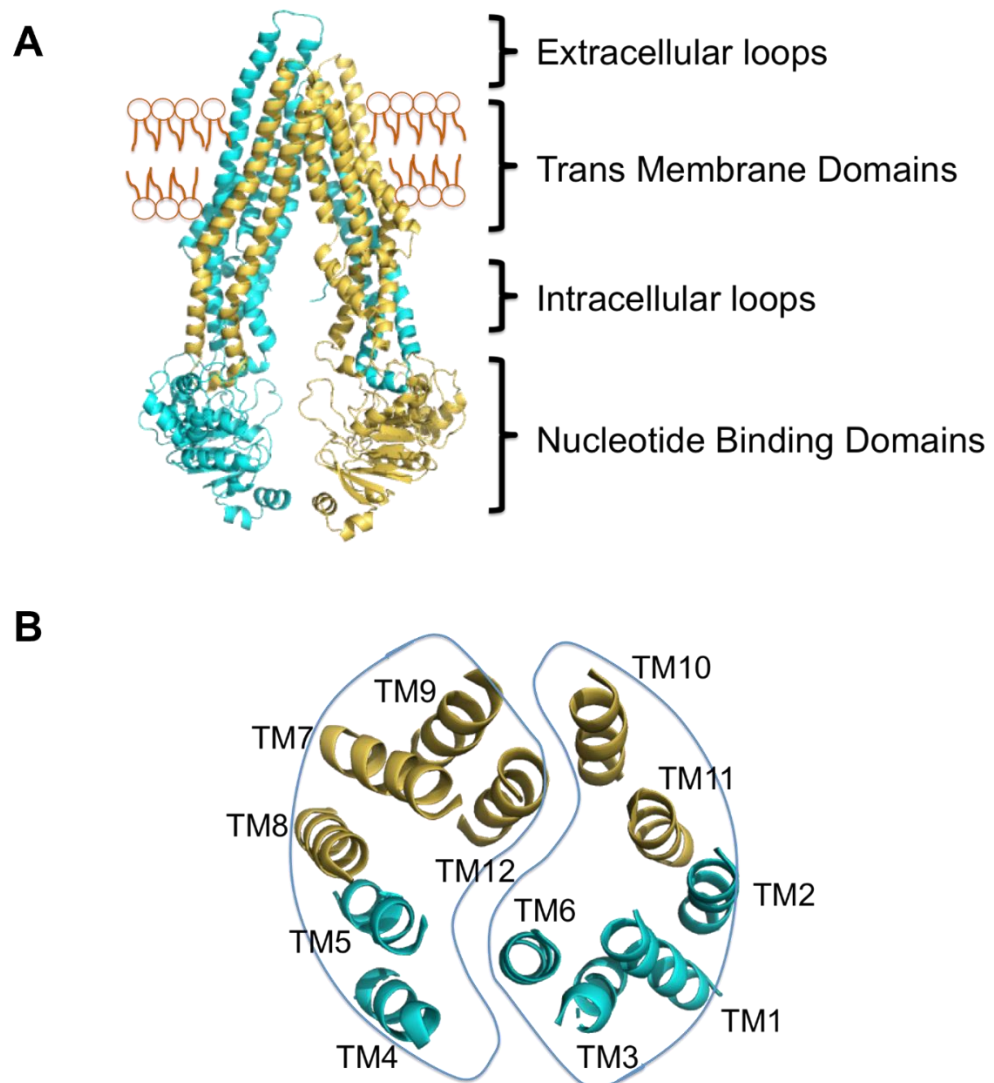
#### 1.1.1.1 *P-glycoprotein*

P-glycoprotein (P-gp/ABCB1) is a 170 kDa protein that was first identified as a permeability factor in a multidrug-resistant Chinese hamster ovary cell line (Juliano and Ling 1976). Later, (Roninson, Chin et al. 1986) reported a similar gene from the human carcinoma cell line which was later classified as ABCB1 (Dean, Rzhetsky et al. 2001). Since its discovery, P-glycoprotein has gained great attention as a multidrug resistance protein (MDR1). Under normal circumstances P-glycoprotein is localized on the apical membrane of enterocytes (von Richter, Burk et al. 2004), the canalicular membrane of hepatocytes (Hoffmaster, Turncliff et al. 2004), the blood-brain barrier (Karssen, Meijer et al. 2004), and blood placenta barrier (Molsa, Heikkinen et al. 2005), implicating its role in xenobiotic protection and regulation of substrate across the barriers. However, this xenobiotic protection and overexpression of P-glycoprotein in cancerous cells have been implicated to be a double-edged sword that results in a drug-resistant phenotype (Juliano and Ling 1976, Gottesman M.M and I Pastan 1988).

##### 1.1.1.1.1 Structure of P-glycoprotein

In the last decade, several structures of human and mouse P-glycoprotein have been reported (see table 1.1, Page 20). In general, all structures have delineated two transmembrane domains (TMDs), two nucleotide-binding domains (NBDs), three extracellular loops (ECL) on each TMD and two intracellular loops (ICL) (see figure 1.1). The TMDs are connected to NBDs through intracellular loop regions. Each TMD consists of 6  $\alpha$ -helical structures that contribute to the substrate-binding sites between the TMDs (Li, Jaimes et al. 2014, Szewczyk, Tao et al. 2015). The NBDs are the powerhouse of the proteins consisting of conserved Walker A and B regions (Kim and Chen 2018, Szöllősi, Chiba et al. 2020). The conserved Walker regions along with the conserved (A-,H-,Q-loops) motifs make the nucleotide-binding and hydrolysis site (Azzaria, Schurr et al. 1989, Kim and Chen 2018). The mutational analysis of MRP1 intracellular

loops suggests that these ICLs can also perform a crucial function in the catalytic cycle of P-glycoprotein (Ren, Furukawa et al. 2006).



*Figure 1.1. Structural representation of P-glycoprotein. (A) Anterior view of representative P-glycoprotein (PDB ID: 4M1M) labelled with its different domains. (B) Top view of the TMDs (PDB ID: 4M1M). TMD1 consists of TM1, 2, 3, 6, 10 and 11 whereas TMD2 consists of TM4, 5, 7, 8, 9, and 12. The two TMDs demonstrate the pseudo-symmetric arrangement along the central axis.*



Table 1.1. List of the available mouse and human P-glycoprotein structure and their states in the Protein database.

PDB ID	Origin	State	Ligand	Adjuvant	Mutations	Expression system	Methods	Resolution (Å)	Reference
7OTG	M	IF	Ivacaftor			Pichia Pastoris	Cryo-EM	5.4	(Barbieri, Thonghin et al. 2021)
7OTI	M	IF	-					4.2	
7A65	H	IF		Nanodisc + MRK16 Fab		HEK293T cells		3.9	(Nosol, Romane et al. 2020)
7A69	H	IF	Vincristine					3.2	
7A6C	H	IF	Elacridar x2					3.6	
7A6E	H	IF	Tariquidar x2					3.6	
7A6F	H	IF	Zosuquidar x2					3.5	
6UJN	M	IF	BDE-100		C952A	Pichia Pastoris	X-ray Diffraction	3.98	(Le, Harvey et al. 2020)
6UJP	M	IF	BDE-100		F979A, C952A			3.98	
6UJR	M	IF	BDE-100		F724A, C952A			4.1	
6UJS	M	IF	BDE-100 x2		F728A, C952A			4.17	
6UJT	M	IF	BDE-100 x2		Y303A, C952A			4.17	
6UJW	M	IF	BDE-100 x2		Y306A, C952A			4.15	
6QEE	HM	IF	Zosuquidar x2 +CHL, 3PE, NAG	Nanodisc + UIC Fab	S893A	HEK293T cells	Cryo-EM	3.9	(Alam, Kowal et al. 2019)
6QEX	H	IF	Taxol +CHL, 3PE, NAG					3.6	
6Q81	M	IF	-			Pichia Pastoris		7.9	(Thonghin, Collins et al. 2018)
6GDI	M	IF	ADP x 2					“	

6FN1	HM	IF	Zosuquidar x2 + NAG	UIC2 Fab		HEK293T cells		3.58	(Alam, Küng et al. 2018)
6FN4	HM	IF	NAG + 3PE						
6C0V	H	OF/OO	ATP x2		E556Q, E1201Q	Bac-HEK293T		3.4	(Kim and Chen 2018)
5KOY	M	IF	ATP x1		-34aa linker deleted	Pichia Pastoris	X-ray Diffraction	3.85	(Esser, Zhou et al. 2017)
5KO2	M	IF	-					3.3	
5KPD	M	IF	-					3.35	
5KPJ	M	IF	-					3.5	
5KPI	M	IF	-					4.01	
4XWK	M	IF	BDE-100			Pichia pastoris		3.5	(Nicklisch, Rees et al. 2016)
4Q9L	M	IF	QZ-Phe x2			Pichia Pastoris		3.8	(Szewczyk, Tao et al. 2015)
4Q9K	M	IF	QZ-Leu					3.8	
4Q9J	M	IF	QZ-Val x2					3.6	
4Q9I	M	IF	QZ-Ala x2					3.78	
4Q9H	M	IF	-					3.4	
4LSG	M	IF	-			Pichia Pastoris		3.8	N/A
4M1M	M	IF	-		(N83Q, N87Q, N90Q, C952A)	Pichia pastoris		3.8	(Li, Jaimes et al. 2014)
4M2S	M	IF	QZ59-RRR					4.4	
4M2T	M	IF	QZ59-SSS					4.35	
4KSB	M	IF	-			Pichia Pastoris		3.8	(Ward, Szewczyk et al. 2013)
4KSC	M	IF	-					4	
4KSD	M	IF	-					4.1	

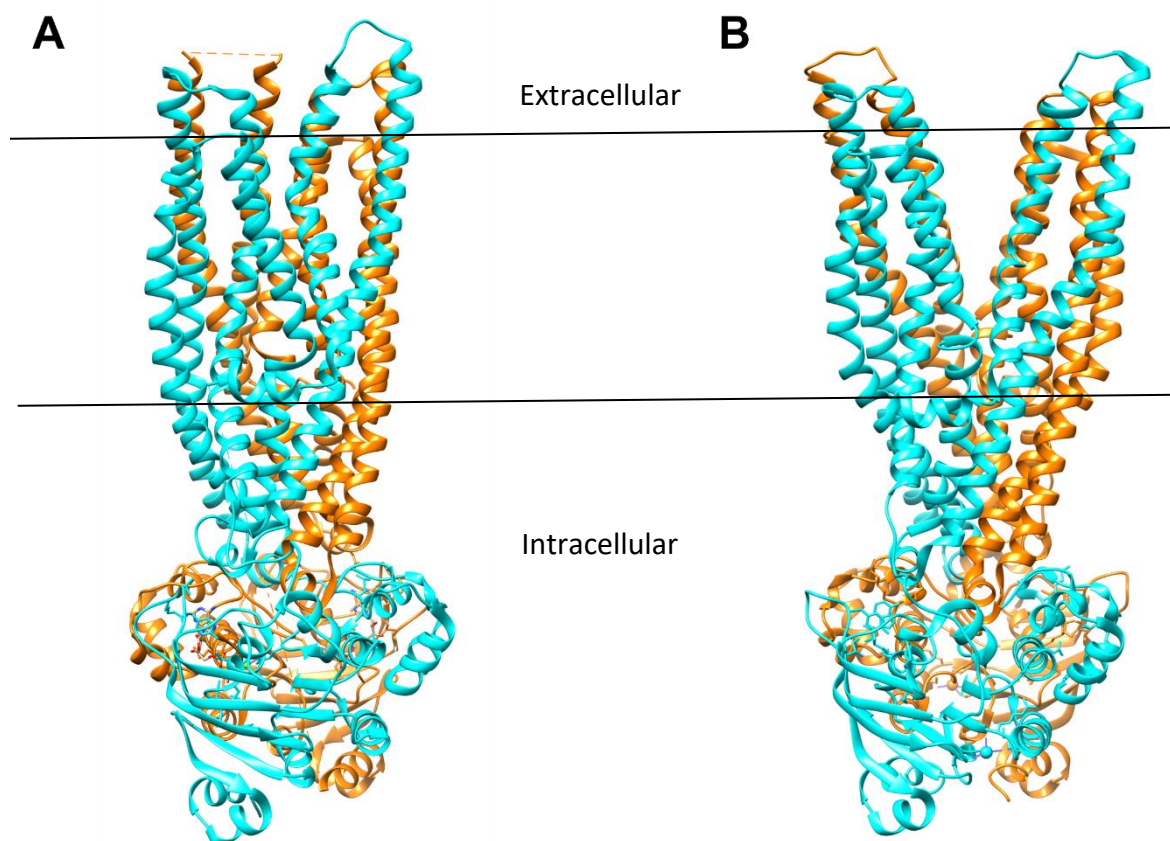
M=mouse, H=human, HM=human/mouse chimera, CHL=cholesterol, PE=Phosphatidylethanolamine, NAG=n acetyl glycosamine

IF=inward facing, OF/OO=outward facing/outward occluded

N/A= not any

Current structures of P-glycoprotein from mouse and humans show different conformational states of P-glycoprotein (Li, Jaimes et al. 2014, Kim and Chen 2018, Thonghin, Collins et al. 2018). The inward-facing conformations (apo and drug bound) have delineated basic two-fold pseudosymmetry of P-gp with an inverted “V” like conformation. The two transmembrane domains are closed at the extracellular region and the two NBDs are far apart. The distance between the two NBDs can vary depending on the state of the protein (Szewczyk, Tao et al. 2015). The substrates may enter the binding cavity via entry portals between the TM helix 4/6 and 10/12 (see Figure 1.1B).

Recently, the structure of human P-gp in the nucleotide bound outward-facing/outward-occluded state has been reported (Kim and Chen 2018). The structure shows a distinct compact state in which the TMDs are arranged in a very closed manner reducing the inter-domain solvent accessible area (see Figure 1.2A). The structure suggests the strong interaction between the intracellular loops that leads to rearrangement of TMDs resulting in an outward-open conformation. The structure suggests a squeeze pump mechanism of the protein where the intracellular region squeezes by generating strong inter-domain interactions through nucleotide binding. This squeeze probably produces the force to export substrate out of the cell. This conformational model is in contrast with the other outward-facing model of SAV1866 (Dawson and Locher 2007) which suggest a floral twist like opening of the extracellular region. Furthermore, the opening of the extracellular loops in SAV1866 are widely separated as compared to the human P-gp state (see Figure 1.2B). This can lead to an argument that either human P-gp has a different molecular mode of action or that the structure is in an open occluded state rather than the completely outward open state.

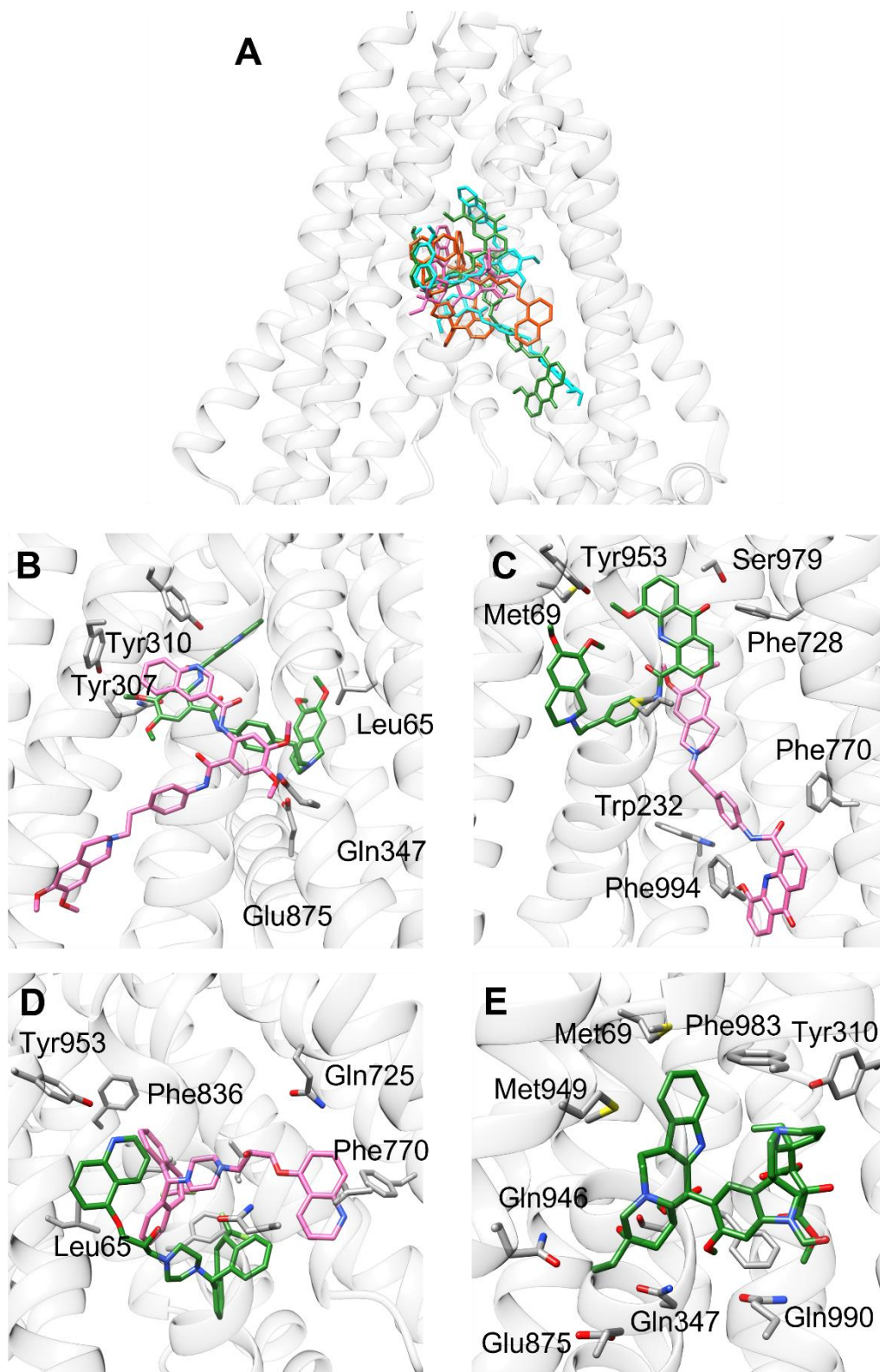


*Figure 1.2 Outward-facing conformations of ABC transporters. (A) Outward (occluded) conformation of human P-glycoprotein demonstrating squeezed inter TMD space. (B) Outward-facing conformation of SAV1866 showing wide extracellular opening.*

#### 1.1.1.1.2 Drug binding cavity

The binding cavity has been characterized using different structural and biochemical approaches (Chufan, Kapoor et al. 2013, Szewczyk, Tao et al. 2015, Nosol, Romane et al. 2020).

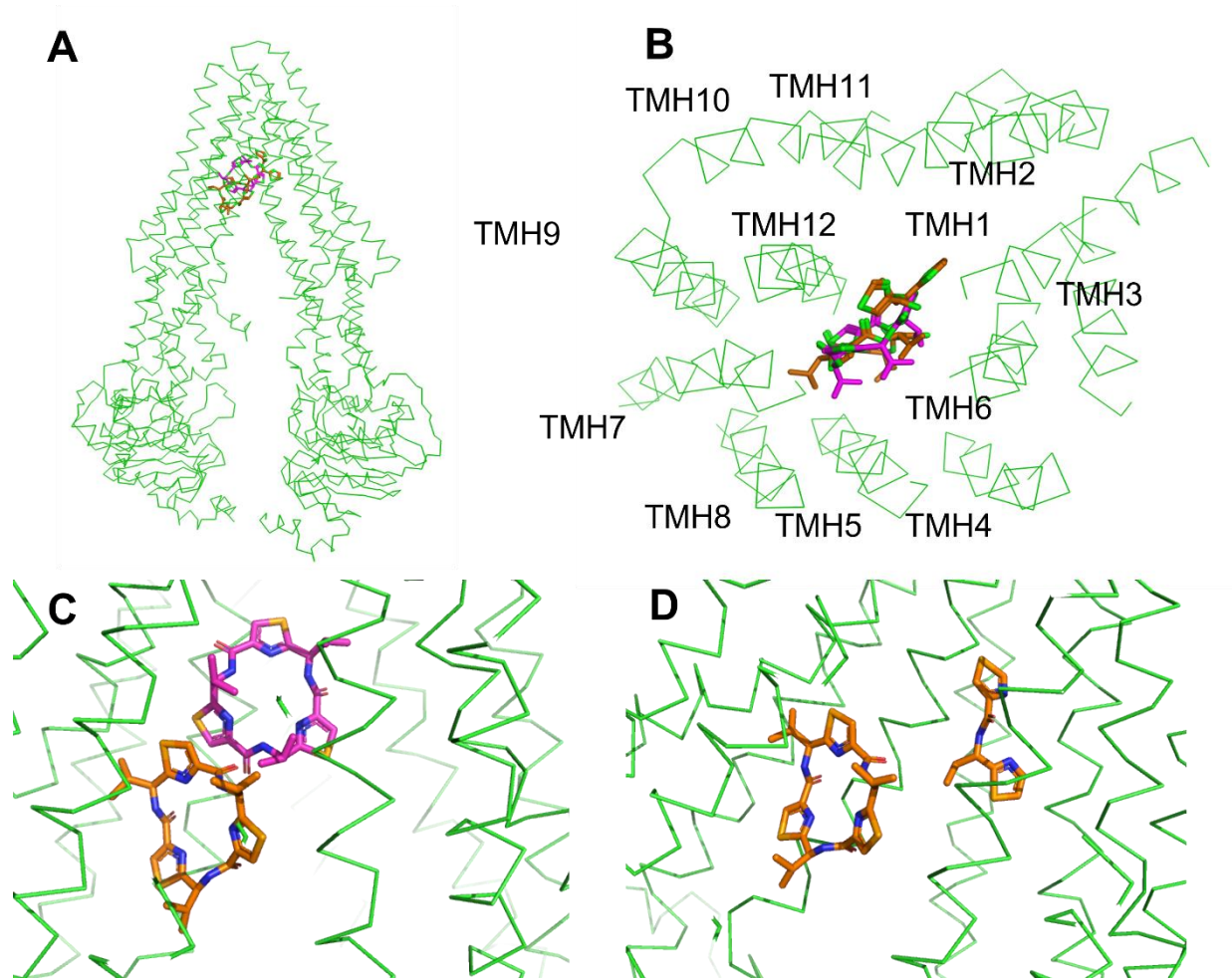
The recent snapshots of human P-glycoprotein with third-generation P-glycoprotein inhibitors have shown that P-glycoprotein can accommodate multiple copies of the drugs in the binding cavity, demonstrating the poly-specificity of the binding cavity. Figure 1.3 shows the poly-specificity of human P-glycoprotein inhibitors in the binding cavity.



*Figure 1.3. Poly-specificity of inhibitor binding in human P-glycoprotein. (A) Binding cavity of human P-glycoprotein identified from the cryo-EM structures. (B) Tariquidar's poly-specific binding with human P-glycoprotein (PDBID: 7A6E). (C) Binding of elacridar in the binding cavity (PDBID: 7A6C). (D) Poly-specificity of zosuquidar in the binding cavity (PDBID: 7A6F). (E) Vincristine binding with human P-glycoprotein (PDBID: 7A69).*

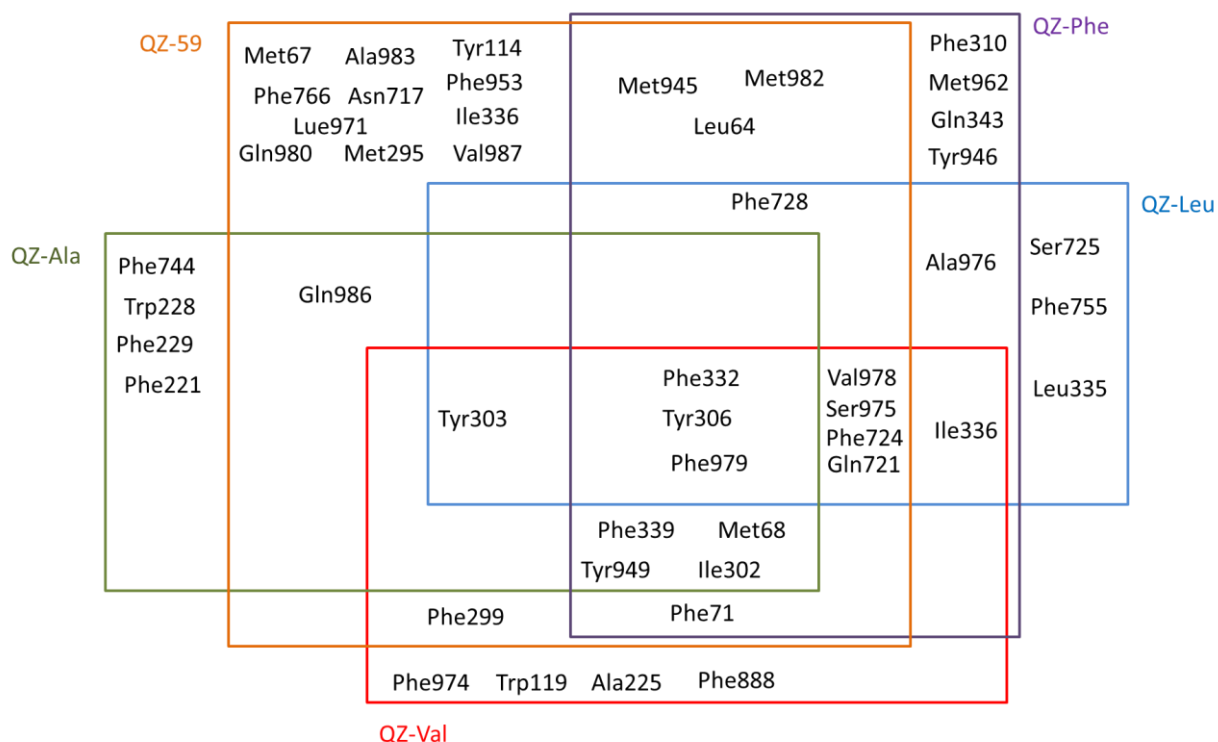
This poly-specific nature of the binding cavity is also established from biochemical methods. Chufan *et al*, used [125I]IodoArylAzidoPrazosin (IAAP) complexed molecules to study the substrate interaction. IAAP molecule upon photoexcitation reacts with binding cavity cysteine residues and generates the disulphide cross-linkages. In the presence of the substrate, IAAP loses its ability to induce cross-linkages. Mutation in the substrate interaction residues results in restoration of the cross-linkage phenomenon, but interestingly, despite mutation at the primary substrate site the substrates were able to translocate across the membrane, which meant that the substrate was still able to interact with the binding cavity (Chufan, Kapoor et al. 2013). Similar results were observed with the mass spectroscopic analysis of P-gp with cyclosporine A, where cyclosporine A was reported to interact with P-gp at two different places (Marcoux, Wang et al. 2013).

This poly-specificity is not limited to only human P-glycoprotein. Crystallographic studies on mouse P-glycoprotein in the presence of peptide-cyclic compounds (QZ59) have also demonstrated the poly-specific nature of the binding cavity (see Figure 1.4) (Li, Jaimes et al. 2014). QZ59 molecules have been shown to interact with the binding cavity at different positions simultaneously which suggests that the same molecule can interact in multiple ways with the binding cavity. Later more crystallographic structures of mouse P-glycoprotein with peptide mimetic (QZ59) compounds were reported (Szewczyk, Tao et al. 2015). Figure 1.5 shows a Venn diagram of residues that were identified to interact with peptide mimetic macrocyclic compounds within the binding cavity of P-glycoprotein. The interaction pattern indicates overlapping multiple binding sites. Interestingly, residues at TM6 and TM12 are common in all interactions. TM6 and TM12 are at the entry portals (see Figure 1.1B).



**Figure 1.4.** Diagrammatic representation of QZ59 in the binding cavity of P-glycoprotein. (A) Binding cavity of QZ59 with mouse P-glycoprotein identified from crystallography. (B) Binding cavity of QZ59 from the top. QZ59 is occupied between transmembrane helix (TMH) 4,5,6,7,12. (C) QZ59-RRR (magenta) and QZ59-SSS (brown) interact with P-glycoprotein at two different places (PDB ID: 4M2S, 4M2T). (D) Poly-specificity of QZ59-SSS in the binding cavity (PDB ID: 4M2T).





*Figure 1.5. Venn diagram of residues involved in ligand-protein interaction of peptide-cyclic compounds with mouse P-glycoprotein. The residues were in 5Å distance of the corresponding compounds. Residues for QZ-59 were obtained from the structures PDBIDs: 4M2S and 4MT2. For QZ-Ala, QZ-Val, QZ-Leu and QZ-Phe, PDBIDs:4Q9I, 4Q9J, 4Q9J and 4Q9L respectively were used*

Later structures of mouse-human chimeric P-gp in presence of the second-generation inhibitor zosuquidar (Alam, Küng et al. 2018) and mouse P-gp in the presence of an environmental pollutant (Nicklisch, Rees et al. 2016) have also shown these compounds to interact with similar residues at TM5, TM6 TM7 and TM12. Figure 1.6 shows the overlapping interaction pattern of zosuquidar, BDE-100 (POP), macrocycles and verapamil. Slightly different interaction pattern of verapamil was due to the residues identified using the cross-linking studies. This argues a different mode of binding/action of verapamil as compared to macrocyclic compounds. However, this requires further insights from high-resolution structural studies.



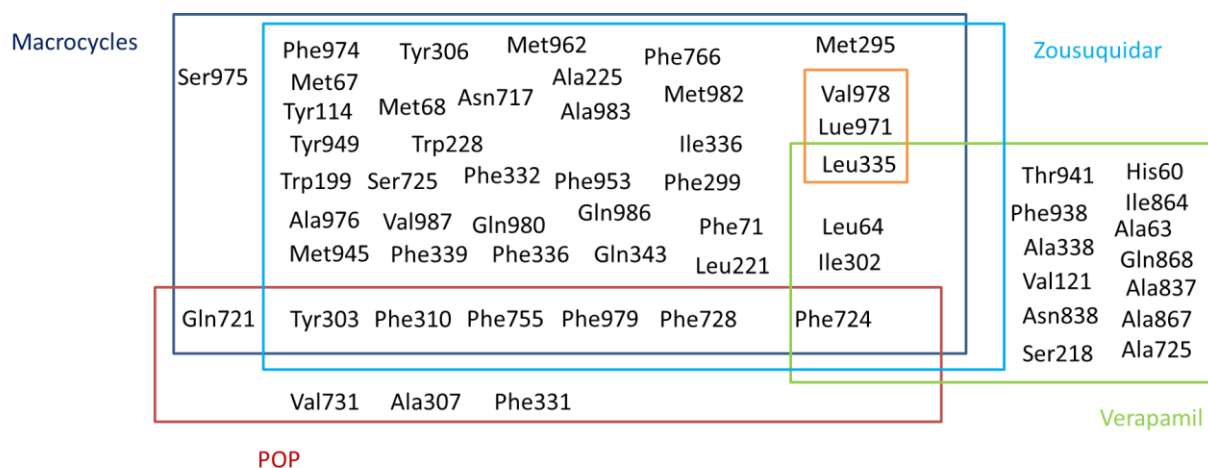


Figure 1.6. Venn diagram of ligand-protein interaction of structurally diverse compounds with mouse P-glycoprotein. Orange subset represents residues for Vinblastine and colchicine.

In general, all these reports delineated a large hydrophobic central cavity with multiple overlapping binding sites. This along with structural flexibility may explain P-glycoproteins's remarkable ability to interact with structurally diverse compounds.

#### 1.1.1.2 ABCG2

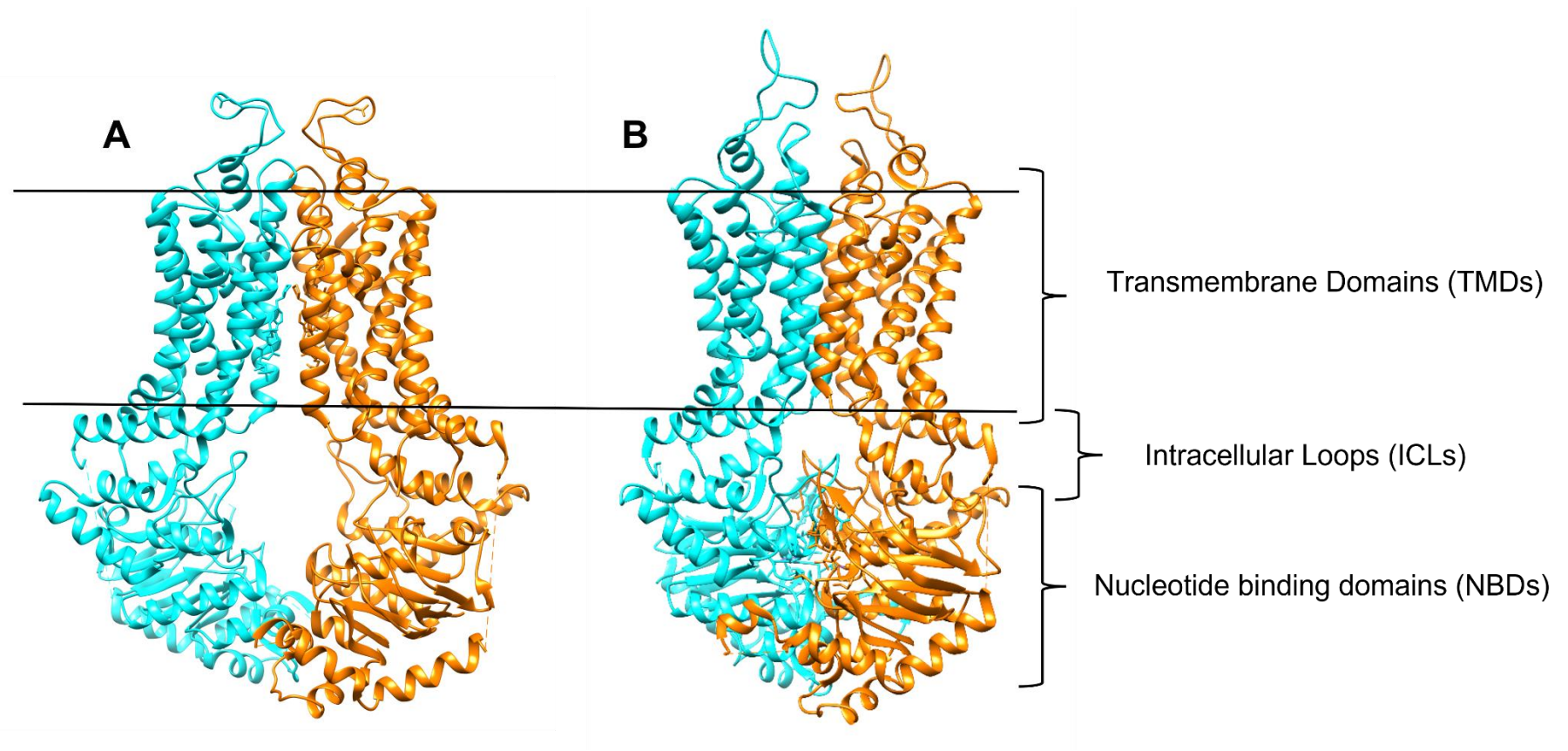
ABCG2 is relatively the most recent MDR protein, identified from a multidrug-resistant Breast cancer cell line (Doyle, Yang et al. 1998). Because of its origin of identification, it is also referred to as breast cancer resistance protein (BCRP). ABCG2 is a 655 amino acid, 70 kDa protein, expressed by the *abcg2* gene (Austin Doyle and Ross 2003). Characteristically ABCG2 is also expressed at the luminal membrane of the intestine, seminal vesicles, placenta as well as the cerebral cortex of the central nervous system and is involved in xenobiotic protection (Uhlén M, Fagerberg L et al.). Sequence analysis of ABCG2 demonstrates it as a half transporter i.e. consisting of a single chain containing a single transmembrane domain and a single nucleotide-binding domain (Miyake, Mickley et al. 1999). In contrast to P-gp, which demonstrate an N-terminal TMD and C-terminal NBD, ABCG2 has an inverse arrangement of the TMDs and NBDs (N-terminal NBD and C terminal TMD). Because of the NBD-TMD domain structure of

monomeric ABCG2, it requires a higher-order dimeric state for proper function (Ozvegy, Litman et al. 2001, Xu, Liu et al. 2004).

#### 1.1.1.2.1 Structure of ABCG2:

Recently, Taylor et al have reported the first high-resolution structure of human ABCG2 using cryo-EM and single-particle analysis (Taylor, Manolaridis et al. 2017). The structure delineates an Apo-closed conformation of ABCG2 bound with the Fab of a monoclonal antibody. Functional ABCG2 consisted of two monomeric chains of ABCG2 as shown in Figure 1.7. Each trans-membrane domain consists of six  $\alpha$ -helical structures that are arranged in symmetric configuration. TM 1, 5 and 6 contains the additional loop structure in the helix that can explain the structural flexibility of ABCG2. The structure shows a very compact arrangement between the TMDs and the NBDs in which the NBDs are very close to the inner leaflet of the lipid membrane (Taylor, Manolaridis et al. 2017). This configuration is very similar to the structural configuration of ABCG5 and ABCG8 structures (Lee, Kinch et al. 2016). Furthermore, the two TMDs are closer to each other as compared to the ABCB1 structures (see Figure 1.1).

The outward-facing conformations of ABCG2 show collapse of the lower binding cavity due to the ATP binding induced rearrangement of the TMDs (Manolaridis, Jackson et al. 2018, Yu, Ni et al. 2021). These rearrangements lead to the opening of the upper binding cavity towards the extracellular side (see Figure 1.7B). The two NBDs also rearrange themselves in a head to tail symmetry to generate two ATP hydrolysis sites. The intracellular loop acts as a pivot in the transfer of information between the TMDs and the NBDs (Manolaridis, Jackson et al. 2018, Khunweeraphong and Kuchler 2020). Table 1.2 shows the list of ABCG2 structures available on the PDB database.



*Figure 1.7. Structural representation of ABCG2. (A) Inward-facing conformation of ABCG2 (PDBID: 6ETI). (B) Outward occlude state showing collapsed drug-binding cavity and exposed extracellular cavity (PDBID: 6HZM). The functional unit of ABCG2 contains two monomers. Each monomer consists of transmembrane domains (TMD), a nucleotide-binding domain (NBD) and an intracellular loop region that links between the TMDs and the NBDs.*

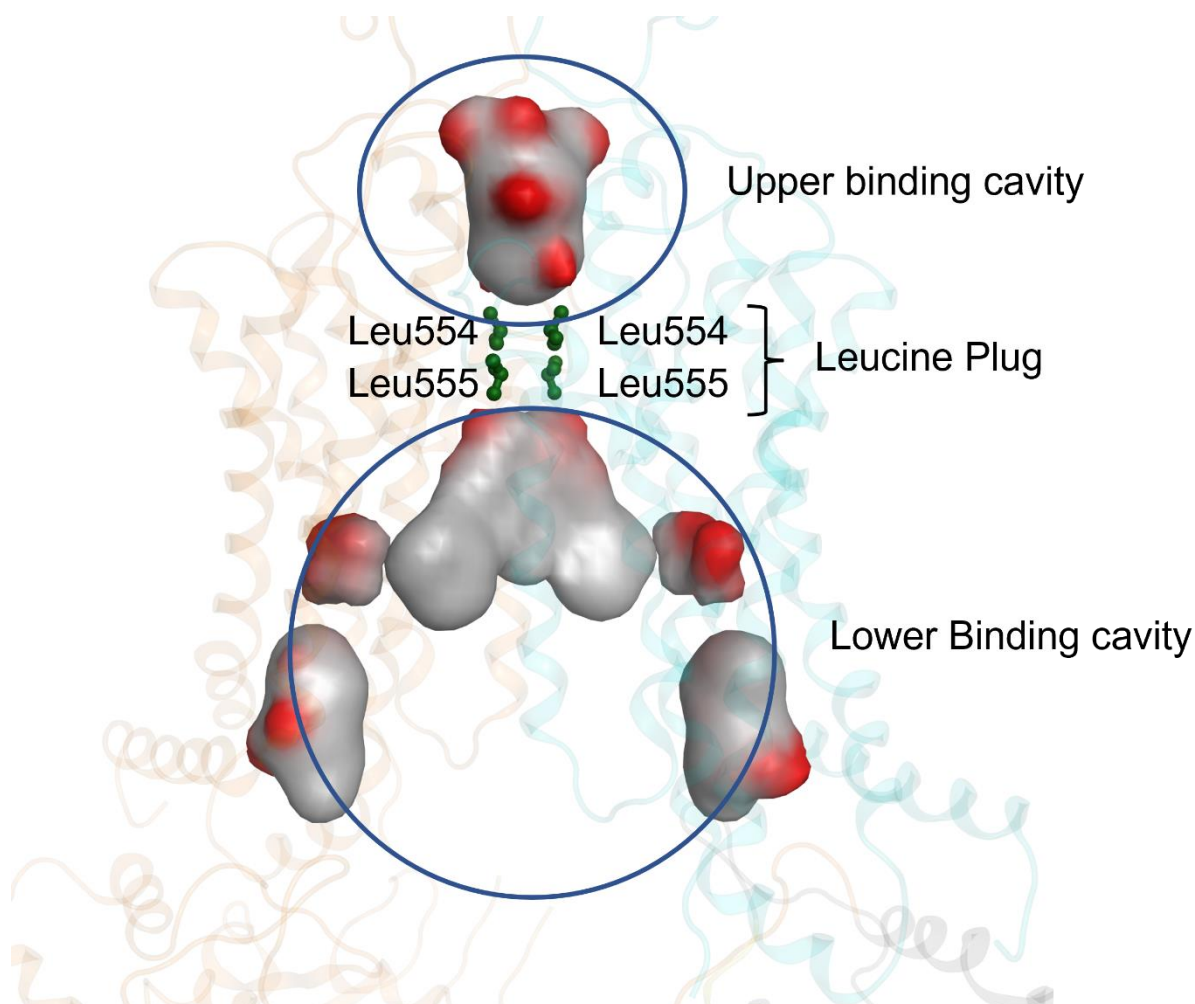
Table 1.2 List of available structures of ABCG2 analogues from the protein data bank (PDB).

PDB ID	Resolution (Å)	State	Drug	Reference
6VXJ	4.1	inward	SN38 +cholesterol	(Orlando and Liao 2020)
6VXI	3.7	inward	Mitoxantrone + cholesterol	
6VXH	4	inward	Imatinib + Cholesterol	
6VXF	3.5	inward		
7NFD	3.51	inward	Mitoxantrone +5D3 antibody fragment	(Kowal, Ni et al. 2021)
7NEZ	3.39	inward	Topotecan +NAG + 5D3 antibpdy	
7NEQ	3.12	inward	Tariquidar + NAG + Cholesterol + 5D3 antibody + U9N	
6FEQ	3.6	Inward	Tariquidar analogue + 5D3 antibody	(Jackson, Manolaridis et al. 2018)
6ETI	3.1	Inward	Ko143 analogue + 5D3 antibody	
6HIJ	3.56	inwards	MZ29 + Cholesterol + PEE	
6FFC	3.56		Ko143 analogue	
6HCO	3.58	Inward	Estron 3 sulphate + 5D3 antibody	
7OJH	3.1	Turnover state 1	Topotecan + 2 x ATP + Cholesterol	(Yu, Ni et al. 2021)
7OJB	3.4	Turnover state 2	Estron 1 sulphate + 2 x ATP + cholesterol	
7OJI	3.4	Turnover state 2	Topotecan + PLC + Cholesterol + 2 x ATP	
5NJ3	3.78	inward	5D3 antibody	(Taylor, Manolaridis et al. 2017)
6HZM	3.09	Outward occluded	2 x ATP + Mg	(Manolaridis, Jackson et al. 2018)
6HCO	3.58	inward	Estron 3-sulphate +5D3 antibody	
6HBU	3.09	Outward occluded	2 x ATP + Mg	

- All studies used HEK293 cell lines with protein reconstituted into the lipid nanodisc.
- All structures were identified by Cryo-EM.

#### 1.1.1.2.2 Binding cavity

In recent times, several studies have characterized the binding cavity of ABCG2 using biochemical and structural approaches (Jackson, Manolaridis et al. 2018, Khunweeraphong, Szöllősi et al. 2019, Gose, Shafi et al. 2020). The latest structures of ABCG2 have delineated a central hydrophobic cavity lined by TM2 and TM5a. The structures propose that the central cavity is divided into two binding chambers via Leu554 and Leu555 referred to as the leucine plug (see Figure 1.8). The upper chamber is proposed to be a low-affinity drug binding site, substrates are extruded out from this cavity (Khunweeraphong, Szöllősi et al. 2019), the lower chamber is the most characterized ABCG2 binding cavity from structural studies (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018, Orlando and Liao 2020). The drugs may access the lower chamber from the cytosol and induce a conformational change that leads to the dimerisation of the two NBDs and movement of the substrate to the upper chamber (Yu, Ni et al. 2021). Consequently, it leads to the opening of the TMDs to the extracellular side effluxing the drug out of the cell (Khunweeraphong, Szöllősi et al. 2019, Yu, Ni et al. 2021).



*Figure 1.8. Diagrammatic representation of ABCG2 binding cavity. The cavities were calculated with LIGSITE implemented in MOE. ABCG2 demonstrate a large central binding cavity divided into two cavities via a Leucine plug (green). The two cavities are represented as the upper binding cavity and lower binding cavity here.*

All the current of ABCG2 structures bound with drug delineate a common point of interaction with the ABCG2 binding cavity. Figure 1.9 shows a Venn diagram of the interacting residues with different classes of drugs. All these structures outline the involvement of TM2 and TM5 in the drug binding and translocation pathways. Biochemical studies also suggest the critical role of conserved residues on TM5 for drug binding and drug translocations (Manolaridis, Jackson et al. 2018, Gose, Shafi et al. 2020). However, these structures suggested multiple modes of substrates binding with the binding cavity. The ABCG2 selective inhibitor Ko143

derivative shows binding of two molecules in the binding cavity, with each molecule bound along with TMH2 and 5 on each monomer. Whereas other compounds bind in a 1:1 molar ratio of drug to dimeric proteins demonstrating an arch-like conformation between the two monomers (see Figure 1.10).

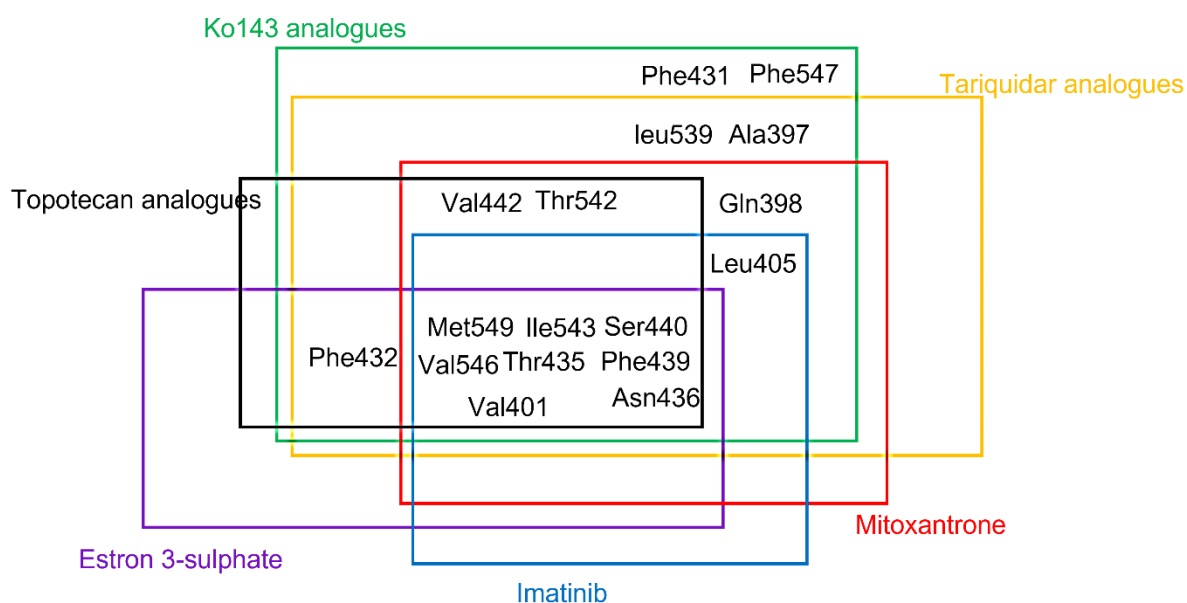
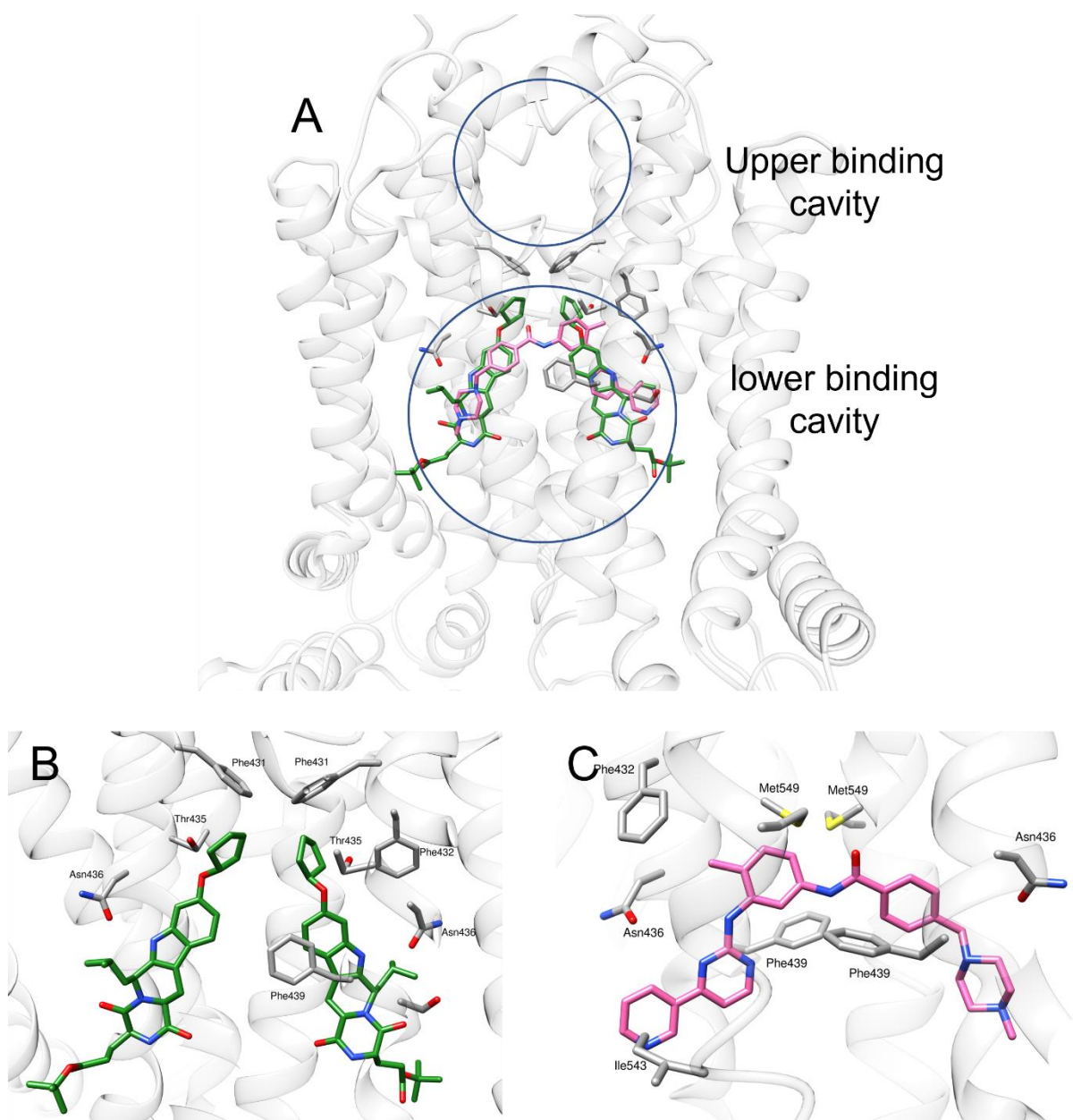


Figure 1.9. Venn diagram of ligand-protein interaction of structurally diverse compounds with ABCG2. All residues were within 5Å distance of the corresponding drug.



*Figure 1.10. Poly-specificity of ligand protein interaction in ABCG2. (A) Interaction of drugs (i.e. Ko143 and Imatinib) in the lower binding cavity. (B) Diagrammatic representation of Ko143 analogue in lower binding cavity of ABCG2 (PDBID: 6ETI). Each monomer binds with one molecule of Ko143. (C) Binding of imatinib with lower binding cavity (PDBID: 6VXH). Single molecule of imatinib make an arch-like conformation between the two monomers.*



## 1.2 Recombinant expression of membrane proteins

The expression of membrane proteins is tightly regulated and they are usually not expressed in high concentrations. Therefore, the production of functional recombinant membrane proteins in high amounts is a challenging task. With the advancement in heterologous expression systems, the large-scale production of membrane protein has become feasible (Wagner, Bader et al. 2006). The cDNA encoding the membrane protein is amplified and cloned into an expression vector. This genetically engineered vector is then transfected into the expression systems to overexpress the protein (Tzvi Tzfira 2005).

Expression of recombinant protein in a bacterial expression system is the desired method for several reasons. (i) The ease to manipulate the bacterial system. (ii) they are cost-effective and (iii) easy to handle (Terpe 2006, Kesidis, Depping et al. 2020). However, eukaryotic membrane proteins require highly regulated pathways of cellular trafficking to be able to express on the surface. In addition to that, most of the membrane proteins are highly glycosylated. The absence of these mechanisms in prokaryotes, therefore, makes the expression of eukaryotic membrane proteins in bacterial systems very difficult (Kesidis, Depping et al. 2020). In the 90's several attempts were made to express the functional P-glycoprotein in the bacterial system (Bibi, Gros et al. 1993, Evans, Ni et al. 1995). However these full-length proteins were unstable and short-lived (reviewed comprehensively in (Evans, Ni et al. 1995)). Up until now, very few studies have reported the successful expression of eukaryotic membrane proteins in prokaryotes (Weiss and Grisshammer 2002, Hiller, Garces et al. 2008, Srinivasan, Pierik et al. 2014).

Eukaryotic systems are recommended heterologous expression systems for the production of complex proteins. Expression is tightly regulated for degradation, membrane trafficking, and

glycosylation. Therefore, eukaryotic systems can provide a stable expression of the protein (Kesidis, Depping et al. 2020, O'Dell 2021). It has now been decades since eukaryotic systems have been used to overexpress recombinant membrane proteins. Yeast is a commonly used eukaryotic expression system for the expression of glycosylated proteins. Yeast demonstrates similar advantages of cost-effectiveness and ease of manipulation to bacteria. They also express a similar protein glycosylation pathway. Yeast has been widely used for the overexpression of membrane proteins (O'Dell 2021). In recent times, several studies have reported successful use of yeast host systems to express human MDR transporters (Lerner-Marmarosh, Gimi et al. 1999, Mao, Conseil et al. 2004, Lingam 2017). Bai *et al.* successfully used a codon optimization strategy to improve the expression of mouse P-glycoprotein in *Pichia pastoris* (Bai, Swartz et al. 2011).

Apart from yeast, insect and mammalian cell lines have also been used to express P-glycoprotein (Sarkadi, Price et al. 1992, Nosol, Romane et al. 2020). They can provide a native-like environment for membrane proteins functionality. However, working with cell lines is more complex, expensive and requires more equipment and better technical expertise.

### 1.3 Purification of membrane protein

Once the protein has been expressed in-vitro, it has to be purified for structural and biochemical analysis. Purification of soluble protein is very straightforward and does not require efficient technical expertise. However, purification of membrane protein is not as trivial as purification of soluble proteins (Smith 2011). Several factors can influence the purification process and repeated trial and error is required to optimize the protocol. In general, three critical steps are involved in the purification of integral membrane-bound proteins.

### 1.3.1 Solubilisation of membrane

An essential criterion for integral membrane proteins is to carefully solubilize the protein from the membrane fragments. This can be done using amphiphilic molecules called detergents (Seddon, Curnow et al. 2004). These molecules reduce the surface tension from the membrane and slowly elute the proteins from the membrane lipids (Lichtenberg, Ahyauch et al. 2013). Typically these detergents show a similar structure to lipids consisting of a polar head region and a hydrophobic tail. Based on their head group detergents can be ionic (e.g. Sodium dodecyl sulphate, SDS), non-ionic (e.g. dodecylmaltoside, DDM) and zwitterionic detergents (e.g. lauryldimethylamine-N-oxide, LDAO). Usually, ionic detergents are not recommended for the purification of the membrane proteins because they can interfere with native interactions resulting in denaturation of protein folds. Zwitterionic detergents have both negative and positive charges on the same head group thus have overall zero net charges. Because of this net-zero charge, they are less harsh than ionic detergents. Non-ionic detergents are mild and do not readily alter the native protein interactions during solubilisation. Mostly non-ionic detergents are used for membrane protein studies. However, the type of detergent used for purification is not trivial and requires repeated experiments to identify suitable compositions (Seddon, Curnow et al. 2004, Lichtenberg, Ahyauch et al. 2013).

The concentration of detergent during protein purification is very critical. At high concentrations, detergent also behaves similarly to lipids in an aqueous environment. They immediately form micelles by arranging themselves with the polar head facing the aqueous region and tail buried deep inside (Lichtenberg, Ahyauch et al. 2013). The minimum concentration at which the micelles are formed is called critical micelle concentration (CMC).

During membrane protein purification, initially, the concentration of protein is kept at least tenfold of the CMC. This not only solubilises the membrane protein complex but also ensure that every heterologous unit micelle contains only a single unit of protein. After solubilisation, the concentration can be reduced to the levels of CMCs to minimize the effects of detergent on protein folds.

### 1.3.2 Protein Isolation

Isolation of desired protein from the cellular proteome is also a critical step in protein purification. Affinity chromatography is a widely used method for protein purification. In Affinity chromatography, the proteins are labelled with tags and run through columns. Columns are specifically designed to attach to the tag residues. Untagged proteins are eluted first then the desired tagged proteins are eluted using buffers that compete for the tag. Size exclusion chromatography is another method for the purification of membrane proteins (Jin, Shen et al. 2021). O’Ryan *et al* demonstrated a two-step method of purification of ABC transporters (O’Ryan, Rimington et al. 2012). In the first step, the protein was purified using the Ni-NTA affinity chromatography, followed by size exclusion chromatography to remove impurities and fragmented CFTR to obtain full-length pure protein. The method showed purification quality of more than 90%. Later, the same two-step protocol for the purification was used for the purification of human and mouse P-glycoprotein (Lingam 2017, Thonghin, Collins et al. 2018).

Nowadays, commercial vectors are loaded with these affinity tags. The desired proteins are cloned upstream of these tags and expressed in suitable expression systems.

## 1.4 Structural techniques to study membrane proteins

Structural analysis cannot only provide the 3 Dimensional structure of the protein. It can also provide insights into the molecular mechanisms of the protein. This kind of information can be vital in structure-aided drug design.

### 1.4.1 X-ray crystallography

X-ray crystallography is a relatively old technique used to determine the structure of macromolecules. X-ray crystallography was first used to solve the structure of membrane protein in 1985 (Deisenhofer, Epp et al. 1985). Since then until the advent of cryo-EM it was the first choice for determining the high-resolution crystal structures of the membrane proteins (White 2009).

During crystallography, the proteins are assembled in a definite pattern of a crystal lattice. X rays are bombarded on these crystals, the X-rays hit the electrons and scattering at different angles. This change of angle from its actual path is termed X-ray diffraction. These diffraction patterns create spots of different intensities on a recording plate. This pattern of diffraction is then translated into a protein structure (Woolfson 1997).

The formation of crystals of membrane protein is a critical rate-limiting factor for membrane proteins (Lacapere, Pebay-Peyroula et al. 2007). Due to this, only a limited number of membrane protein structures are reported using X-ray crystallography. One of the major problems during crystal formation is the presence of detergent in the crystal that was used during the solubilisation. The presence of these solubilization molecules in the crystals makes them fragile and prone to radiation damage during diffraction analysis (Loll 2014). Other difficulties include the lack of sufficient quantities of protein after the purification and the poor stability of membrane proteins in artificial conditions (Sonoda, Newstead et al. 2011). In

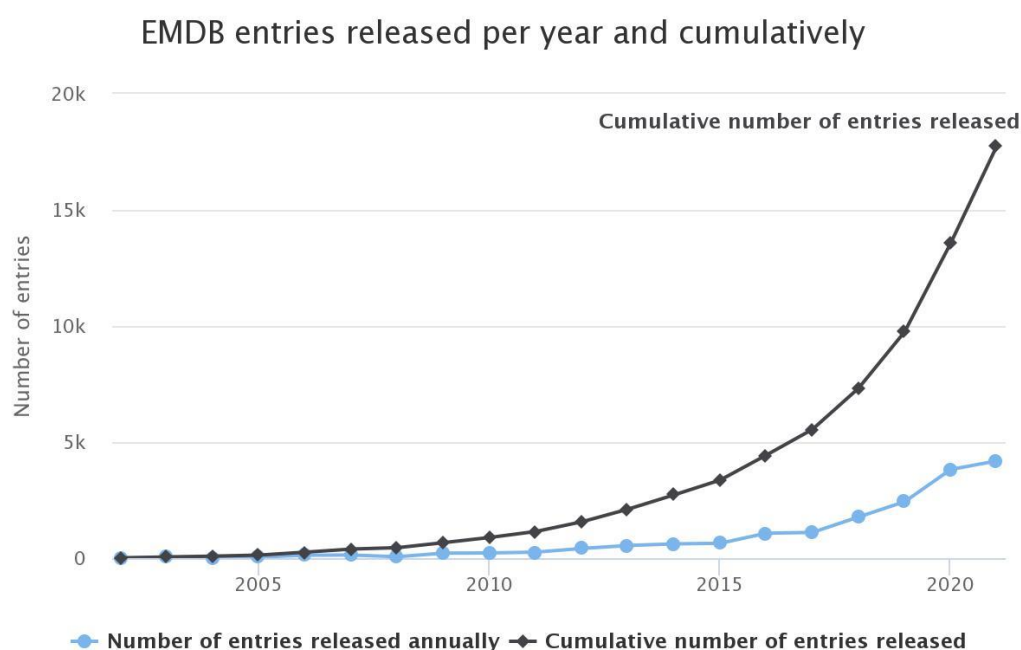
recent times, several techniques have been identified to reduce the problems e.g. use of less harsh detergents (Lacapere, Pebay-Peyroula et al. 2007, Sonoda, Newstead et al. 2011), availability of commercial kits for robust purifications and crystallization (e.g. MemGold Molecular Dimensions), improved expression systems to increase the yield of proteins. Yet, 3D crystal formation is still a daunting task and often yields low-resolution diffraction patterns (McPherson 2004).

#### 1.4.2 Cryo-Electron Microscopy (cryo-EM)

In the last decade, the cryo-EM technique has gone through a resolution revolution (Kühlbrandt 2014, Crowther 2016). The new direct electron detection cameras can give fast readouts that enables us to compensate for beam-induced drifts. Dozens of images of the same area are recorded in the form of a movie which allows us to track the particle motion (Kühlbrandt 2014). Furthermore, improvements in the hardware of the microscope have also allowed crossing the resolution barriers. A fundamental improvement is the introduction of a phase plate which increases the signal to noise ratio especially when working with proteins of lower molecular masses (Majorovits, Barton et al. 2007, Müller, Jin et al. 2010). Furthermore, The new sample stage designs allow better vacuum and temperature stability inside the cryo-EM chamber thus reducing the specimen drift. Also, the modern microscope allows remote operation and automated data collections to eliminate the instabilities caused by human interference (Grigorieff and Harrison 2011).

Cryo-EM is a widely used technique for determining the structures of macromolecules and membrane proteins are the typical clientele of cryo-EM (Thonghin, Kargas et al. 2018). It provides multiple advantages over the other conventional high resolution approaches i.e. (i) only requires a few micrograms of protein sample (ii) resolves protein in a near native

environment. (iii) can resolve proteins ranging from a few kilo Daltons to Mega Daltons to near-atomic resolutions. (iv) automation of the process (Benjin and Ling 2020). Because of these advantages, recent years have seen an exponential increase in depositions of protein structures resolved by cryo-EM (see Figure 1.11).



*Figure 1.11 Graphical representation of total (soluble and membrane protein) EM structures reported in the last two decades. (taken from EMDB website)*

#### 1.4.2.1 Grid Preparation

Sample vitrification is a critical step in cryo-EM that determines the quality of data collection and final resolution of the biomolecule (Sgro and Costa 2018, Weissenberger, Henderikx et al. 2021). The principle of vitrification is to flash freeze the sample to cryogenic temperatures to avoid the formation of ice crystals. Instead, the water makes a very thin layer of amorphous ice that is embedded with protein in random orientations (Dubochet and McDowell 1981, Weissenberger, Henderikx et al. 2021). In practice, the sample is applied to 3mm diameter

grids customized for cryo-EM. The excess liquid is removed by blotting the grid with filter papers. The time of blotting is a critical condition in determining the thickness of ice and sample distributions. Optimization of this factor is critical in obtaining high-quality data. The grid is then plunged frozen in a cryogenic liquid (e.g. liquid ethane) to form a thin layer of amorphous ice with protein embedded in it (Passmore and Russo 2016, Sgro and Costa 2018). The modern vitrification tools are very sophisticated and achieve vitrification in a very controlled environment which has made this process highly efficient and robust process (Frederik and Hubert 2005, Noble, Wei et al. 2018).

#### *1.4.2.2 Single-particle analysis*

Single-particle analysis (SPA) was developed to resolve biological structures that were not amiable with crystallography (Murata and Wolf 2018). Single-particle analysis is a systematic process that relies on averaging the 2D objects framed in different orientations and reconstructing them into a 3D structure (Vinothkumar 2015). The dataset for SPA is a collection of movies obtained from different locations on the cryo-preserved grids. Each movie is a stack of frames from the same area that contains information on beam-induced particles motions and stage drifts (Murata and Wolf 2018). Therefore, motion correction is a standard procedure in SPA to neutralise the beam-induced blurring effect (Brilot, Chen et al. 2012, Li, Mooney et al. 2013).

Cryo-EM images are generally collected using defocus to increase phase contrast and typically modified in frequency space via contrast transfer function (CTF). CTF mainly describes how the microscope transfers the information of a true 3D object into a 2D object in the micrographs (Jeong, Park et al. 2013). Thus data needs to be computationally transformed into reciprocal space via Fourier transformation to calculate the distortion in phase-



amplitudes. This eventually facilitates reconstruction of the true object at high resolution (Crowther, DeRosier et al. 1970). Typically, CTF correction is represented as an oscillation function with decreasing amplitudes at high frequencies. The experimental CTF profiles are compared with the theoretical profiles and the value for the best fit is considered as the defocus value. To date, a variety of tools have been developed which can perform CTF corrections and defocus estimations i.e. CTFFIND (Rohou and Grigorieff 2015). gCTF (Zhang 2016) for GPU accelerated estimations and e2ctf (Bell, Chen et al. 2018).

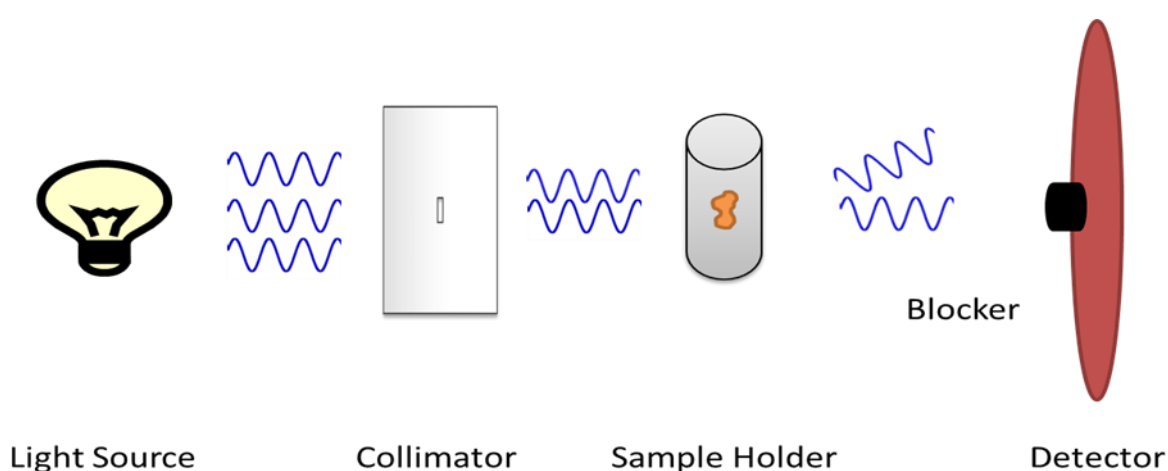
After CTF estimation, particles are picked, classified and averaged into 2D classes that represent the different orientations. These classes are used in the 3D reconstruction of the protein. 3D reconstruction refers to a procedure that combines particles from 2D classes in an iterative process called projection matching (Penczek, Grassucci et al. 1994). Simply the 3D reconstruction relies on the projection slice theorem where 2D images are positioned into a 3D projection (Crowther, DeRosier et al. 1970, Nogales and Scheres 2015). This is based on the fact that the 2D Fourier transform of each particle is a slice that passes from the centre of the 3D Fourier transform of the fundamental structure. In other words, the first structure is the rough estimation of 3D projection based on the 2D projections. During the 3D refinement, the experimental 2D projections are compared with the initial 3D projection to find the best match of an individual particle with the estimated structure (Sigworth 2016).

#### 1.4.3 Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) analysis is a powerful and robust technique to study biomolecules in solution (Blanchet and Svergun 2013). Contrary to crystallographic and cryo-EM analyses which are limited by the size of the proteins and challenges with specimen preparation, SAXS is a relatively easy technique that can be used to analyse proteins from a

few kDa to GDa. It can provide quantitative measurements of the size, shape, and surface of proteins in a solution. Furthermore, it also provides information about the intrinsic disorder, dynamics and conformational polydispersity of proteins in a native state (Blanchet and Svergun 2013, Boldon, Laliberte et al. 2015).

A simple SAXS instrument consists of a source of photons, a collimator system, sample holder, beam blocker and a scattering detection system. Figure 1.12 shows a schematic diagram of SAXS instruments. In small instruments, X-ray tubes are used as a photon source. However, in synchrotron facilities highly accelerated electron beams under high voltage are used to generate a constant wavelength source of X-ray photons (I. Svergun, Koch et al. 2013). These photons are then passed through the collimating system that reduces the highly divergent beams and only a coherent beam is focused on the sample holder where the sample is placed in a quartz capillary. Un-scattered high-intensity waves are blocked by a beamstop to reduce noise detection on the screen. The scattered photons are recorded on a detector screen. Resolution and frame rates per second of a detector are critical factors in obtaining a high quality scatter in SAXS (I. Svergun, Koch et al. 2013).



*Figure 1.12. Diagrammatic representation of SAXS instrument.*

During the experiment, protein in a solution is exposed to a beam of X-rays that are scattered in an isotropic manner and are recorded relative to a momentum of transfer given as

$$q = \frac{2(\pi \cdot \sin 2\theta)}{\lambda}$$

Here,  $2\theta$  is the angle between the incident and scattered radiation. And  $\lambda$  is the wavelength of the incident radiation.

The mathematical theory of analysing such data was first explained by Guinier and Fournet (Rice 1956). At low  $q$  values, the scattering profile can give qualitative measures of Radius of Gyration ( $R_g$ ) and forward scatter which can be given by the following relation

$$i(q) = i(0) \left( \frac{-s^2 R_g^2}{3} \right)$$

$R_g$  is a basic structural parameter that can be deduced from the scattering profile. It is the root mean square of the distance from the centre of mass of the sample. It defines the overall distribution of the mass of the protein around its centre of gravity. The limitation of this Guinier analysis is that it is strictly dependent on the low “ $q$ ” values therefore, they are highly prone to any discrepancy at large scattering angles or where there is inter-particle interaction. Another robust way of representing the SAXS data is Pairwise Distance Distribution Function (PDDF) (Debye 1915) given as

$$i(q) = \int_0^{D_{max}} dr \cdot p(r) \cdot \frac{\sin(qr)}{qr}$$

This PDDF is a Fourier transformation function that converts the reciprocal space data from scattering into a real space distance distribution histogram in which each point is a distance

between two atoms of the given sample. This distance distribution graph tells us about the overall shape of the particle along with size. Both information from  $R_g$  and pairwise distance distribution function can be further used in ab-initio modelling of the protein (Daniel and I. 2009).

Ideally, a SAXS sample should have a high concentration of protein that is monodisperse in a solution with no inter-particle interactions (Henderson 2011, Jeffries, Graewert et al. 2016). These ideal properties are not true for all proteins, as they usually show some intrinsic inter-particle interactions. To reduce these interactions either size exclusion chromatography linked SAXS experiments are designed (Ryan, Trewhella et al. 2018) or alternatively, different concentrations of the sample are analysed and extrapolated to infinite dilution (Jacques and Trewhella 2010). Analysis of negative control or blank buffer control is critical in SAXS experiments (Jacques and Trewhella 2010, Henderson 2011). The control solution should be identical to the buffer containing the sample. A mismatch in the concentration of the two solutions can result in a misleading interpretation (Jeffries, Graewert et al. 2016).

## Aims and objectives

The general aim of this project was to elucidate ligand-protein interaction of drugs with P-glycoprotein and ABCG2 and to understand the drug translocation pathway using structural and computational approaches.

The specific aims were

- Purification of human and mouse P-glycoprotein from the yeast expression system.
- Examine the conformational changes in P-glycoprotein using the low-resolution X-ray scattering profiles of human and mouse P-glycoprotein in different catalytic states.
- Investigate the ligand-protein interactions of ivacaftor with human and mouse P-glycoprotein.
- Optimization of virtual drug screening protocol and screening of drug library against human P-glycoprotein.
- 

In collaboration with the Schuetz lab at St. Jude Children's Hospital

- To investigate the role of Phe439: a highly conserved residue of ABCG2 in drug translocation pathway.
- Optimization of virtual drug screening protocol and screening of drug library against human ABCG2.

## Material and Methods

### 2.1 Materials and Equipments

#### 2.1.1 Chemicals

Yeast Nitrogenous Base (YNB)	Sigma
Ammonium persulphate	Fisher Scientific
Uracil dropout supplement	Formedium
Glycerol	Acros organics
Agar	formedium
Zeocin	Fisher Scientific
Di-potassium hydrogen phosphate ( $K_2HPO_4$ )	Sigma
Phosphate dihydrogen phosphate ( $KH_2PO_4$ )	Sigma
Glucose	Fisher-Scientific
Galactose	Sigma
Tris	Fisher-Scientific
Sucrose	Fisher-Scientific
Ethylenediaminetetraacetic acid (EDTA)	Fisher-Scientific
Dithiothreitol (DTT)	Fisher-Scientific
Sodium Chloride (NaCl)	Fisher-Scientific
n-Dodecyl $\beta$ -D-maltoside (DDM)	Anatrace
Cholesterol hemisuccinate (CHS)	Sigma
Imidazole	Fisher-Scientific
Glycine	Fisher-Scientific
Sodium dodecyl sulphate (SDS)	Sigma
Sodium Orthovanadate	Sigma
Surecast 40% poly acrylamide	Thermo-Fisher
Instant Blue stain (coomassie blue)	Sigma-Aldrich
4-20% stain free gradient gels	Bio-Rad
PAGERULER PLUS pre-stain protein ladder	ThermoFisher-Scientific
Bradford reagent	ThermoFisher-Scientific
Methanol	Fisher Scientific
Acetic acid	Fisher-Scientific

Coomassie blue G250	Fisher Scientific
Bromophenol blue	Sigma Aldrich
Tween-20	Sigma
TNP-ATP	Tocris Bioscience
Uranyl acetate	Agar Scientific

### 2.1.2 Equipment and Consumable

Acid Washed glass beads	Sigma Aldrich
Akta FPLC	Akta
Bead beater	Biospec
Bioreactor	Applikon
Carbon coated 300 nickel mesh grid	Agar Scientific
Quantifoil 1.2/1.3 Cu grids	Agar Scientific
96 well plate reader (fluorometer)	Biotek
HisTrap column	Generon
SepFast™ SEC 6-6000kDa	Generon
JP 33 film processor	JPI healthcare
K100X glow discharge system	Quorum Technologies
Uncle spectroscopic system	Unchained Lab
Viva spin concentrators	Sartorius
VitroBot Mark IV	ThermoFisher Scientific
Glacios cryo-EM	ThermoFisher Scientific
Fluorometer	Cary Eclipse
Optima XE90 ultracentrifuge	Beckman Coulter
Avanti J26S XP	Beckman Coulter
Optima Max-XP ultracentrifuge	Beckman Coulter
Ti-45 45000 rpm rotar	Beckman
JLA8.1 8000 rpm rotar	Beckman
TLA 120.1 rotar	Beckman

## 2.2 Buffer recipes

### **MGY media for *Pichia Pastoris***

MGY agar: 1.34% YNB, 1% glycerol, 2% agar, 500µg/ml zeocin.

MGY media: 1.34% YNB, 1% glycerol, 0.1M potassium phosphate buffer.

### **YNB media for *Saccharomyces cerevisiae*:**

YNB agar: 0.69% YNB, 0.077% uracil dropout supplement, 2% agar, 2% glucose

YNB media: 0.69% YNB, 0.077% uracil dropout supplement, 0.2/2% glucose

### **Buffer for microsome preparations:**

mPIB buffer: 0.25M tris-HCl (pH 8), 0.25M sucrose, 1mM EDTA, 2mM DTT

High Salt buffer: 50mM tris-HCl (pH 8), 500mM NaCl, 10% glycerol

Solubilization buffer: 50mM tris-HCl (pH 8), 200mM NaCl, 10% glycerol, 2% DDM

### **Ni-NTA purification:**

mpgp-buffer A: 50mM tris-HCl. (pH 8), 200mM NaCl, 10% glycerol, 1mM DTT, 0.1% DDM

mpgp-buffer B: 50mM tris-HCl. (pH 8), 200mM NaCl, 10% glycerol, 1mM DTT, 500mM imidazole, 0.1% DDM

hpgp- buffer A 50mM tris-HCl (pH 8), 150mM NaCl, 20% glycerol, 0.1% DDM

hpgp buffer B 50mM tris-HCl (pH 8), 150mM NaCl, 20% glycerol, 500mM imidazole, 0.1% DDM

**SEC buffer:** 50mM tris-HCl (pH 8), 150mM NaCl, 10% glycerol, 0.05% DDM

**Cryo buffer:** 20mM tris-HCl (pH 8), 20mM NaCl

### **SDS-PAGE**

10X running buffer 250mM tris, 1.92M glycine, 1% SDS

1X running buffer 25mM tris, 192mM glycine, 0.1% SDS

resolving buffer 1.5M tris-HCl (pH 8.8), 0.4% SDS



stacking buffer	0.5M tris-HCl (pH 6.8), 0.4% SDS
Coomassie solution	50% methanol, 10% acetic acid, 0.1% Coomassie-blue G-250
Destain solution	50% methanol, 10% acetic acid
Sample buffer	0.375M tris (pH 6.8), 60% glycerol, 12% SDS, 0.3M DTT, 0.06% bromophenol blue

### **Western blotting**

transfer buffer	25mM tris, 192mM glycine, 20% methanol
10X TBS	200mM tris, 1.5M NaCl
1X TBS	20mM tris, 150mM NaCl
TBS-T	20mM tris, 150mM NaCl, 0.1% tween
Ponceau Stain	0.1% ponceau dye, 5% acetic acid
blocking buffer milk	20mM tris, 150mM NaCl, 0.1% tween, 5% powdered skimmed milk

### **Inhibitor Cocktail 100x**

<b>AEBSF</b>	200mM	48mg in 1ml distilled water
<b>Benzamidine</b>	300mM	36mg in 1ml distilled water
<b>Chymostatin</b>	4mM	2.5mg in 1ml DMSO
<b>E-64</b>	7mM	2.5mg in 1ml distilled water
<b>Leupeptin</b>	20mM	10mg in 1ml distilled water
<b>Pepstatin A</b>	15mM	10mg in DMSO
<b>PMSF</b>	1M	174mg in DMSO

## 2.3 Methodology

### 2.3.1 Expression and purifications

#### 2.3.1.1 *Constructs*

Human P-gp was cloned in  $\Delta p424GAL1$  and expressed in *Saccharomyces cerevisiae* by Dr Swathi Lingham (Lingam 2017). Briefly, the gene was directly cloned under a strong galactose promotor (GAL1) and was transformed into XL-10 GOLD ultracompetent cells. p424GAL1 is designed to express the protein *in vivo* however, does not contain tags for protein purification. Therefore, the cloned vector was modified to add a 10X histidine tag and a stop tag at the C terminal end after the GFP tag for protein purification. The modified plasmid was then transfected into *Saccharomyces cerevisiae* using the protocol as described by Prof David Drew (Drew, Newstead et al. 2008) and O’Ryan, *et al* (O’Ryan, Rimington et al. 2012). Only genetically modified colonies were able to grow in uracil drop-out medium.

A Mouse P-gp construct expressing mP-gp in *Pichia pastoris* (KH71M stain) was provided by Prof. Ina Urbatsch (Bai, Swartz et al. 2011). The codon-optimized gene was added with a 6-histidine tag at the C-terminus and inserted into pHIL vector under the auxotrophic AOX1 promotor. The genetically modified organisms were screened using 100 $\mu$ g/ml Zeocin selection plates.

#### 2.3.1.2 *Expression of mP-gp*

Mouse P-gp expression was carried out using the protocol described by Lerner-Marmarosh et al (Lerner-Marmarosh, Gimi et al. 1999) and Thonghin et al (Thonghin, Collins et al. 2018). In brief, frozen glycerol stocks were thawed at room temperature and evenly spread on a zeocin selective MGY plate. The cells were allowed to grow at 30°C for 2-3 days or until the whole plate was covered with culture. The cells were then inoculated using a loop into 100ml initial seed culture and left overnight on a shaking incubator at 30°C. Later that day, six 2L baffled flasks

with 1L MGY media each were prepared and autoclaved. The next morning the 10ml of seed culture was equally divided into each 2L flask with 1L autoclaved MGY media and incubated at 30°C for 24 hours with 225rpm shaking. After 12-16 hours, a 0.5% (v/v) pre-induction glycerol feed was added to each flask to boost the cell growth and culture was allowed to grow at 30°C for 5-6 hours. The protein expression was induced by adding 0.5% v/v to each flask followed by booster doses of 0.5% (v/v) methanol inductions at 24hours and 48hours. The cells were harvested after 72 hours of induction by centrifuging cultures at 2300x *g* for 10 minutes. Cells were resuspended in ice-cold mPIB buffer and snap frozen with liquid nitrogen and stored at -80°C for subsequent processing.

#### *2.3.1.3 Expression of human P-gp*

Small-scale expression was carried out in a 2L baffled flask. In short, the glycerol stock was thawed at room temperature, re-suspended and evenly spread over a uracil dropout YNB agar plate. The cells were allowed to grow at 30°C for 2-3 days. The cells were streaked and inoculated in a 1L YNB media containing 0.2% glucose and was left to grow overnight at 30°C with continuous shaking at 225 rpm in baffled flasks. Glucose consumption is critical in protein expression and must be consumed before galactose induction (O'Connor 2021). Glucose levels were checked using glucose strips at periodic intervals after 12 hours until no colour change observed on the glucose testing strip, indicating total glucose has been consumed. The protein expression was induced by adding 8% (v/v) glycerol and 2% (w/v) galactose (100ml/L of 20% (w/v) filtered galactose stock solution). The cells were harvested after 12 hours by centrifugating at 3000 rpm for 10 minutes, and pellets were resuspended in mPIB buffer.

Large scale expression was carried out using an Applikon 18L fermenter system with Bio-Controller 1010 as explained by (O'Ryan, Rimington et al. 2012). The initial seed was prepared

in three 2L baffled flasks with 0.75L YNB media in each flask. The media was autoclaved and inoculated with *S. cerevisiae* expressing the human P-glycoprotein. The flasks were incubated overnight at 30°C on an orbital shaker incubator at 225 RPM. Meanwhile, 12L of YNB media was pumped into the sterile fermenter vessel through a peristaltic pump. The temperature was regulated at 30°C by a heat-blanket covered around the glass chamber and cooling water passing through a heat exchanger. The media was left overnight with aeration through a 0.22µm filter pump and continuous stirring at 700rpm with a p1000 motor. Continuous stirring ensured the even distribution of oxygen and media. The next morning the 2.25L of initial seed was inoculated into the fermenter using the peristaltic pump and sterile piping. Optical density at 600nm (OD600) and glucose levels were periodically monitored to measure the growth of yeast. Optical density should be higher than 1.2 when glucose levels are completely depleted (O'Ryan, Rimington et al. 2012). If OD600 was less than 1.2 more glucose was added to reach the desired optical density. Meanwhile, 1.5L of glycerol (8%v/v) and 360 grams of galactose (2%w/v) were dissolved in 2L of YNB media and filtered through a 0.22µm filter for protein induction. Once the OD600 reached the desired levels and glucose was fully consumed. The above mentioned induction media was injected to initiate the protein expression. The cells were harvested after 12 hours post-induction by centrifuging at 2300 x g for 10min at 4°C and resuspended in ice-cold mPIB buffer. The harvested cells were snap frozen using liquid nitrogen and stored at -80°C until further use.

For time scale experiment, 50ml of induced culture was collected in fresh falcon tube at each periodic time point. The cells were harvested by centrifugating at 3000 rpm for 10 mins. Supernatant was discarded and cells were resuspended in 5ml ice cold mPIB buffer. The cells were fixed on the glass cover slips using the poly-lysine coating. In summary, the cover slips

were incubated with 100  $\mu$ L of poly-lysine solution for 10 minutes. Subsequently, the cover slips were washed with water to remove excess of the poly-lysine and dried on room temperature. Afterwards, 50 $\mu$ L of re-suspended cells were applied to the poly-lysine coated cover slip and incubated for 10 min. excess sample was washed and cell were analysed using Olympus BX60 microscope using 60X objective lens. Images were taken with CoolDNAP EZ CCD camera (Photometrics) using MetaMorph software (MDS Analytical Technologies). Specific band pass GFP filter set was used and exposure time was set to 500m/s for each set of images. Images were then analysed using Image J (<http://rsbweb.nih.gov/ij/>). The mean intensity of each field of view was calculated

#### *2.3.1.4 Crude membrane preparation*

Microsomes were prepared by using slight modifications of the bead mill method as explained by (O'Ryan, Rimington et al. 2012) and (Pollock, Cant et al. 2014). Briefly, frozen cells were thawed on ice and diluted to a final concentration of 0.3-0.4 g/ml (wet weight) with ice-cold mPIB buffer. The mixture was supplemented with a 1X protease inhibitor cocktail to eliminate internal protease activity during the bead beating process. To rupture the cells 200g of acid-washed glass beads (400-600 $\mu$ m in diameter) were added to the steel chamber of a midi bead beater system (Biospec). Subsequently, the cell-inhibitor slurry was added to the chamber with beads in a 1:1 ratio (200grams of beads: 200ml of cell slurry). The bead beating was performed for cycles of 2 minutes with 1 min of resting for a total of 10-12 minutes of bead beating time. Throughout the process, the steel chamber was covered with dry ice to keep the sample as cold as possible.

After bead beating, the slurry mixture was allowed to rest on ice for 5 min to sediment the glass beads from the slurry. The suspension was carefully collected and centrifuged at 14000 x g for

10 min at 4°C to remove unbroken cells and cellular debris. The supernatant was centrifuged at 120,000 x g for 90 min at 4°C to pellet the membrane. Subsequently, the supernatant was discarded and the microsome pellet was washed and resuspended with high salt buffer using a Teflon in a glass homogenizer. The suspension was centrifuged again at 100,000 x g for 45 min and the microsome pellet was resuspended in solubilization buffer. Total protein concentration was estimated using the Bradford assay. The presence of protein of interest was confirmed by Western blotting using anti-His antibody and GFP scanning after electrophoresis. Microsome suspension was flash-frozen in liquid nitrogen and stored at -80°C till further process.

For solubilization, 10% (w/v) DDM stock solution in solubilization buffer was prepared and mixed with microsomes to a final concentration of 2.5mg/ml total protein and 2% (w/v) DDM. Solubilization was performed at 4°C with gentle agitation for 1 hour. The suspension was centrifuged at 100,000 x g for 45 min at 4°C. the supernatant was directly loaded onto the Ni-NTA affinity column for affinity purification. (However the sample could be flash frozen and stored at -80°C for long-term storage at this stage).

#### *2.3.1.5 Immobilized metal affinity chromatography*

Purification was performed on GE healthcare AKTA FPLC system with a Frac950 fraction collector system. The sample was loaded onto the pre-equilibrated Ni-Affinity column (Generon HiFliQ) with a flow rate of 0.5ml/min. Flow-through was collected in falcon tube for further analysis. The column was washed with four percent (20mM imidazole) elution buffer B and subsequently with 16 percent (80mM imidazole) elution buffer B (see page 48 for composition) to remove loosely bound impurities. Purified P-glycoprotein was eluted at 200mM imidazole concentration (40 percent elution buffer B). Fractions were analysed on SDS-PAGE with Coomassie Blue staining and for human P-gp also scanned for the GFP fluorescence signal.

The elution fractions with purified protein were pooled and concentrated using a 100kDa cut-off centrifugal filter (Vivaspin20). The protein was concentrated to a point where further concentration produce precipitation in the concentrator. The concentrated protein was centrifuged at 200,000 x g for 10 minutes to remove large aggregates and the supernatant was aliquoted into 500µl aliquots (max sample capacity of the SepFast 6-6000kDa, small-scale SEC columns) and snap frozen for multiple rounds of SEC.

#### *2.3.1.6 Size Exclusion Chromatography (SEC)*

Size exclusion chromatography was performed on a GE healthcare AKTA FPLC system with Frac900 collection system. The column (Generon SepFast™ 6-5000 kDa) was calibrated with 2 column volume degassed ultrapure water and followed by 1.5 column volume of degassed SEC buffer. The sample was injected into a 500µl super loop, and which was then loaded onto the calibrated column. The purification was performed at 0.5ml/min flowrate with collection of 1ml fractions for 1.2 CV. Absorbance at UV280 and in the case of human P-gp GFP fluorescence was employed to track protein and was later confirmed by on SDS-PAGE. Protein fractions enriched with pure protein were pooled and concentrated using a 100kDa cut-off (Vivaspin 6) filter and aliquoted and snap frozen for further use.

#### *2.3.1.7 SDS-PAGE*

Quantification and analysis of the purification procedure was performed using 8% polyacrylamide gels (Maniatis, Fritsch et al. 1982). The gels were cast in a BioRad mini gel casting system. 30% bis-acrylamide (1:29) was diluted to 8% in resolution buffer (1.5M tris-HCl (pH 8.8), 0.4% SDS) followed by the addition of 0.1% APS and 0.2% TEMED to initiate the polymerization. The stacking gel was prepared using 4% acrylamide and 0.1% APS and 0.2%

TEMED in stacking buffer (0.5M tris-HCl (pH 6.8), 0.4% SDS). Polymerized gels were stored at 4°C up to a week in a moist environment.

The samples were mixed with 1X sample SDS loading buffer and incubated for one hour at room temperature. Polymerized gels were installed in a Bio-Rad gel electrophoresis system and subsequently filled with 1X SDS running buffer. The prepared samples along with the standards were loaded into the wells and run at 80V for two hours or until the forward dye front was near the edge of the gel. The resolved gels were taken out of the cassette and allowed to stain overnight in Coomassie Blue stain (50% methanol, 10% acetic acid, 0.1% Coomassie-blue G-250). In the case of human P-gp labelled with GFP, gels were scanned under Cys5, Cys3, and Alexa488 using BioRad Chemidoc MP imager before the Coomassie blue staining protocol. Afterwards, Gels were stained using Coomassie Blue stain. The gels were de-stained using 10% (v/v) acetic acid and 50% (v/v) methanol solution until Coomassie labelled bands are visible. For glycosylation assays, where the negative control was a 20kda soya trypsin inhibitor protein, 4-20% precast gradient gels (BioRad) were used instead of standard 8% gel.

#### *2.3.1.8 Protein quantification*

Protein quantity was determined using the gel densitometry method. Serial dilutions of protein along with standard dilutions of BSA were run in SDS PAGE. The gel was incubated with Coomassie-Blue stain overnight. The stained gel was washed with de-stain solution till background noise was minimal. The image was scanned and imported to ImageJ for quantification protocol as explained at [https://openwetware.org/wiki/Protein\\_Quantification\\_Using\\_ImageJ](https://openwetware.org/wiki/Protein_Quantification_Using_ImageJ). in summary, the image was imported into the ImageJ package and lanes with proteins bands were selected. Intensity



histogram plot for each lane was generated. The area under the peak represented the intensity of the band. The intensity of each standard band was calculated and used for plotting a standard plot. The concentration of purified protein was estimated from the linear slope function of the plot.

### 2.3.2 Protein characterization

#### 2.3.2.1 *Western blotting*

Western blotting analysis was performed using the wet-blotting method (Mahmood and Yang 2012). In brief, samples were prepared and run as described earlier for SDS-PAGE (see section 2.3.1.7). The protein was transferred to a nylon membrane in a wet electroblotting protocol. The gel was sandwiched with the membrane in a transfer cassette with gel on the negative terminal and membrane towards the positive electrode. The transfer was run overnight in a chilled transfer buffer in a cold room with a constant voltage (30V). Protein transfer was checked with Ponceau stain and washed using PBST. The washed membrane is immersed in blocking buffer for 1-2 hours on a rocking plate and subsequently, incubated with 1:10,000 primary (anti-his) antibody solution overnight at 4°C. The membrane was washed three times with PBST for ten minutes. For blots with HRP labelled primary antibodies, the membrane was incubated with a chemiluminescence solution (1:1 mixture of luminol solution and HRP substrate) for 45 seconds. Images were obtained in BioRad Chemidoc MP imager system with chemiluminescence filter. In some cases, a secondary dye labelled antibody was used. The blot membranes after washing as described above were incubated with 1:10,000 secondary antibody solution for 1 hour at room temperature. The membrane was again washed twice with PBST and imaged using a Li-Cor Odyssey CLx infrared imaging system.

### *2.3.2.2 Thermal stability assay*

Thermal protein stability assay was used to determine the melting temperature ( $T_m$ ) of the protein.  $T_m$  represents the temperature at which 50 percent of the protein is denatured, in this case by heating. In general, 2  $\mu$ g purified protein was heated 4°C to 90°C for 5 min followed by 5 min on ice. Samples were then centrifuged at 100,000 x g for 15 min to remove aggregated protein. Protein in the supernatant was determined by running the samples on an SDS-PAGE gel and relative protein levels were quantified as described earlier (see section 2.3.1.7 and 2.1.3.8).

### *2.3.2.3 Glycosylation assay*

Glycosylation of protein was examined using the pierce glycosylation staining kit method (Zacharius, Zell et al. 1969). In summary, protein samples were prepared with sample buffer and run on a Bio-rad stain-free gradient SDS PAGE gel. Horseradish peroxidase and soybean trypsin inhibitor provided in the kit were used as positive and negative controls, respectively. After running, the gel was washed with 50% methanol solution for 30 minutes followed by gel fixation with 3% acetic acid for 10mins. The gel was incubated with 25ml of oxidizing solution (provided in the kit) for 15 minutes on a rocking plate. The gel was washed with a 3% acetic acid solution. The gel was then transferred into the glycoprotein stain (provided in the kit) for 15 minutes. The staining solution was removed and gel was placed in the reducing agent (provided in the kit) for 5 min with gentle agitation. The gel was washed with 3% (v/v) acetic acid solution then by ultrapure water. Glycosylated proteins appeared magenta stained bands.

The type of N-glycosylation was probed using the PNGase and EndoH treatment which can cleave the N-linked glycans completely (Mellors and Sutherland 1994). The cleavage of

glycosylation can be observed as a mobility shift on the SDS-PAGE. EndoH and PNGase protocols were performed with slight modification to those described in the manufacturer protocol. 5µg protein was incubated with EndoH and PNGase in sample buffer for eight hours at room temperature. The sample was centrifuged at 100,000 x g for 15 minutes to remove any aggregation and analyzed using SDS-PAGE followed by immunoblotting as explained earlier.

#### *2.3.2.4 Drug binding assay*

Drug binding and mouse ABCB1a thermostability were measured in three ways using an UNCLE (Unchained laboratories) instrument: Concentrated ABCB1a protein (1µg per sample) was diluted into 10µl of buffer (100mM Tris-HCl pH 8, 150mM NaCl, 10% glycerol, 0.1% DDM and 0.02% CHS) containing different concentrations of the drug. The experiments were also repeated with the further addition of 100ng of CPM dye (7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin) which acts as a reporter of solvent-exposed cysteine residues (Kohlstaedt, von der Hocht et al. 2015, Lingam 2017). The mixtures were loaded into capillaries at 4°C and the latter was inserted into pre-chilled 16 well UNCLE capillary cassettes. Fluorescence emission spectra were recorded between 200nm and 700nm with an excitation wavelength of 266nm. Static light scattering was recorded simultaneously using the 266nm laser source. Data was gathered between 16°C and 90°C with a stepwise increase in the temperature (2°C increments) followed by the recording of spectra for all the samples. The total heating run required approximately 70min. Raw data was exported and then analysed for tryptophan fluorescence, static light scattering and CPM fluorescence using the GraphPad Prism software package.

#### 2.3.2.5 *ATP binding assay*

ATP binding assay was determined using TNP-ATP fluorescence as described by (LaConte, Srivastava et al. 2017) TNP-ATP alone is non-fluorescent however when it binds to the nucleotide-binding domain it emits fluorescence of wavelength between 500-600nm upon excitation. In brief, 5µg protein was incubated with different concentrations of TNP-ATP in a 96 well plate. Fluorescence emission spectra were recorded using Biotek multi-mode spectrophotometer. The sample was excited at 410nm, and the emission spectra were recorded at 500, 510, 520, 530, 540, 550 and 560nm. Data was analyzed and plotted using GraphPad prism.

### 2.3.3 Structural studies

#### 2.3.3.1 *Negative staining electron microscopy*

Initial structural validations were performed using negative stain electron microscopy. The specimen was prepared using the method as described previously by Booth *et al* with minor modifications (Booth, Avila-Sakar et al. 2011). 300-mesh carbon-coated copper grids were glow discharged under partial vacuum using a K100X glow discharge system at 25mA for 45 seconds. Purified protein was diluted to 10-50µg/ml in SEC buffer and applied to the grid towards the carbon-coated side. After 2 minutes excess protein was blotted away. The grid was washed with ultrapure water and excess water was removed using blotting paper. After washing the grid two to three times the grid was applied to a 5% uranyl acetate solution for 30 seconds. The excess staining solution was removed with blotting paper (Grids can be stored in a grid box at room temperature until use). Grids were screened using Thermo Fisher Talos L120C transmission electron microscope with Ceta CMOS camera (EM facility, FBMH)

### 2.3.3.2 *Small Angle X-ray Scattering (SAXS) analysis*

SAXS analysis was performed at beamline B21 at Diamond light source (Didcot UK) using the automated batch mode. The sample was exposed with radiation of  $10^{12}$  photons/sec and scattering was recorded on an Eiger 4M detector with a fixed camera length of 4.014m at 12.4keV. The data was collected between angular  $q$  ranges of 0.0035 to  $0.37 \text{ \AA}^{-1}$ . The data was recorded in an NXS file format using the Generic Data acquisition software and converted to 1D (.dat) profiles with Data Analysis WorkbeNch (DAWN)

The data was analysed using ScÅtter (Forster, Apostol et al. 2010) and ATSAS (Franke, Petoukhov et al. 2017). The basic data reduction and subtraction, primary analysis of radius of gyration ( $R_g$ ) flexibility and pairwise distance distribution was performed in ScÅtter. Data extrapolation and low-resolution structural analysis were performed in ATSAS. Comparison of theoretical SAXS curves and experimental curves was performed with the FoXS algorithm (Schneidman-Duhovny, Hammel et al. 2010, Schneidman-Duhovny, Hammel et al. 2013) implemented in the Chimera software suite (Pettersen, Goddard et al. 2004).

### 2.3.3.3 *Cryo-electron microscopy (cryo-EM)*

#### 2.3.3.3.1 Grid preparation

Cryo specimen were prepared as described earlier by freezing biological samples in vitreous ice using liquid ethane as a rapid coolant (Thonghin, Collins et al. 2018). In summary, 200 mesh 1.2/1.3 carbon-coated copper grids were washed with chloroform and left overnight to dry. The sample was exchanged into glycerol and detergent-free buffer (20mM Tris (pH 8), 20mM NaCl) by diafiltration using 100 kDa cut-off Vivaspin500 centrifugal filter before applying protein solution onto the grid. The sample was supplemented with drugs and/or ATP and incubated on ice for 30 min. Before loading and freezing, grids were glow discharged at

25mA for 2 minutes using a K100X glow discharge system. The glow discharged grid was clipped onto Dumont tweezers and loaded into FEI vitrobot Mark IV. The chamber was maintained at 95 percent humidity and 20°C. 3µl of 2.5mg/ml sample was applied on the carbon side of the grid. Excess liquid was blotted off with blotting strength 2. Different blotting times from 3-6 seconds were screened to get thin and even ice on the grids. The grids were plunged immediately into liquid ethane. The grids were stored in grid boxes and stored in liquid nitrogen until further screening or data collection.

#### 2.3.3.3.2 Data acquisition

Data was collected on an in-house Thermo Fisher Glacios Cryo-TEM. Grid loading and preliminary settings were performed by Dr Richard Collins (senior experimental scientist, FBMH). Data was collected on 190,000 X magnification corresponding to 0.77 Å per pixel size. the data was recorded using an automated data collection software EPU (FEI company) with a high electron dose of 100e with a defocus value between -2 to -3 µm.

#### 2.3.3.3.3 Single-particle analysis

The data was motion-corrected using the motionCor2 algorithm (Zheng, Palovcak et al. 2017). The images were further compressed for data handling and processing. The compressed, motion-corrected images were imported into cisTEM (Grant, Rohou et al. 2018) for particle picking and classification. The motion-corrected images were run for CTF correction. Images were manually screened for CTF fit and images with excessive carbon backgrounds, thick ice, or contamination were removed from the dataset. The screened dataset of images was used for *ab-initio* particle picking implemented in CistEM. P-glycoprotein has a maximum particle size of 140Å. Therefore, particles with a characteristic radius of 68Å and maximum radius of 72Å were picked. The particle search threshold was set at 5 with an interparticle distance of

100 pixels. The particles were extracted and 2D classified and averaged. Classes with empty micelles and showing carbon support film were eliminated during subset selection and the remaining particles were used in a subsequent round of 2D classification. The 2D classification was repeated until bad particles were eliminated and final dataset contained only P-gp particles. The final set of particles was used in 3D classification and generation of three-dimensional density maps.

## 2.3.4 Computational studies

### 2.3.4.1 *Computational Studies on P-glycoprotein (ABCB1)*

#### 2.3.4.1.1 Ivacaftor-P-gp interactions studies

Ivacaftor interaction studies were performed using Gold suite version 2020.3.0 (Jones, Willett et al. 1997, Verdonk, Cole et al. 2003). In-brief, in-house, generated mouse P-gp, crystal structure of mouse P-gp in apo state (PDBID: 4KSB) (Ward, Szewczyk et al. 2013), and structure of human P-gp bound with vincristine (PDBID: 7A69) were prepared by using the structure preparation module implemented in MOE (ULC 2019) followed by protonation at physiological pH (7.4). Two-dimensional structure files of ivacaftor and DDM were downloaded from the PubChem database (Kim, Chen et al. 2021) and were imported into the MOE database. Partial charges were calculated using MMFF94x (Halgren 1996). The charge calculated ligands were energy minimized to obtain a three-dimensional conformation. Ligands and proteins were imported in the GOLD docking wizard. To explore conformational diversity whole transmembrane region was considered as the binding cavity. Both protein and ligands were considered flexible during the docking simulations. To ensure reproducibility as well as to maintain the computational cost, a total of 100 conformations were generated using standard accurate (slow) calculations and ranked according to their scoring functions.

#### 2.3.4.1.2 Virtual screening

Virtual screening experiments were performed using flexible docking implemented in the GOLD suite (Verdonk, Cole et al. 2003, Cole, Nissink et al. 2005). Structures of human P-glycoprotein (PDBIDs: 7A69) bound with ligand (Nosol, Romane et al. 2020) was obtained from the PDB database (Berman, Westbrook et al. 2000). The structure was imported in MOE (ULC 2019) and corrected for structural discrepancies using the structure correction module. Furthermore, additional antibody chains and cholesterol molecules were deleted from the structure files. The corrected structure was protonated at physiological pH (7.4) and saved as mol2 file. A dataset of known P-gp substrates and non-substrate from literature was developed. The compounds were imported in MOE and prepared using MMFF94x followed by local energy minimization to obtain relaxed three-dimensional conformations. Optimization of screening conditions and cut-off threshold values were determined and tested using the known dataset.

#### 2.3.4.2 Computational studies on ABCBG2 (BCRP)

##### 2.3.4.2.1 Docking of kinase inhibitors with ABCG2

Docking experiments were performed on the inward-facing EM structure of ABCG2 (PDB: 6ETI) (Jackson, Manolaridis et al. 2018) using MOE 2019v01 (ULC 2019). The protein structure was prepared by protonating at physiological pH (7.4) followed by partial charge calculations using default parameters in MOE. A site finder implemented in the software was employed to identify docking residues in the TMDs which was then manually selected to restrict the binding cavity to the solvent-accessible residues inside the binding cavity. Ligands were sketched using the builder application implemented in MOE which were then stochastically minimized for 3D conformations. Partial charges on ligands were calculated using the Gasteiger PEOE method (Gasteiger and Marsili 1980) implemented in the software. The docking experiment employed



the alpha triangle placement method along with London-dG (ULC 2019) as a scoring function. A total of 100 poses (without duplication) were generated for respective ligands. Each pose was then minimized and rescored using the GBVI/WSA-dG scoring function (Naïm, Bhat et al. 2007).

#### 2.3.4.2.2 Pharmacophore modelling

Pharmacophore models were developed using imatinib conformation obtained from Cryo-EM structure 6VXH (Orlando and Liao 2020) and topotecan conformation was obtained from 7NEZ (Kowal, Ni et al. 2021). The pharmacophore features were calculated using the pharmacophore query editor and screened against the known dataset to identify similar features and to separate the active molecule from the rest of the dataset. The final models was evaluated based on precision and accuracy as given below.

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

TP = true positive

TN= true negative

FP = false positive

FN=false negative

#### 2.3.4.2.3 Residue mutagenesis analysis

Residue mutagenesis scanning was performed on the inward-facing EM structure of ABCG2 (PDB: 6ETI) (Jackson, Manolaridis et al. 2018) using MOE v2019.01 (ULC 2019). the protein structure was prepared by protonating at physiological pH (7.4) followed by partial charge calculations using default parameters in MOE. Residues identified at the homodimer interface

were mutated using a residue mutagenesis scanning algorithm implemented in MOE. Each residue was mutated to alanine or aromatic amino acids and protein stability was calculated *In-silico*. ABCG2 structure consists of two monomers and to ensure that both residues from each monomer were mutated in the homodimer an additional condition of double mutagenesis was introduced.

#### 2.3.4.2.4 Virtual screening

Virtual screening experiments were performed using flexible docking implemented in the GOLD suite (Verdonk, Cole et al. 2003, Cole, Nissink et al. 2005). Structures of human ABCG2 (PDBIDs: 6ETI) bound with ligand (Jackson, Manolaridis et al. 2018) was obtained from the PDB database (Berman, Westbrook et al. 2000). The Structure was imported in MOE (ULC 2019) and corrected for structural discrepancies using the structure correction module. Furthermore, additional antibody chains and cholesterol molecules were deleted from the structure files. The corrected structure was protonated at physiological pH (7.4) and saved as mol2 files. A dataset of known ABCG2 substrates and non-substrates from literature was developed. The compounds were imported in MOE and prepared using MMFF94x followed by local energy minimization to obtain relaxed three-dimensional conformations. Optimization of conditions, and cut-off threshold values were determined and tested using the known dataset.

## Chapter 3: Protein Expression and Purification:

### 3.1 Expression of mouse P-glycoprotein

*Pichia pastoris* expressing the mouse P-glycoprotein was provided in a form of a frozen pellet by Prof. Ina Urbatsch. The cells were thawed on ice and plated on MGY-agar plates with 100µg/ml zeocin. The transformed cells were zeocin resistant and were able to grow on the zeocin selective plates. The cells were streaked from the plate and inoculated in a 100ml small scale culture as explained by Thonghin *et al* (Thonghin, Collins et al. 2018). The cells were induced with 0.5% methanol and harvested. To confirm the expression of mouse P-glycoprotein a small portion of the cells were ruptured using the small scale bead beating method as explained in the methodology section 2.3.1.2. The crude membrane was analysed for mouse P-glycoprotein expression using the western blotting analysis. Figure 3.1 shows the western blotting of mouse P-glycoprotein with anti-His antibody labelled with horseradish peroxidase (HRP). The remaining cells were aliquoted to make glycerol stocks for future large-scale expressions and purifications.

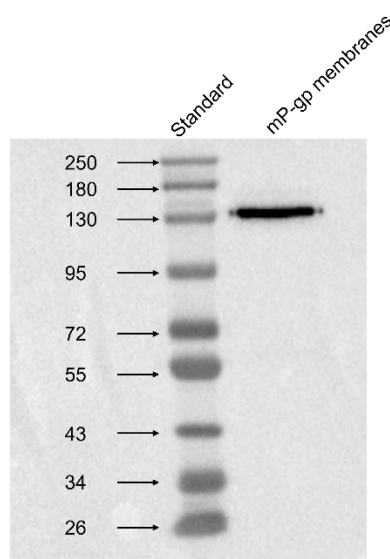


Figure 3.1. Validation of mouse P-glycoprotein expression in *Pichia pastoris*. Crude membranes were solubilized with 1% tween and analysed using western blotting.

For large scale expression, A shake-flask fed-batch approach was adopted as explained by (Lerner-Marmarosh, Gimi et al. 1999). The *Pichia* glycerol stocks were plated and incubated at 30°C until plates were covered completely with cells. A fresh MGY-agar plate was used for each run of expression and purification. The cells were streaked using a loop and inoculated into a 100ml MGY media with phosphate buffer at pH 6 (Filtered phosphate buffer was added after autoclaving the media). The culture was incubated overnight at 30°C. The following day the seed culture was equally divided into the main 6 x 1L cultures in 2L baffled flasks and incubated 24hours with orbital shaking of 225 RPM. A short booster feed of 5ml glycerol was added 5 hours prior to the first induction. Once the OD<sub>600</sub> is over 10, 5ml of 100% methanol induction was added and left on shaking at 30°C. Booster feed of 5ml methanol was added at 24hrs and 48hrs post-induction to increase the protein expression. Cells were harvested and wet cell weight was measured by subtracting the buffer.

The crude membrane was prepared using the bead beating method explained in section 2.3.1.4. The protein was solubilized using 2% DDM with final protein concentrations at 2.5mg/ml. Purification was performed using a 2-step protocol as explained by (Pollock, Cant et al. 2014). In the first step, The solubilized protein was purified using Ni-NTA chromatography followed by purification by size. Figure 3.2a shows the chromatogram of nickel affinity chromatography. The first peak represents the flow-through when the sample was applied to the column. In wash 1 the column is washed 8CV of elution buffer with 20mM imidazole to remove nonspecific interactions as well as the unbounded protein left during flow-through collection. Most of the impurities are washed at this point. The remaining loosely bound contaminants are eluted at wash two which corresponds to the second peak in the chromatogram. Pure mouse P-glycoprotein is eluted during the elution fraction represented

with A1-A5 on chromatogram between volume 40ml to 65 ml. The column was washed with 500mM imidazole to elute all protein out of the column.

The corresponding fractions obtained from the fraction collector were then analysed using the SDS-PAGE see figure 3.2b. mouse P-glycoprotein is represented by a 140KDa band on SDS\_PAGE (uniport ID: P21447) (Consortium 2021). The majority of contaminant proteins are eluted during the flow-through fractions and wash 1 fraction. However, loosely bound proteins were eluted at higher imidazole concentration in wash 2. Pure protein was purified at 200mM imidazole concentration. However, minor quantities of smaller bands are also visible in elution fractions. Western blotting delineates these bands as fragmented mouse P-glycoprotein with multiple bands of HRP chemiluminescence on the blotting membrane. Furthermore, when comparing the amount of mouse P-glycoprotein present in the soluble fraction with the amount of mouse P-glycoprotein in flow-through collected, Ni-NTA affinity chromatography shows approx. 80% efficiency in protein retention. This efficiency remained comparable in each run. For a few minor occasions due to in-complete washing of the Ni-NTA column, internal Pichia protein was also co-purified (data not shown).

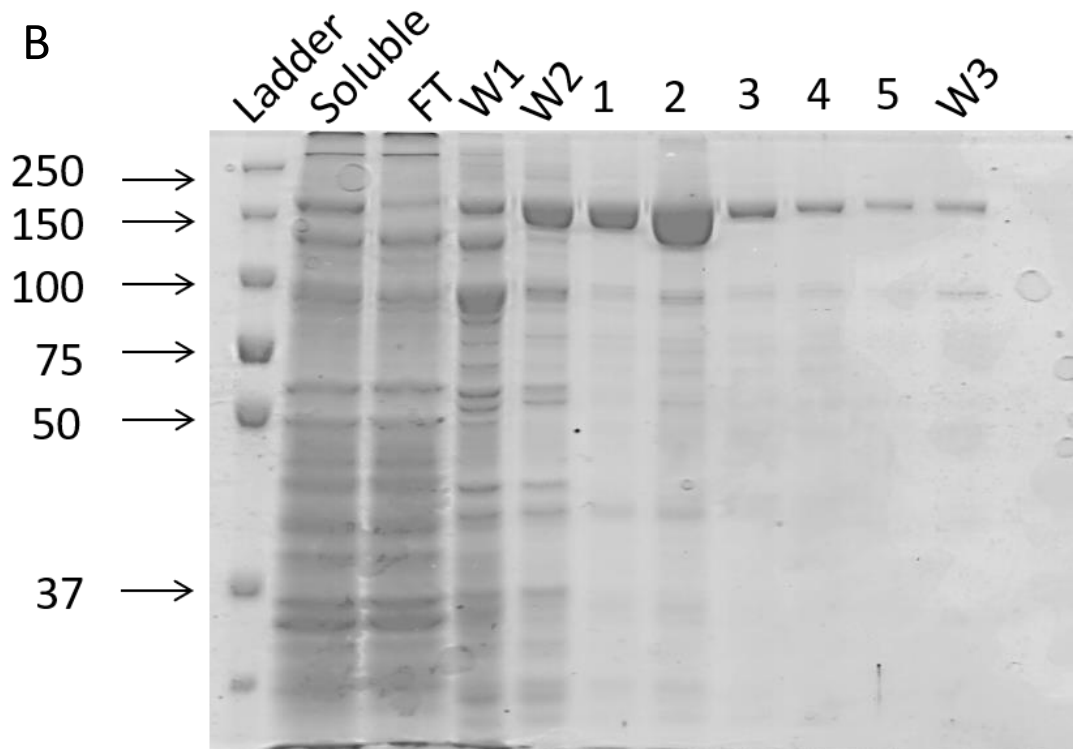
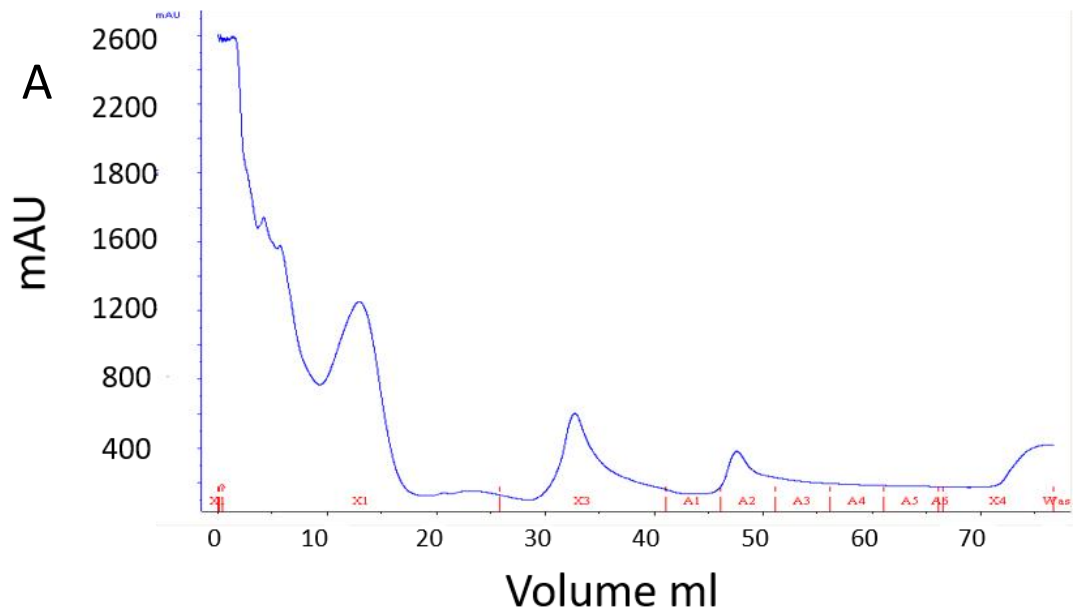
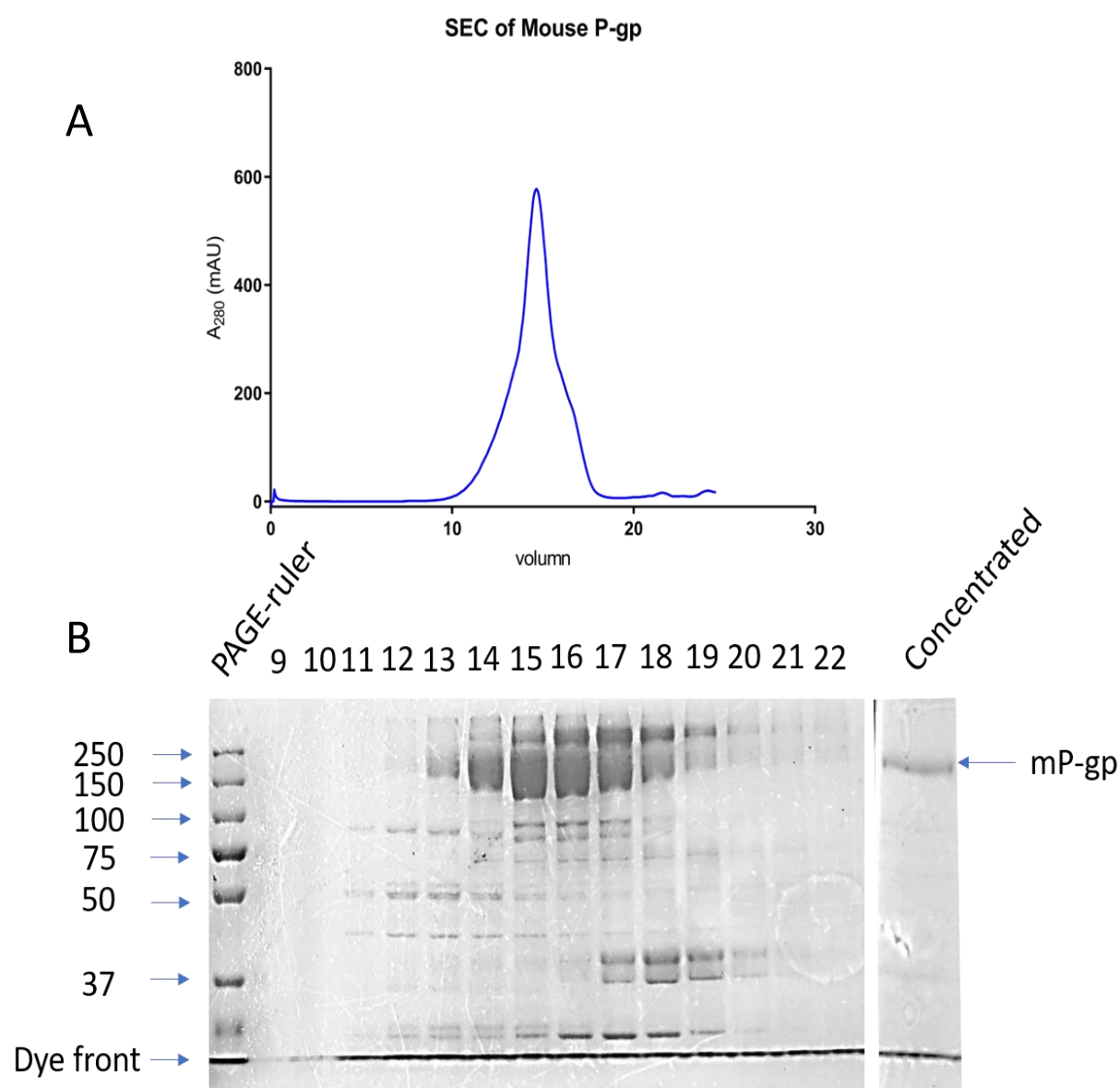


Figure 3.2. Nickel affinity purification of mouse P-glycoprotein. (A) Representative chromatogram of Ni-affinity chromatography. Y-axis represents the UV absorbance of protein at 280nm, employed to track the protein movement through the column whereas the X-axis represents the volume in ml. protein is eluted during fractions A1 to A5. (B) Representative image of corresponding fractions on SDS-PAGE under Coomassie Blue stain.

The elution fractions were pooled and diluted with purification buffer A to reduce the imidazole concentrations from 200mM to less than 50mM. The diluted fractions were then concentrated using 100kDa cut-off vivaspin20 concentrators to a final volume of 1-2ml (from approximately 100ml) or to the point where further concentration caused aggregation in the concentrator.

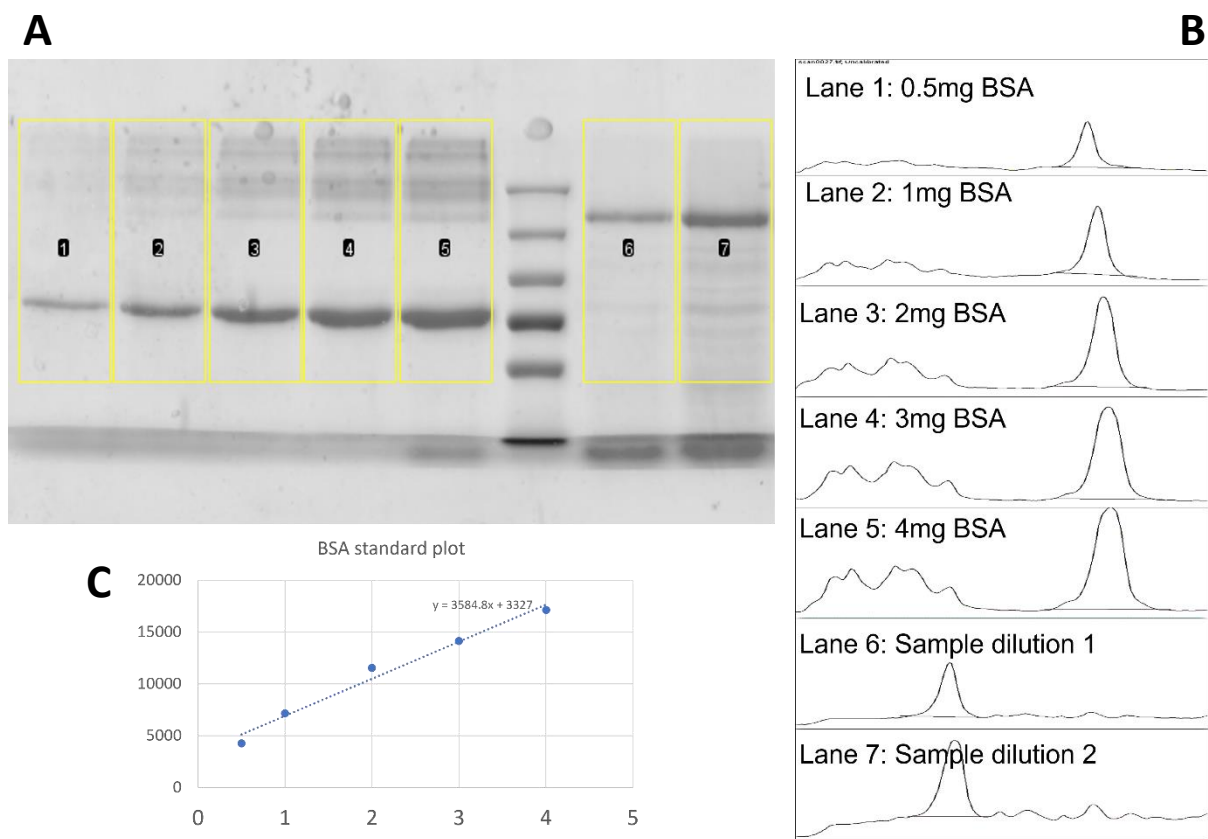
The second step of purification was SEC, which separates the protein on the basis of size. 500µl of concentrated protein was centrifuged as explained in the methodology section 2.3.1.6 and the sample was loaded onto the Generon SepFast 6-5000kDa column. Purification was performed for 1.2 CV at a flow rate of 0.4-0.5 ml/minutes. No protein was detected in the void volume suggesting no aggregation in the sample see Figure 3.3. Protein was detected after ~10ml. Mouse P-glycoprotein was detected in a single broad peak between 10ml to 18ml. On a few occasions, the chromatograms showed a strong peak between 9ml to 13 ml fractions and a large shoulder between 14ml to 18ml suggesting aggregation of the protein. this problem was common when the temperature during the bead beating or purification was not properly maintained. Fractions 9 to 22 were further analysed using SDS PAGE (see figure 3.3b), fractions 9 and 10 were void volumes. The protein started to appear at fraction 11 and maximum protein was eluted in fractions 14 to 17. SDS-PAGE analysis also showed the presence of aggregates at the start of the resolving gel. This aggregation is not present in earlier fractions and starts to appear in fractions 14 and later. This may be due to the overloading of the gel with purified protein in the later fractions. Fractions with purified protein were pooled and concentrated using Vivaspin6 100kDa cut-off filters. The lower molecular mass protein contaminants were passed through the membrane leaving >95% pure

protein in the concentrator. The concentrated protein was analysed for purity and quantification on SDS PAGE (see Fig 3.4).



*Figure 3.3. SEC purification of mouse P-glycoprotein. (A) Representative chromatogram of mouse P-glycoprotein SEC. the X-axis represents fractions in terms of volume. Each fraction represents 1ml of buffer collected through the fraction collector. The Y-axis represent the UV absorbance of protein at wavelength 280nm. protein elutes between fractions 10 to 20. (B) Representative SDS-PAGE of factions collected through SEC. The purified protein is pooled and concentrated and analysed for purity.*





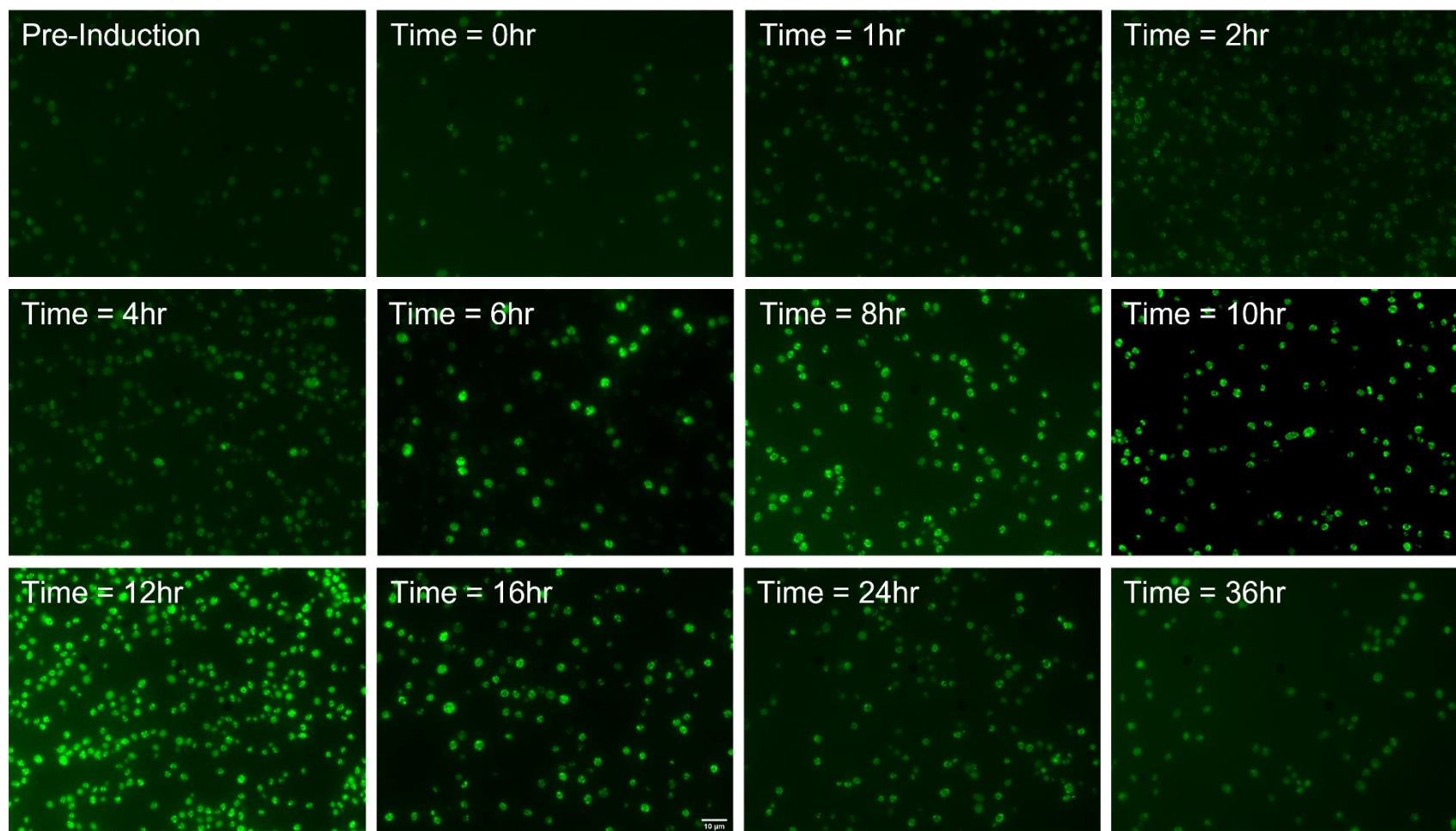
**Figure 3.4 Schematic diagram of protein quantification using ImageJ Package. (A) Lane selection of standard BSA concentrations and unknown concentration sample in ImageJ (B) Histogram plot for signal intensity from different concentrations of BSA standard and unknown sample. Area under the peak for each concentration was calculated from the histogram plot. (C) Scatter plot of signal intensity BSA standards as a function of concentrations. 1 $\mu$ l of each sample dilutions was mixed with the loading dye and loaded in the well. The concentration of the purified protein was calculated from the linear slope function.**

## 3.2 Expression of human P-glycoprotein

### 3.2.1 Time course experiment

A Time course experiment was performed to determine the optimal expression time for human P-glycoprotein fused with green fluorescent protein (GFP) (hP-gp-GFP) in *S cerevisiae*. Yeast was grown in YNB culture media with 0.2% glucose in a 2L baffled flask. The media was left overnight at 30°C with shaking at 225 RPM or until the glucose was consumed. Cells were induced with galactose and glycerol as explained in the methodology section 2.3.1.3. 50 ml of culture were collected at each time point and harvested in mPIB buffer with protease inhibitor cocktail. The cells were fixed on a glass slide using poly-lysine and analysed with an Olympus BX60 microscope with a GFP filter as described in section 2.3.1.3. Yeast cells showed equally distributed internal signals under the blue light filter used for GFP fluorescence (Figure 3.5, pre-induction) probably due to internally produced intrinsically fluorescent proteins such as succinate dehydrogenase and other FAD-containing proteins (Frey, Sott et al. 2011). After induction, the GFP signal intensifies. Interestingly, yeast cells start expressing human P-glycoprotein even at the first hour after induction (see Figure 3.5, time =1hr) (Drew, Newstead et al. 2008). Yeast has its internal endoplasmic reticulum-associated degradation (ERAD) pathway for protein quality control for membrane proteins. Misfolded proteins are trapped in the endoplasmic reticulum and degraded whereas only properly folded proteins can reach the plasma membrane (PM) (Brodsky and Skach 2011). Therefore, the expression of the GFP tag signal can be an indirect indicator of heterologously expressed membrane protein folding (Drew, von Heijne et al. 2001, Newstead, Kim et al. 2007). At time point 6 hours the expression of hP-gp-GFP becomes prominent. The expression reaches a maximum at 12 hours time point and starts to decrease at the time point of 16 hours. This decrease of signal continues and at time point 36 hours and 48 hours post-induction almost reaches the starting levels. When the

graph of corrected total cell fluorescence was plotted against time, peak expression was observed between 10 to 16 hours (Figure 3.6). This time is similar to the time point observed by Dr Lingam (Lingam 2017). However, she reported maximum expression at 10 hours whereas in this study it was observed at approx. 12 hours. This difference can be explained on the basis of the technique used for quantification. At the 10<sup>th</sup> hour, the maximum expression is present at the peripheral membrane. At the 12<sup>th</sup> hour, the GFP signals are located inside the cytoplasm in addition to the peripheral localization. This suggests the early start of protein internalization and degradation. However, during microscopy, the GFP signal was recorded. This needs to be validated using confocal microscopy.



*Figure 3.5 Time course experiment of human P-glycoprotein fused with GFP (hP-gp-GFP) expressed in yeast.*

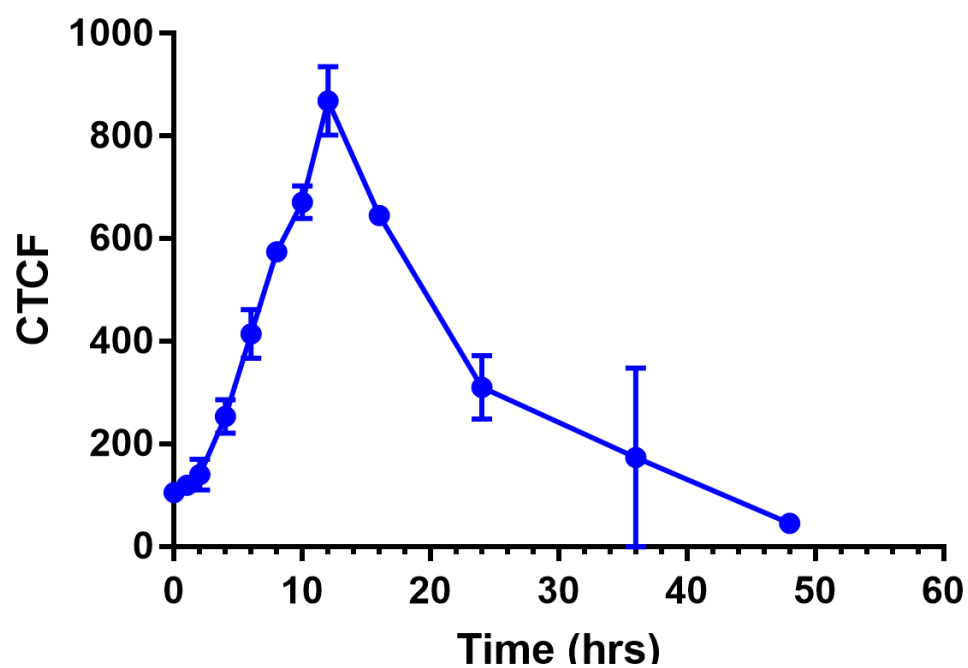


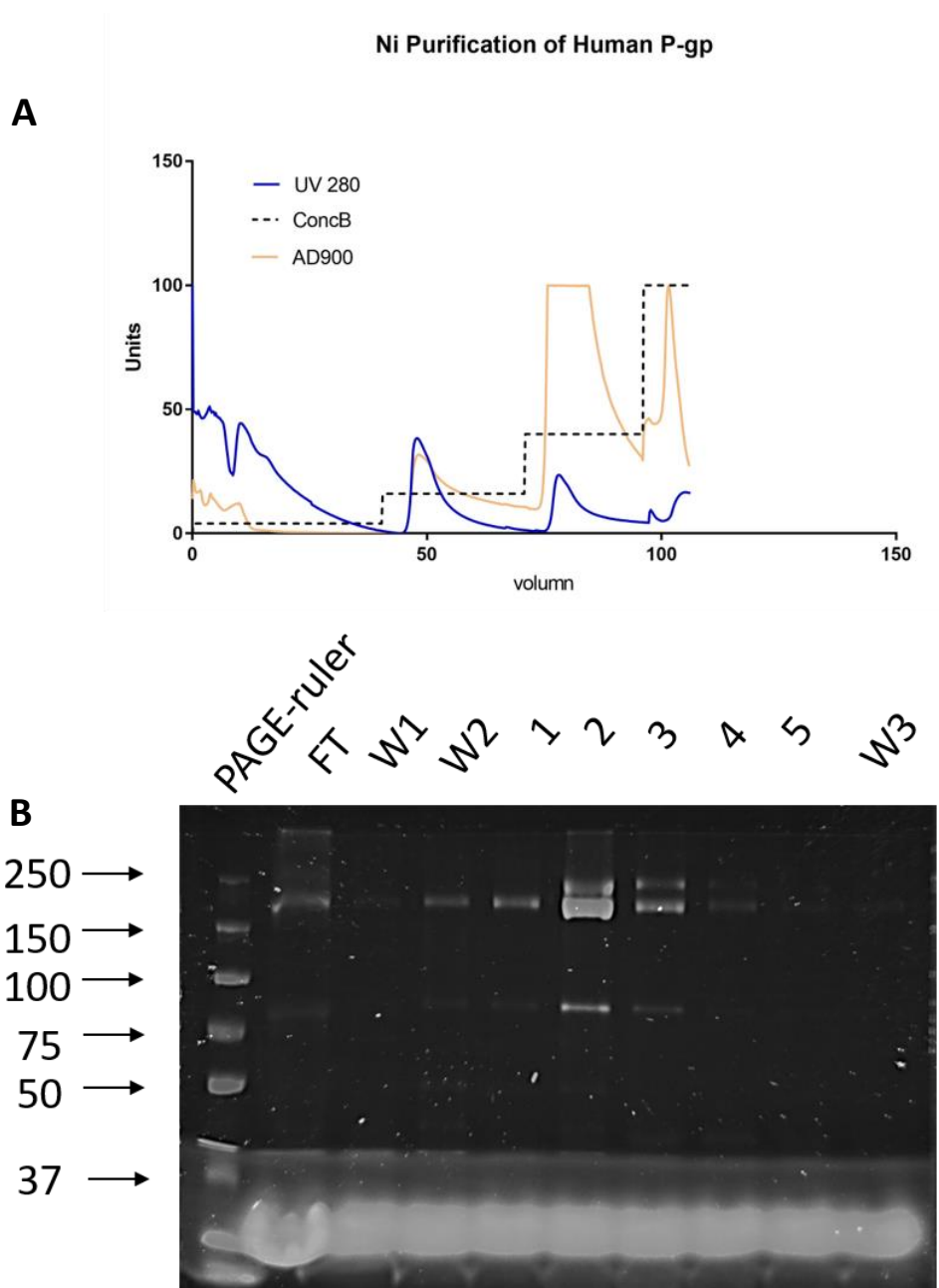
Figure 3.6. Graphical representation of time course experiment for human P-glycoprotein (hP-gp-GFP). The images were imported Into ImageJ and calculated for Corrected Total Cell Flouresence (CTCF). i.e. Total signal of GFP from single cell after subtracting the background noise. Human P-glycoprotein expression peaks at 12 hours

### 3.2.2 Crude membrane preparation and purification

Large scale cultures were performed in a 20L fermenter as explained by (Lingam 2017) and (Pollock, Cant et al. 2014). The media was pumped into a sterile fermenter and added with seed culture. The OD600 was measured in periodical intervals until an optical density of 2 was achieved. Once the glucose was consumed, protein expression was induced by the addition of glycerol and galactose. The cells were harvested between 10hrs to 12hrs of induction as explained in the previous section. Approximately  $100 \pm 15$  grams of cells (approx. 6g/L) was harvested per round of the 18L fermenter. The quantity of cells harvested is dependent on nutrient distribution and amount of aeration through the culture. The number of cells harvested in 1L cultures prepared in a 2L baffled flask was higher (approx. 8g/L) as compared to the 18L fermenter. However, due to pragmatic limitations of protein expression in 1L cultures, an 18L fermenter was the preferred protocol. The harvested cells were resuspended in mPIB buffer supplemented with a protease cocktail to inhibit the yeast proteases. The crude membranes were prepared using a bead-beating protocol as explained (see section 2.3.1.4), and dissolved in solubilization buffer without the detergent. On average 200-250mg of total protein was obtained per 80 grams (maximum amount of cells that can be broken in one round of bead beating) of cells broken.

The prepared membranes were solubilized using DDM as explained in methodology section 2.3.1.4. The protein sample was applied onto a 5ml Ni-NTA column with a flow rate of 0.5ml/minute. The flowthrough was collected and further used in efficiency analysis. 80% of the recombinant protein (hP-gp-GFP) was recovered after the Ni-affinity chromatography. The GFP tracking shows that a small amount of human P-glycoprotein was eluted in wash 2 (see Figure 3.7A). Most human P-glycoprotein was eluted at 200mM imidazole concentration. The

fractions were diluted with purification buffer without the imidazole to reduce the imidazole concentrations.



*Figure 3.7.* Ni affinity chromatography of human P-glycoprotein (hP-gp-GFP). (A) A representative affinity chromatogram of nickel affinity chromatography. The X-axis represents volume whereas Y-axis represents normalized absorbance. GFP tracking (yellow) and UV280 absorbance were employed to track the elution of human P-glycoprotein. (B) representative gel analysis of Ni-NTA fractions on SDS-PAGE. The gel was analysed for GFP fluorescence under visible excitation range of 460-490 nm and Emission range of 518-546 nm and excitation wavelength of 625-650 nm and emission range of 675-725 nm wavelength for biorad prestained marker.

The purified protein was concentrated using 100kDa cut off concentrator to a final volume of 500 $\mu$ l and loaded onto the SEC column with a flow rate of 0.4ml/minute. The SEC chromatograms with GFP tracking shows that P-glycoprotein elutes between fraction 8 to fraction 18 (see Figure 3.8A). However, the chromatogram of UV<sub>280</sub> shows a sharp peak between fraction 8 to fraction 10 and a wider peak between fraction 11 to fraction 18. When analysed on SDS-PAGE with Coomassie Blue staining the sharp peak between fraction 10 to fraction 13 shows aggregated protein (Figure 3.8B). Whereas a broad shoulder peak observed between fractions 13 to 16 showed monomeric protein. A minor amount of small fragments were also observed which may delineate partial degradation during purification. These minor bands were further reduced during filtration concentration. Finally, the monomeric fractions were pooled and concentrated using 100kDa cut off concentrator to a final concentration of 1-2mg/ml estimated via GFP fluorescence spectrophotometry. The concentration was further validated using gel densitometry methods on SDS\_PAGE. The final yield of human P-glycoprotein (hP-gp-GFP) was approximately 300-400 $\mu$ g from an 18L culture.



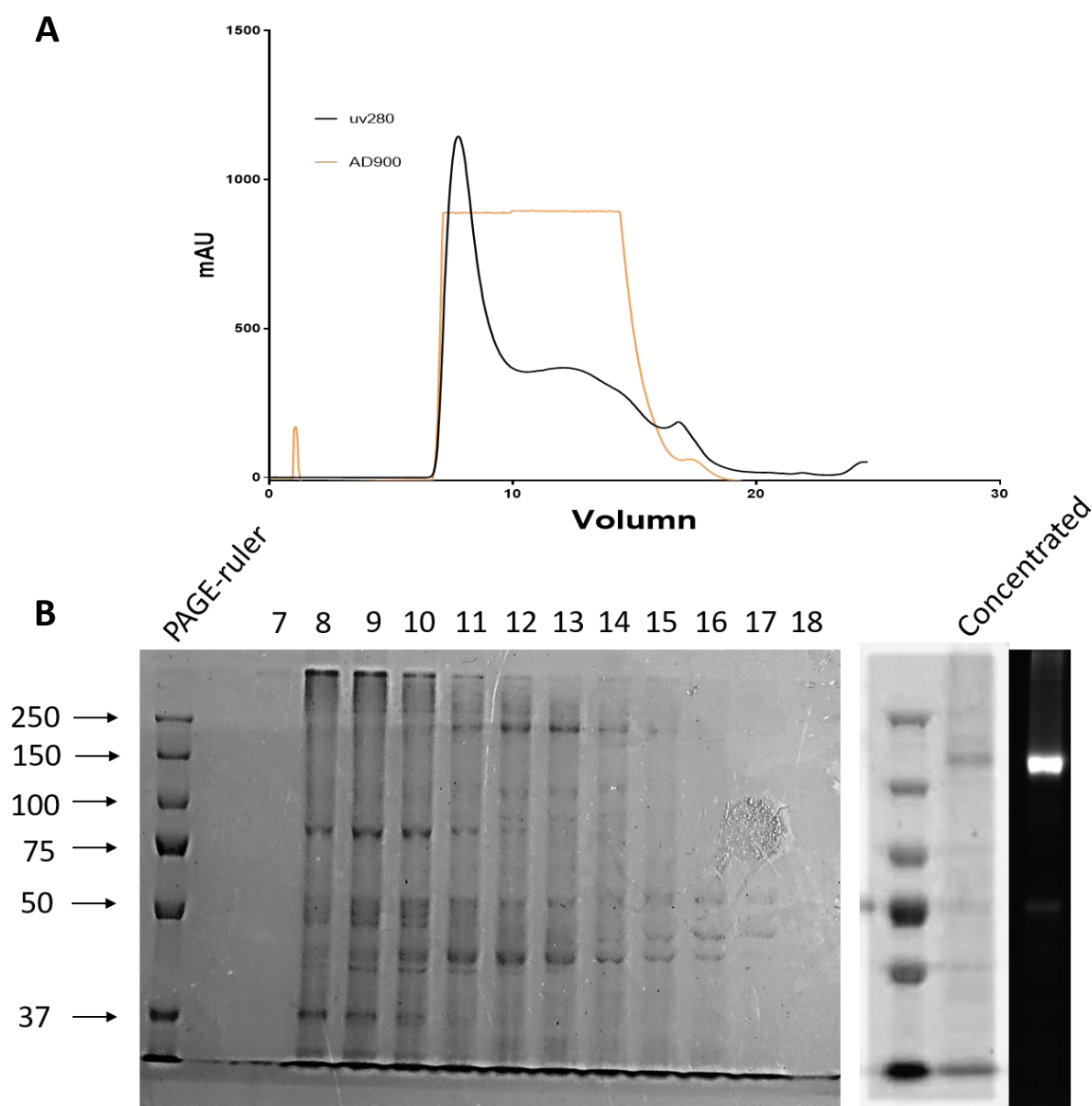


Figure 3.8. Size exclusion chromatography of Human P-glycoprotein (hP-gp-GFP). (A) Representative chromatogram of human P-glycoprotein. The X-axis represent the volume whereas, Y-axis represents absorbance at UV280 and GFP fluorescence. (B) Representative Coomassie Blue stained gel. the left panel showed fractions collected during SEC. The right panel showed concentrated human P-glycoprotein stained with Coomassie Blue and under visible excitation wavelength of 460-490 nm and emission wavelength of 518-546 nm in a Chemidoc MP system.

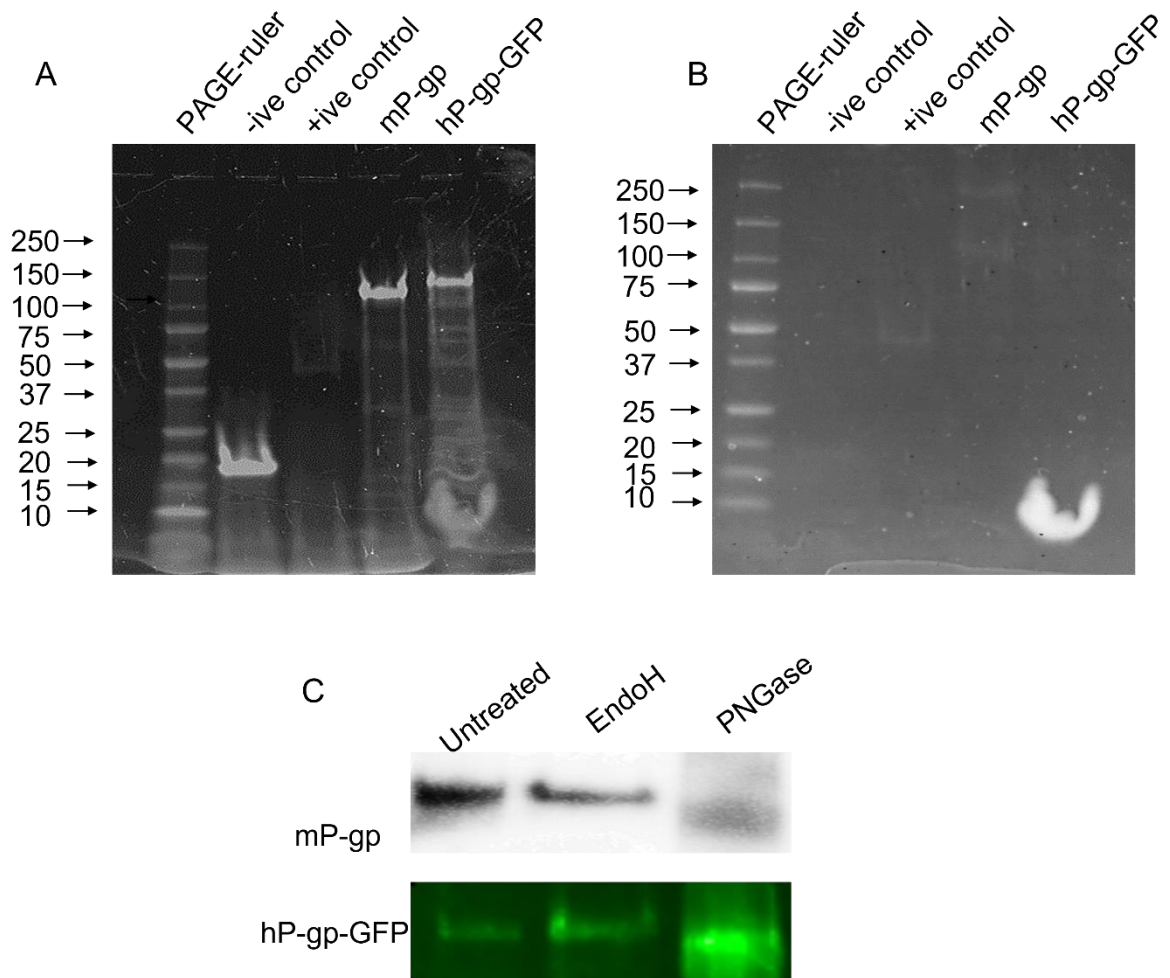
### 3.3 Characterization of protein

#### 3.3.1 Post-translational modifications

During protein synthesis, P-glycoprotein undergoes post-translational modifications that are important for protein stability, folding and transportation to the cell surface via ER and Golgi pathways (Czuba, Hillgren et al. 2018). Yeast can perform many eukaryotic PTMs but may not correctly process the heterologously expressed proteins. Therefore, it is important to identify the PTMs on our heterologously expressed protein. Firstly, the glycosylation state was analysed using the glycoprotein staining kit (Pierce glycosylation staining Kit). Figure 3.9B shows SDS-PAGE stained with acidic Fuchsin sulphite stain. Fuchsin stain can bind to oxidized carbohydrates to demonstrate the presence of glycosylation on the protein. the soya trypsin inhibitor protein was used as a negative control. Mouse P-glycoprotein showed the presence of glycosylation with Fuchsin stain suggesting that *Pichia* can do post-translational glycosylation of P-glycoprotein. No stain was observed in human P-glycoprotein expressed in *S. cerevisiae*. However, the staining was very weak as compared to the amount of protein loaded on the gel (see Figure 3.9A) same gel observed under Stain-free filter on Bio-Rad Chemi-Doc) even in the case of the positive control. Therefore, additional verification through the enzymatic method was necessary.

To validate the results and to probe the type of glycosylation, the pure protein was treated with EndoH and PNGaseF. Both EndoH and PNGaseF can cleave most N-linked glycans (Freeze and Kranz 2010). EndoH can cleave higher mannose glycoforms whereas, PNGaseF work as glycoamidase cleaves almost all types of N glycosylations converting the Asn to Asp. Cleaving glycosylation can produce mobility shift on an SDS PAGE. Figure 3.9C. shows enzymatic cleavage of glycosylation from mouse and human P-glycoprotein. Untreated P-glycoprotein was used as a control to measure the shift. both human and mouse P-glycoprotein showed

mobility shift after PNGaseF treatment suggesting the presence of higher-order N glycosylation.



*Figure 3.9. Glycosylation of human and mouse P-glycoprotein, (A) Detection of total protein using the stain-free filter on chemidoc image analyser. (B) SDS-PAGE stained with fuchsin stain. (C) Mobility shift assay for detection of N-linked glycosylation. Mouse P-glycoprotein was detected with HRP chemiluminescence on BioRad chemidoc image analyser, whereas human P-glycoprotein was detected for green dye with Licor Odyssey Clx.*

### 3.3.2 ATP binding assay

Biochemical and structural evidence delineate that P-glycoprotein consist of two ATP binding domains that can trigger the catalytic cycle of substrate translocation. Therefore, the binding of ATP is a crucial feature of P-glycoprotein. A fluorescent analogue of ATP (TNP-ATP) as described by Francis J. Sharom was used to study the ATP binding (Qu, Russell et al. 2003). TNP-ATP shows a wide excitation spectrum ranging from 350 to 520nm with excitation maxima at 410nm and 470nm and emission spectrum between 500nm to 600nm.

In this study, an excitation peak of 410nm was used to avoid the overlapping GFP signal in human P-glycoprotein (hP-gp-GFP). The emission spectrum was recorded at periodic intervals between 510nm to 560nm. The binding assay was performed in black well clear bottom 96 well plates to avoid any cross signalling between the wells. Figure 3.10A shows the fluorescence of TNP-ATP in the presence of detergent and/or protein. BSA was used as a negative control for ATP binding. TNP-ATP in water shows an emission peak around 550nm to 560nm. This emission spectrum was similar to the emission spectrum in BSA. When P-glycoprotein was added to TNP-ATP, an increase in the fluorescence signal was observed moreover the emission peak was shifted from 560nm to 530-540nm. Interestingly, this increase of fluorescence and shifting of the peak was also observed in the sample containing only DDM buffer. However, the amount of fluorescence was less than the fluorescence observed in the case of P-glycoprotein. This was further investigated by using different TNP-ATP concentrations in the DDM buffer. The graph showed an increasing trend till 100 $\mu$ M TNP-ATP concentration (see Figure 3.10C). The graph remains constant till 200 $\mu$ M but starts to self quench at 400 $\mu$ M TNP-ATP concentration. This quenching can be explained through a wide excitation spectrum reaching 520nm which overlaps with the excitation spectrum. The emission spectrum can self-excite the TNP-ATP molecules.

The role of  $Mg^{2+}$  in ATP hydrolysis is well established in P-glycoprotein. But the role of  $Mg^{2+}$  in ATP binding has not been elaborated. Therefore, TNP-ATP binding with P-glycoprotein was performed in the presence and absence of  $Mg^{2+}$  ions (Figure 3.10C). The results showed that  $Mg^{2+}$  ions are critical in the binding of ATPs with the nucleotide-binding domains. Furthermore, TNP-ATP shows higher fluorescence in presence of  $Mg^{2+}$  ions in the solutions as compared to control where no  $Mg^{2+}$  ion was present which suggest that TNP-ATP fluorescence is highly affected by reaction mixture composition.

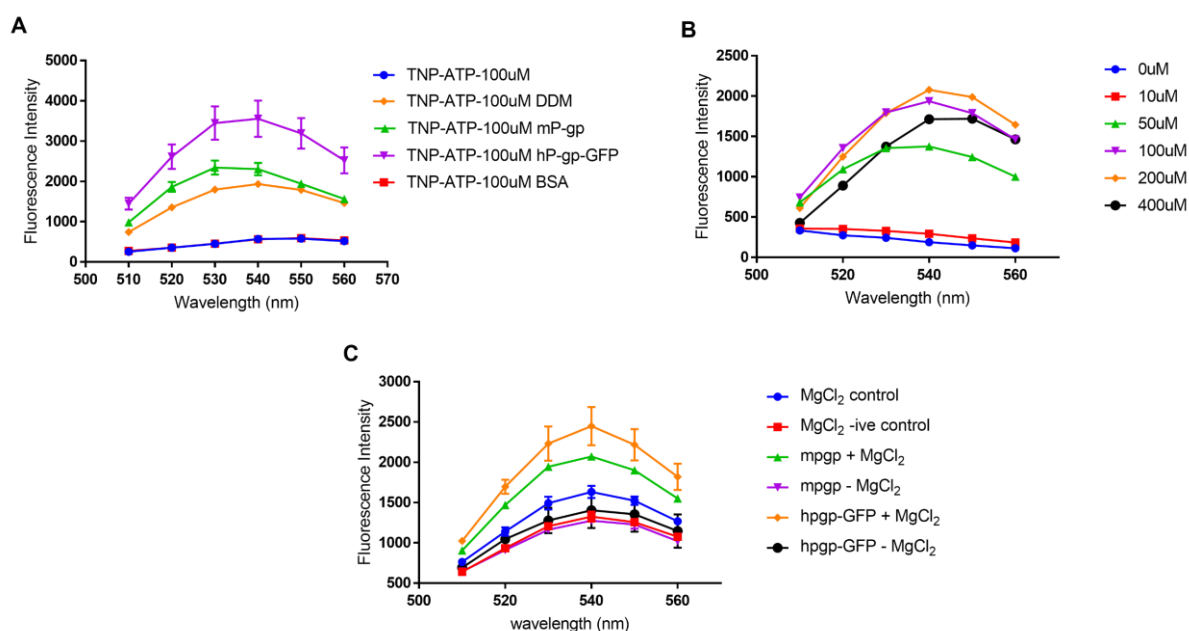


Figure 3.10. Graphical representation of TNP-ATP binding assay used for nucleotide binding. (A) Shows the binding of TNP-ATP with mouse P-glycoprotein and human P-glycoprotein (hP-gp-GFP). (B) Concentration-dependent analysis of TNP-ATP in DDM containing buffer. (C) Effect of  $Mg^{2+}$  ion in TNP-ATP (100μM) binding.

### 3.4 Discussion and conclusions

Yeast offers great advantages over conventional eukaryotic cell systems. They require less expertise as well as provide low-cost benefits at a large scale. Several studies have

demonstrated that they can also be efficient in expressing membrane proteins. During this study, we expressed and purified mouse and human P-glycoproteins (hPgP-GFP) using the *Pichia pastoris* and yeast expression system respectively. which were then used in structural studies and ligand-protein interactions studies. We also demonstrated that yeast can regulate post-translational modifications. These post-translational modifications can be vital in protein stability and functionality. Human P-glycoprotein tagged with green fluorescence protein (GFP) provided two advantages during expression first GFP signals can indicate the expression of the protein. Secondly, it also acted as a tracer during purification.

Temperature control during purification was a critical factor in protein quality control. Human P-glycoprotein has shown melting temperatures of 57°C (Lingam 2017). While the temperature in bead beating can go as high as 65-70°C during the two minutes bead beating cycle. Therefore, the bead mill was covered with dry ice during the whole bead beating process. Furthermore, the slurry was supplemented with protease inhibitors to block yeast intrinsic proteases which can degrade our recombinant protein. Previously two-step purification has shown more than 90% purity in CFTR analogues (Pollock, Cant et al. 2014). We also observed 95% pure protein after size exclusion chromatography. Negative stain microscopy has shown that mouse P-glycoprotein is relatively more homogenous in detergent-solubilized form. In contrast, human P-glycoprotein showed a mixture of monomeric and aggregated proteins (figure 3.10). Aggregation in human P-glycoprotein was also evident from the size exclusion chromatograms. Even after ultracentrifugation and repeated size exclusions runs. These results suggested an interparticle equilibrium between the aggregated and monomeric forms in the aqueous solutions. Another possibility is that

human P-glycoprotein is tagged with GFP. GFP has been reported to alter the interparticle interactions in recombinant proteins (Montecinos-Franjola, Bauer et al. 2020).

ATP binding and hydrolysis are vital activities for functional ABC transporters. TNP-ATP binding assay has been used to establish ATP binding in P-glycoprotein. Detergent purified P-glycoprotein can demonstrate an ATP binding capacity in the presence of  $Mg^{2+}$  ions. However, in absence of  $Mg^{2+}$ , there was no signal for TNP-ATP. Interestingly, this finding is in contrast with the findings of (Lusvarghi and Ambudkar 2019), where they reported that ATP can bind to NBDs even in non-hydrolytic conditions. The possible explanation can be that fluorescence derivate of ATP (TNP-ATP) is selective for  $Mg^{2+}$  ions for binding.

Purification of eukaryotic membrane protein is not trivial and requires highly regulated expression systems. In this chapter, we expressed and purified human and mouse P-glycoprotein from *Saccharomyces cerevisiae* strain FGY217, and *Pichia Pastoris* respectively. Both mouse and human P-glycoprotein showed a mixture of monomeric and aggregated P-glycoprotein after purification. Aggregation was more dominant in human P-glycoprotein whereas the mouse analogue was more stable in detergent-solubilized form. Our results demonstrated that yeast expression systems can be useful in expressing heterologous membrane proteins.

## Chapter 4: Small-Angle X-ray Scattering (SAXS) studies on P-glycoprotein catalytic states

Small Angle X-ray Scattering (SAXS) analysis is a powerful and robust technique to study biomolecules in solution (Svergun, Koch et al. 2013). Contrary to the crystallographic studies where crystal formation is the rate limiting factor, and cryo-EM analysis which are limited by the size of the proteins and challenges with the distribution of particles on the grid surface as well as within the amorphous ice. SAXS is a relatively easy technique that can be used to analyse proteins from a few kDa to GDa (Svergun, Koch et al. 2013). It can provide quantitative measurements of structural parameters like size and shapes of proteins in a solution. Furthermore, it can also provide information about the intrinsic disorder, dynamics and conformational polydispersity of proteins in a native state (Bernadó and Svergun 2012).

In the last decade, several snapshots of mouse and human P-glycoprotein have been reported (Li, Jaimes et al. 2014, Jackson, Manolaridis et al. 2018, Nosol, Romane et al. 2020). These snapshots have provided high-resolution information on ligand-protein interactions and catalytic states. Still, these studies were not able to explain the conformational dynamics of the protein. With recent advancements in cryo-EM technology, exploring these dynamics have been possible, however, this requires handling and processing of massive datasets. Moreover, these studies also depend on how stable different catalytic states are before capturing the image. P-glycoprotein is one of the examples of such a protein. Verhelen *et al* demonstrated that vanadate can trap the protein in the outward conformation (Verhalen, Dastvan et al. 2017). However, when similar conditions were used in cryo-EM and single particle analysis only an inward-facing state was observed (Thonghin, Collins et al. 2018). Which implies that Inward-facing states are very stable as compared to outward-facing states. Recently, the outward-facing structure of human P-glycoprotein was obtained by trapping the protein into



an outward occluded state by mutating the glutamates at 556 and 1201 to glutamine in the conserved Walker B sequence. These mutations render the proteins into a non-catalytic state (Kim and Chen 2018).

This study aimed to generate low-resolution structural profiles of P-glycoprotein using SAXS under different experimental conditions. These results could help in understanding the molecular dynamics of the protein.

#### 4.1 Preliminary analysis

For preliminary analysis, the structures of ABC transporters in different states pre-embedded in lipid bilayer were downloaded from the MemProtMD database (Stansfeld, Goose et al. 2015). For the inward-facing apo state, the crystallographic structure of mouse P-glycoprotein (4KSB) (Ward, Szewczyk et al. 2013) was used. The structure represents the canonical inverted 'V' like conformation open from the cytosolic side. For outward-facing conformations, the Sav1866 structure (2HYD) with TMDs open towards extracellular matrix (Dawson and Locher 2006) and Walker B mutant of human P-glycoprotein (6COV) in outward-occluded state were used (Kim and Chen 2018). To mimic the disc-like micelle, a 50Å disc (calculated from electron density map of mouse p-glycoprotein) of the bilayer was cropped around the structures. The embedded structure was used to generate the SAXS profile using the FoXS server (Schneidman-Duhovny, Hammel et al. 2010). Figure 4.1 shows the theoretical 2D profiles of different catalytic states of ABC transporters. The theoretical profiles show that both inward and outward-facing conformations show similar scattering profiles at longer scattering angles ( $s=0.3-0.5$ ) and smaller scattering angles ( $s=0-0.5$ ), however, the SAXS profile was able to distinguish between inward-facing and outward-facing conformations suggesting the robustness of SAXS in resolving different catalytic states.

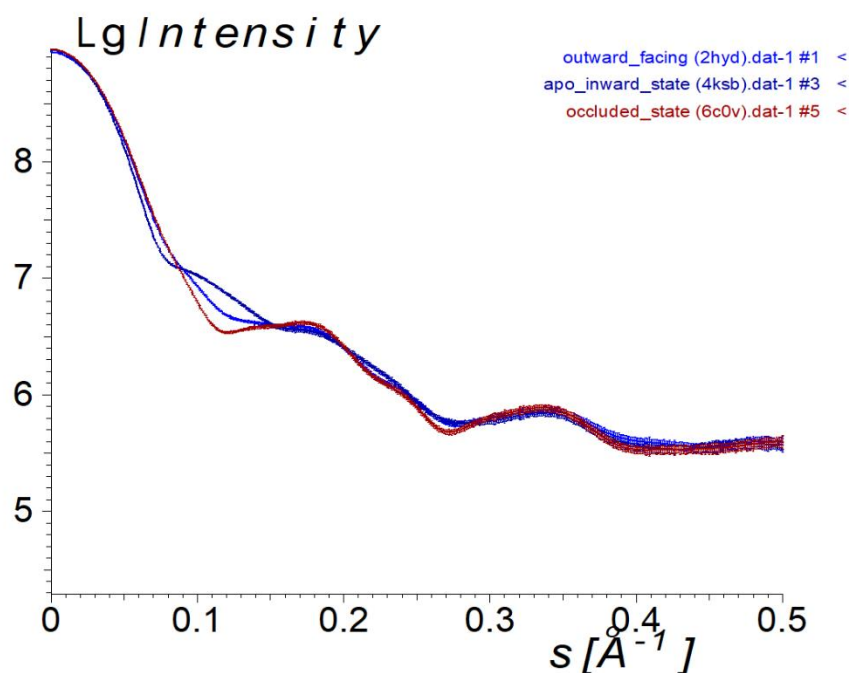


Figure 4.1 Theoretical scattering profile of ABC transporters in apo (PDBID 4KSB), outward (PDBID 2HYD) and occluded (PDBID 6C0V) as a function of log intensity (y-axis) and momentum transfer on x-axis ( $s=4\pi\sin\theta/\lambda=1/\text{\AA}$ ) generated using FoXS server.

#### 4.2 Experiment 1: P-glycoprotein in apo and ADP trapped state

SAXS data on mouse P-glycoprotein (mP-gp) and human P-glycoprotein fused with GFP (hPgp-GFP) were collected at Beamline 21 at Diamond Light Source (Harwell, UK). For the apo state P-glycoprotein at 5mg/ml in SEC buffer was used whereas for ADP-Vi trapped conditions (Verhalen, Dastvan et al. 2017), 5mg/ml protein was incubated with 2mM Na.ATP and Na<sub>3</sub>VO<sub>4</sub> for 30 min at room temperature. The data was collected in a batch mode with 10 frames per sample to ensure consistency and identify any radiation damage. For data processing, data analysis software ATSAS (Manalastas-Cantos, Konarev et al. 2021) from EMBL was used. Each frame was imported into the PRIMUS (Konarev, Volkov et al. 2003) interface and manually examined for frames with radiation damage. The frames were then averaged before buffer subtraction. Buffer without the ATP and vanadate was subtracted from human and mouse apo samples whereas, buffer with ATP and vanadate was used for subtraction in ADP trapped

sample. Figure 4.2 shows the 2D scattering profiles of human and mouse P-glycoprotein in apo and ADP trapped states.

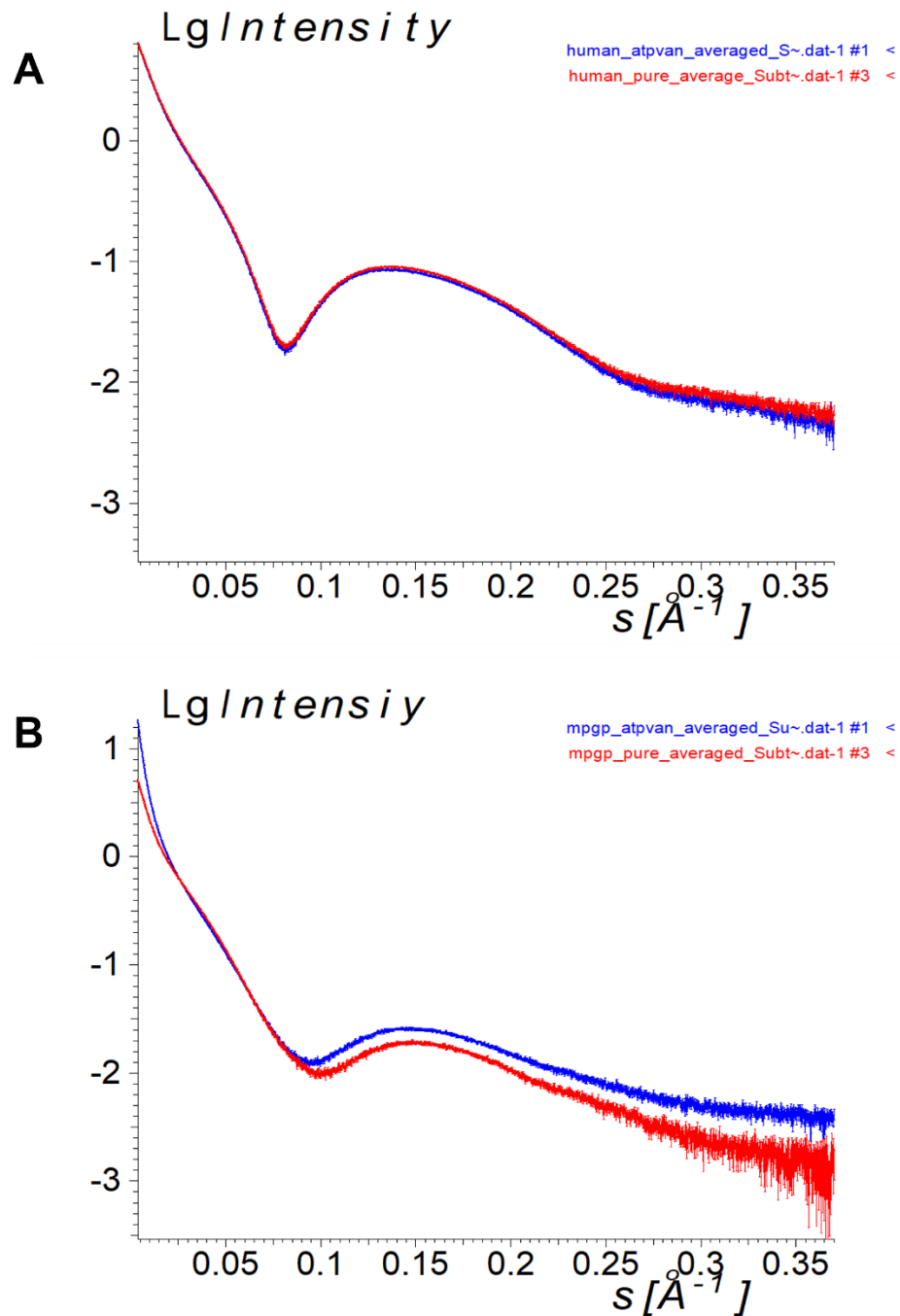


Figure 4.2 Scattering profiles of P-glycoprotein in apo and ADP trapped state. (A) Scattering profile of human P-glycoprotein fused with GFP (hP-gp-GFP). The red curve represents the scattering profile in apo states and the blue curve shows the ADP trapped state (B) Scattering profile of mouse P-glycoprotein. The x-axis represents the moment transfer ( $4\pi \sin \Theta / \lambda = 1/Å$ ) and the y-axis represent the intensity in log scale

Human P-glycoprotein in apo and ADP trapped states showed similar 2D scattering profiles. Whereas mouse P-glycoprotein showed the difference in apo and ADP trapped states. The scattering intensities at lower momentum transfers values suggest aggregation in the sample. This aggregation was also reported by (Shintre 2009) when they used protein at 6mg/ml for small-angle X-ray scattering data collection and 0.5mg/ml for wide-angle data. The radius of gyration ( $R_g$ ) calculated using the Guinier function also suggested aggregation. Figure 4.3 and Figure 4.4 shows the Guinier plot for the radius of gyration ( $R_g$ ) estimation for mouse and human P-glycoprotein respectively. Interestingly, both human and mouse P-glycoprotein showed a characteristic hump in the scattering profile. This feature was hypothesized due to the detergent micelle present around the protein. Therefore, theoretical profiles of P-glycoprotein with bicelle nanodisc as generated in the preliminary studies were compared with the theoretical profile of P-glycoprotein without the bicelle. Figure 4.5 shows the theoretical scattering profiles of P-glycoprotein without the bicelle and with the bicelle nanodisc in comparison with the experimental profile. The theoretical profile with bicelle also shows a characteristic hump similar to the curve observed in the experimental profile. A similar scattering profile was also recorded with aquaporin0 solubilized with  $\beta$ DDM (Molodenskiy, Mertens et al. 2020).

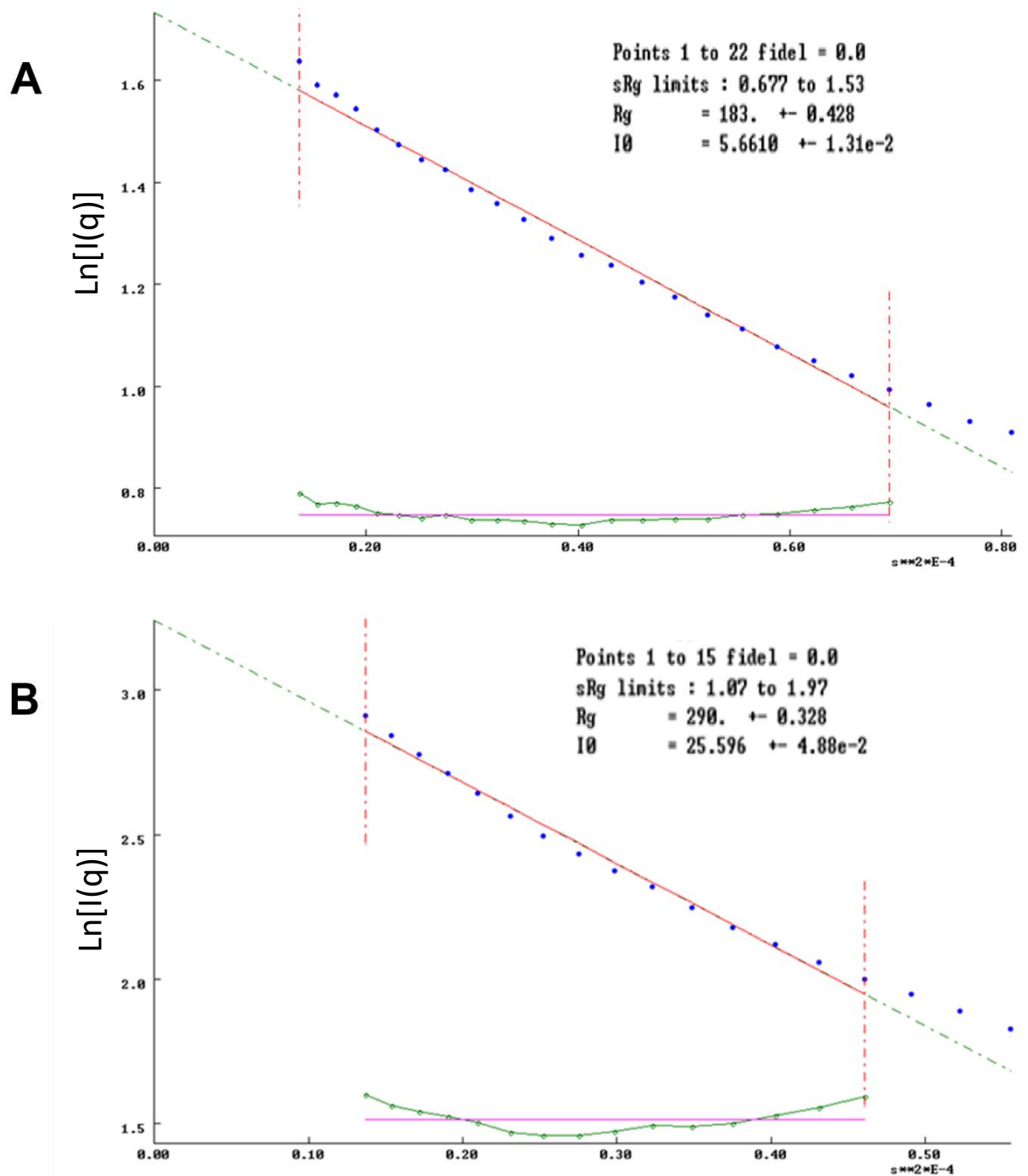


Figure 4.3 Guinier analysis of mouse P-glycoprotein in apo (A) and ADP trapped states (B).

Blue dots represent the experimentally determined scattering intensities, red line is the theoretical fit of scattering intensities. The green line residual noise in comparison to the noise-free data (pink). The curved residual noise pattern suggest aggregation in the sample.

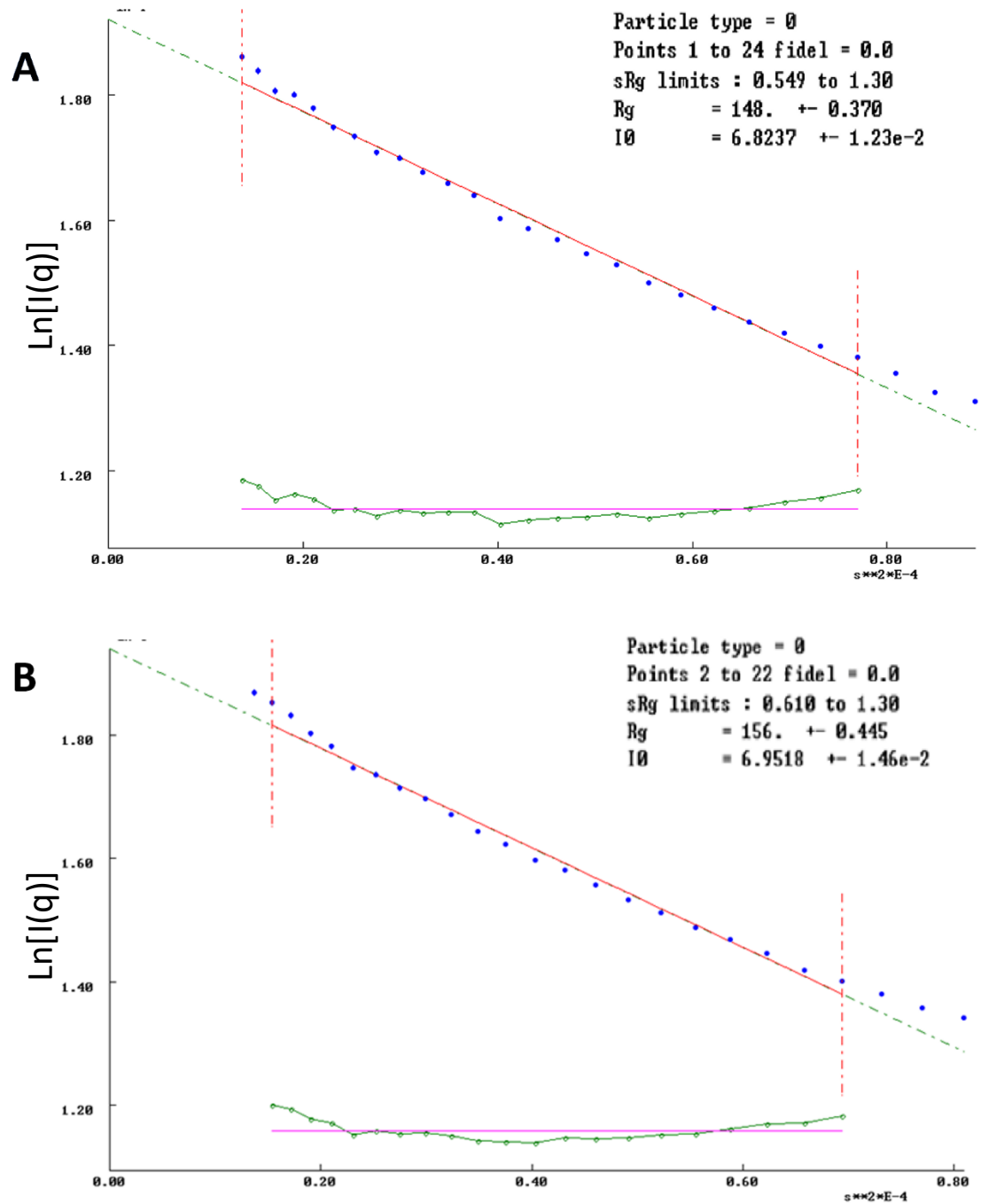
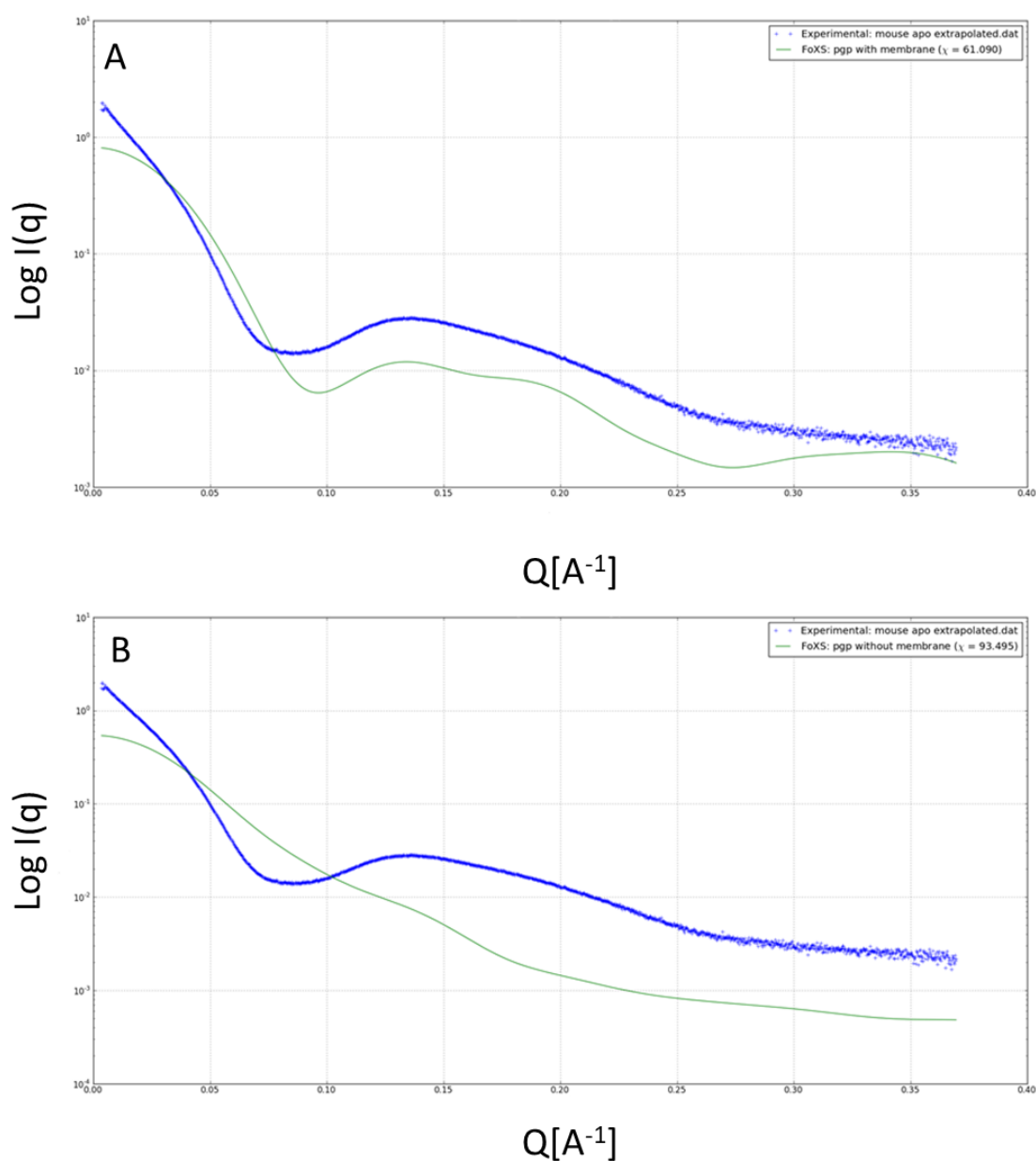


Figure 4.4 Guinier analysis of human *P*-glycoprotein in apo (A) and ADP trapped states (B).

Blue dots represent the experimentally determined scattering intensities, red line is the theoretical fit of scattering intensities. The green line residual noise in comparison to the noise-free data (pink). The curved residual noise pattern suggests aggregation in the sample.



*Figure 4.5 Comparison of experimental scattering profile and theoretically calculated profiles. A) Theoretically calculated profile (green) of P-glycoprotein embedded in membrane Nano disc with the experimental curve of mouse P-glycoprotein (blue). B) Theoretical scattering profile of Mouse P-gp without detergent micelle or Nano disc (green) with the experimental curve (blue)*

Table 4.1. Shows the structural parameters obtained from scattering profiles.

*Table 4.1 Summary of SAXS derived structural parameters*

	<b>R<sub>g</sub> (Guinier) (Å)</b>	<b>AutoR<sub>g</sub> (GNOM) (Å)</b>	<b>Real Space R<sub>g</sub> (Å)</b>	<b>D<sub>max</sub> (Å)</b>
<b>Mouse Apo</b>	183	215.5	216.17	780
<b>Mouse ATP/vana</b>	290	318.91	320.13	990
<b>Human Apo</b>	148	170.52	171.21	675
<b>Human ATP/vana</b>	156	179.06	179.95	700

Structural parameters from the scattering profile suggest large aggregations in mouse and human P-glycoprotein. SAXS is extremely sensitive to interparticle interactions (Zhou and Bilse 2014), at higher concentrations both human and mouse P-glycoprotein show aggregation. This aggregation is visible in human P-glycoprotein purification, even after ultracentrifugation at 100,000 x g or size exclusion chromatography aggregates were visible in the size-exclusion chromatogram (see section 3.2.2). The Guinier analysis of mouse P-glycoprotein showed a radius of gyration of 183Å and 290Å for apo and ADP trapped states respectively. This was also similar to the radius of gyration calculated from pairwise distance calculations. However, with manipulated D<sub>max</sub> of 150 (theoretical D<sub>max</sub> from P-glycoprotein structures) the apo-state showed R<sub>g</sub> of 59Å and ADP trapped state showed 62Å. which is close to the mathematically calculated theoretical R<sub>g</sub> (45.83Å) of mouse P-glycoprotein. Similarly, human P-glycoprotein also showed D<sub>max</sub> four times the theoretical D<sub>max</sub> indicating aggregation.



### 4.3 Experiment 2: Concentration-dependent SAXS profiles

To improve the scattering profiles, further sets of data on different catalytic states of mouse and human P-glycoprotein was collected. Furthermore, different concentrations were used for each catalytic state. The 2D profiles were averaged and subtracted for their respective buffers as explained in section 4.2. The profiles were extrapolated to infinite dilution in PRIMUS software (Konarev, Volkov et al. 2003) in the ATSAS package (Manalastas-Cantos, Konarev et al. 2021). Figure 4.6A shows the representative concentration-dependent profiles of mouse P-glycoprotein in apo-state. P-glycoprotein at 5mg/ml again demonstrated an aggregation-induced scattering profile at the lower 's' region ( $s=0-0.02$ ). This effect was reduced with a decrease in concentrations. However, at lower concentrations, the scattering profiles demonstrated increased deviations in the high 's' region. Figure 4.6B shows the extrapolated profiles (section 2.3.3.2) of mouse P-glycoprotein in different catalytic states. The difference of scale in drug bound states i.e. mP-gp-verapamil and mP-gp-ATP-verapamil and non-drug bound states is due to the fact that the scattering profiles for Apo and ATP-Vanadate, protein concentrations of 0.5, 1, 2, and 5 mg/ml were used. Whereas for scattering profiles of mP-gp-verapamil and mP-gp-ATP-verapamil only 0.5, 1, and 2mg/ml concentrations were used.

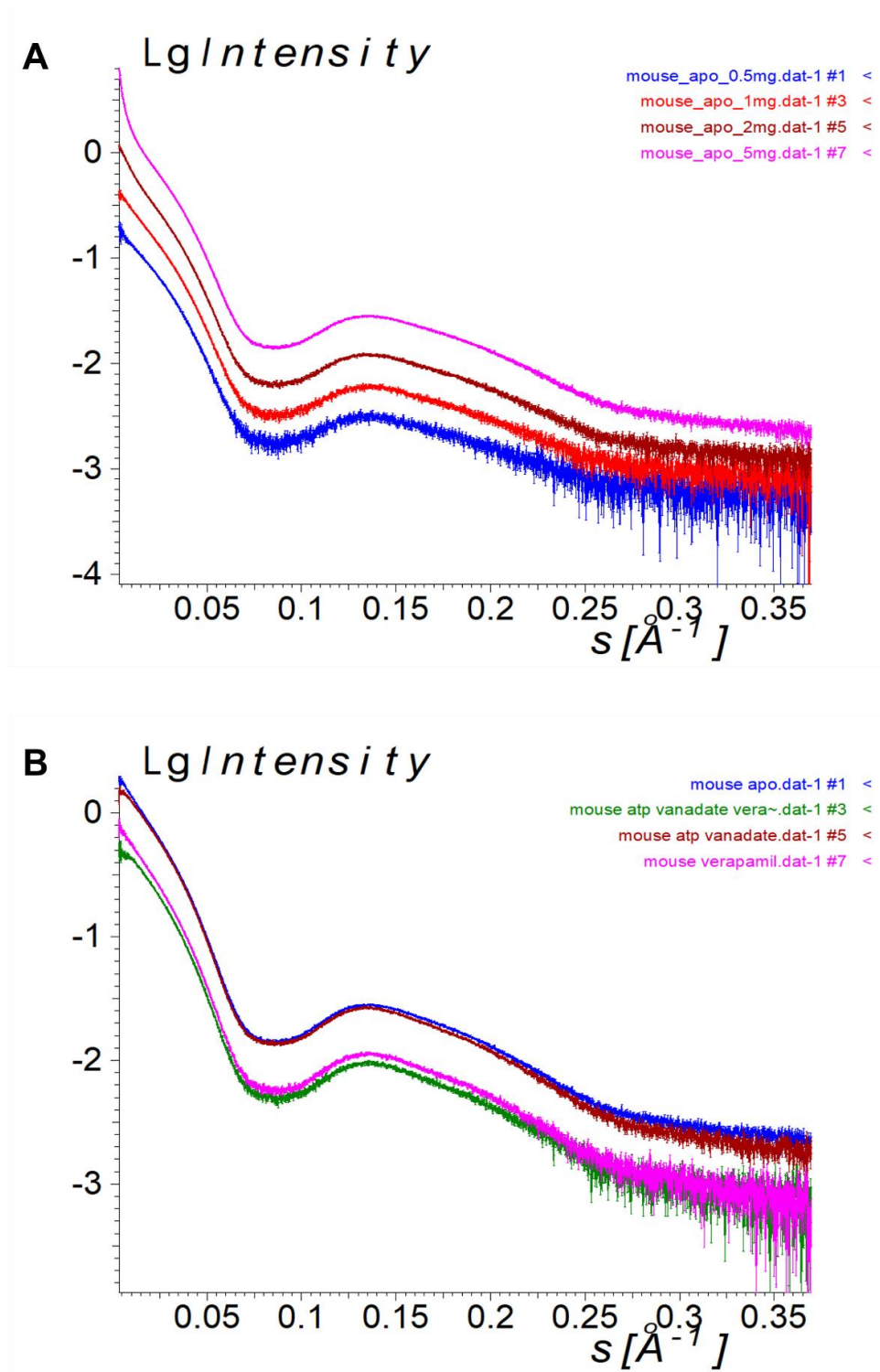


Figure 4.6 Representative scattering profile of mouse P-glycoprotein, plotted as a function of intensities vs momentum transfer ( $4\pi\sin\theta/\lambda$ ). A) Concentration-dependent analysis of mouse P-gp in the apo-state. the offset in intensities recorded is due to the differences in the concentrations used. B) Extrapolated (projected at zero concentrations) scattering profiles of mouse P-glycoprotein in different conditions. The offset of scale was due to the difference in the concentration used for extrapolation in each condition.

Figure 4.7 shows the extrapolated profiles of human P-glycoprotein in apo, and ADP trapped states. The extrapolated profiles also suggested large aggregation ( $d_{\text{max}} \sim 900 \text{ \AA}$ ) in human P-glycoprotein. Even the Guinier analysis of human P-glycoprotein at 0.5mg/ml, showed the presence of large aggregation. This aggregation can be intrinsic to the detergent purified protein. Prima-facie the size-exclusion results also suggested the intrinsic aggregation propensity of human ABCB1 (see section 3.2.1.2). Despite ultra-centrifugation and repeated SEC a small portion of the protein was found to be aggregated in final samples. This implies a intrinsic propensity of detergent purified human P-glycoprotein to aggregate. This might be a reason that most of the successful structural studies reported on human P-glycoprotein have been done with nano-disc reconstituted protein (Verhalen, Dastvan et al. 2017, Alam, Kowal et al. 2019, Nosol, Romane et al. 2020).

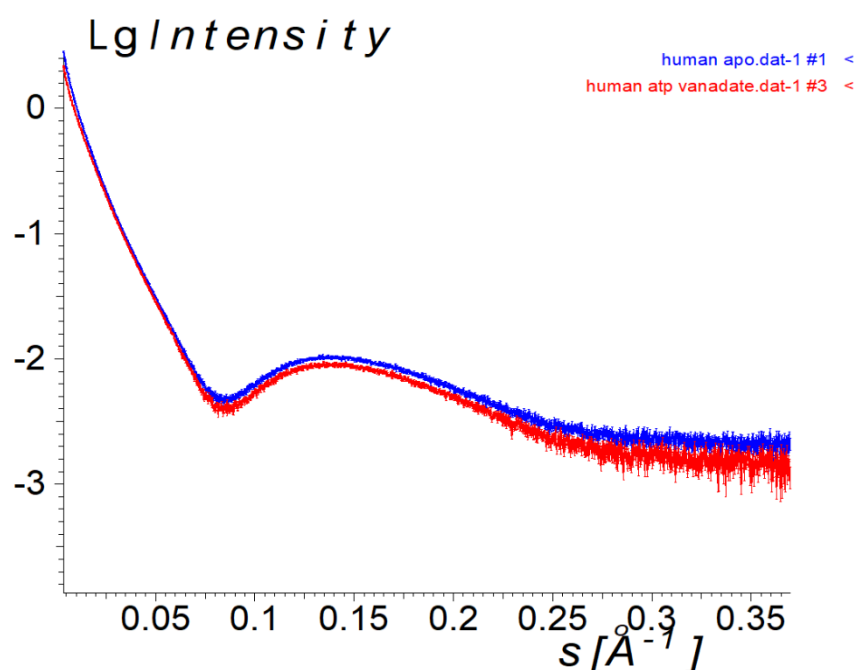


Figure 4.7 Extrapolated scattering profiles of human P-gp in Apo (red) and ADP trapped state (blue).

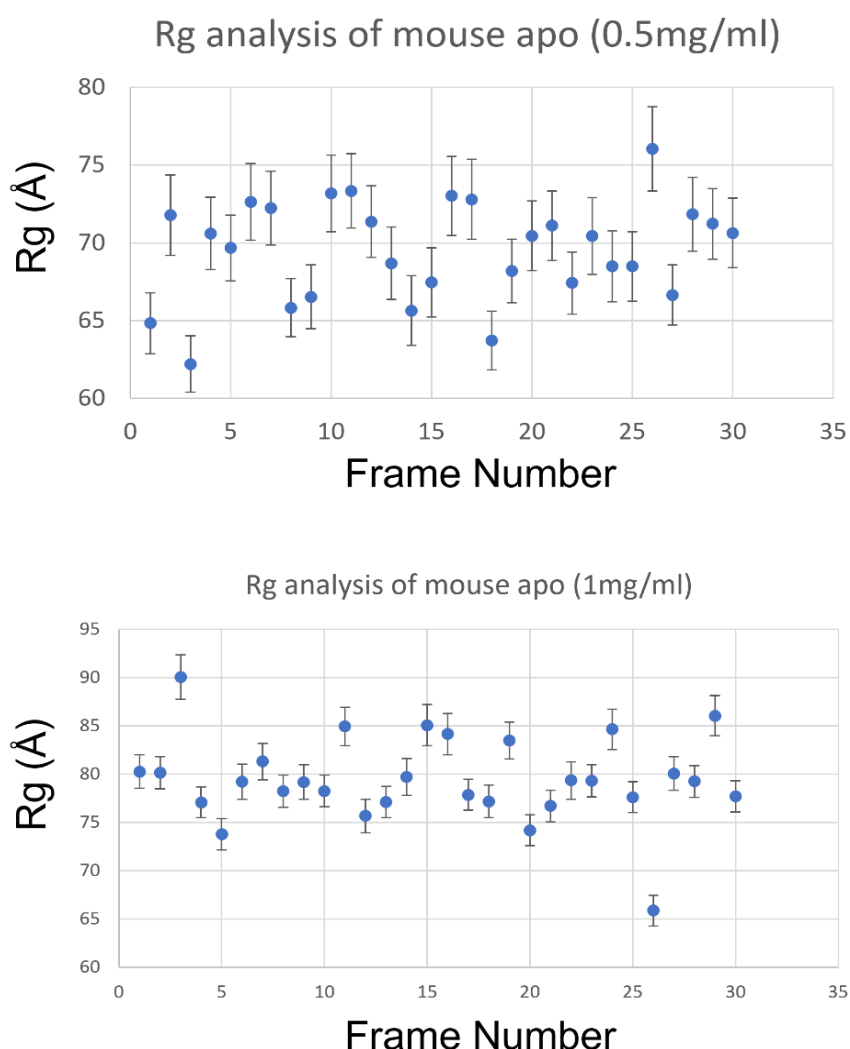
Table 4.2 shows the structural parameters obtained from the extrapolated scattering profiles. Human P-glycoprotein showed  $D_{\max}$  of  $\sim 900\text{\AA}$  in both apo and ADP trapped state, however, mouse P-glycoprotein showed less aggregation as compared to the human analogue. Interestingly, the structural parameters of mouse P-glycoprotein in the presence of ATP showed lower  $R_g$  and  $D_{\max}$  values as compared to the catalytic states where ATP was absent suggesting that ATP can stabilize P-glycoprotein against aggregation. This is consistent with the findings of Lusvarghi et al where P-glycoprotein shows increased thermostability in presence of ATP under non-hydrolytic conditions (Lusvarghi and Ambudkar 2019).

*Table 4.2 Structural parameters obtained from extrapolated scattering profiles.*

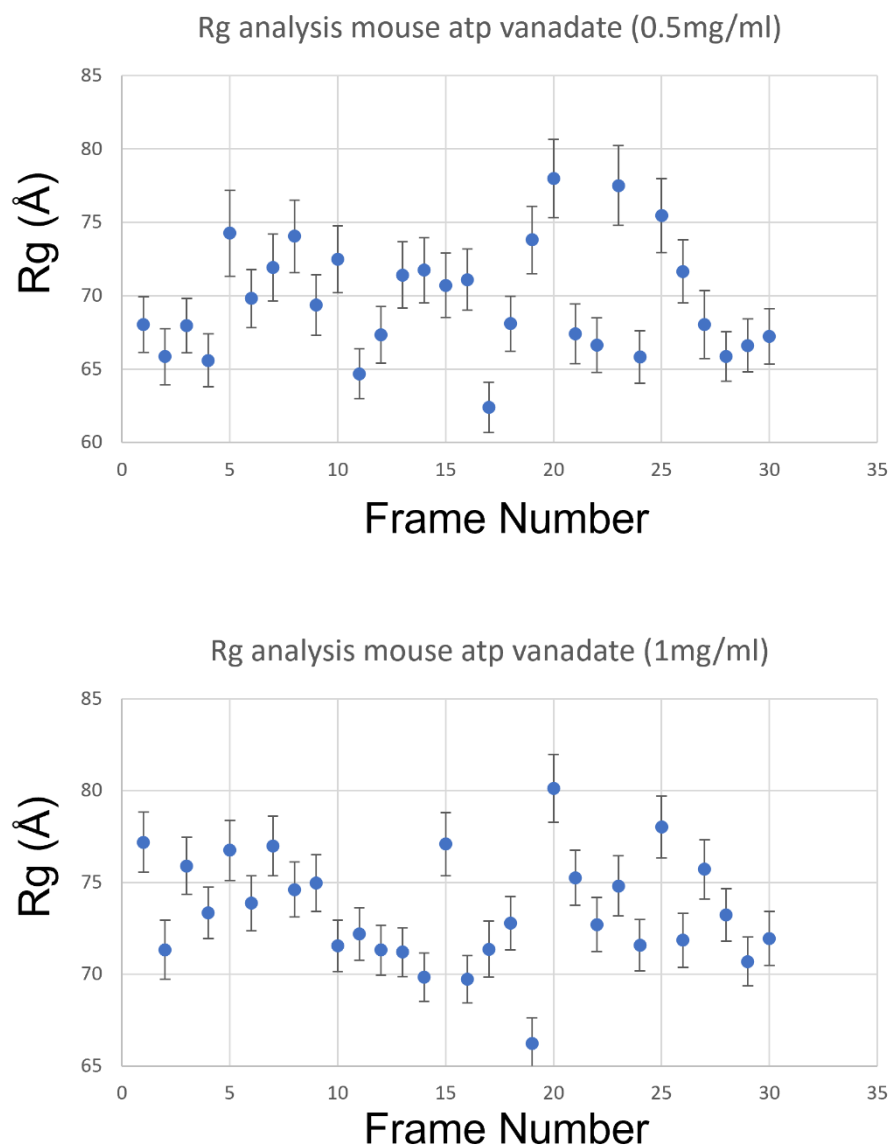
	Apo		ADP trapped		Verapamil		verapamil simulated ADP trapped	
	Human	Mouse	Human	Mouse	Human	Mouse	Human	Mouse
$R_g$ (guinier) $\text{\AA}$	245.81	89.2	201.4	80.1	---	86.8	---	69.4
$R_g$ (Pr) $\text{\AA}$	254.69	92.96	248.15	85.14	---	104.79	---	75.43
$D_{\max}$ $\text{\AA}$	894	462	885	367	---	445	---	321
$Q_{\min}$ (point)	1	1	1	1	---	1	---	1
$Q_{\max}$ (point)	1648	1648	1648	1648	---	1648	---	1648

Concentration-dependent analysis of human P-glycoprotein still showed aggregation as a major confounding factor for the interpretation of the data. Even at concentrations of 0.5 mg/ml, the low momentum transfer values ( $s=0-0.01$ ) do not plateau towards Y-axis suggest the presence of large aggregates. Nevertheless, structural parameters of mouse P-

glycoprotein showed improvements as compared to the concentration-independent dataset (see Table 4.1) however, the data still showed a concentration-dependent effect on  $R_g$ . To further analyse the dataset, the  $R_g$  of individual data frames from mouse data set at 0.5 mg/ml and 1 mg/ml were estimated. Figure 4.8 shows the  $R_g$  of mouse P-glycoprotein in the apo state. The plot suggests a flat trend which suggests no evidence of radiation damages nor aggregation effects during data collection. However,  $R_g$  values showed a concentration-dependent effect with average  $R_g$  at 0.5mg/ml of 69Å in the apo state and 79Å at 1mg/ml.



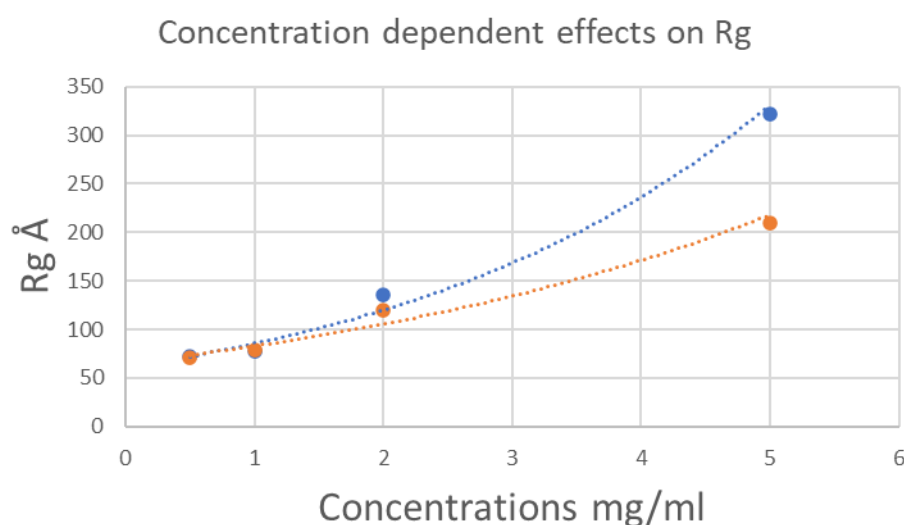
*Figure 4.8 Radius of gyration estimate from individual frames from the mouse apo dataset. (Top panel) Scatter plot of radius of gyration and data frame in mouse P-glycoprotein in apo state at concentration of 0.5mg/ml. (Bottom panel) Scatter plot of radius of gyration vs data frames in 1mg/ml concentration.*



*Figure 4.9 Radius of gyration estimated from individual frames from the mouse ADP trapped dataset. (Top panel): Scatter plot of the radius of gyration and data frame in mouse P-glycoprotein at a concentration of 0.5mg/ml. (Bottom Panel): Scatter plot of the radius of gyration vs data frames in 1mg/ml concentration.*

#### 4.4 Dataset reduction

The radius of gyration analysis of each concentrated showed that at lower concentrations the effects of aggregation was lower as compared the 2mg/ml and 5mg/ml (see Figure 4.10). The data suggested an exponential increase in the  $R_g$  values after 1mg/ml. Indicating that the scattering profiles are dominated by aggregations. Therefore, the data was reduced and further analysis was restricted to 0.5 mg/ml and 1 mg/ml concentrations. The averaged, subtracted scattering profiles of 0.5 mg/ml and 1 mg/ml for mouse P-glycoprotein in different states were imported into the PRIMUS interface and were extrapolated to zero concentration. Table 4.3 shows the structural parameters of extrapolated profiles from the reduced datasets. The new dataset showed  $R_g$ s values closer to the theoretical  $R_g$  (45.5 Å). However, the  $d_{\max}$  again suggested aggregation, with estimates twice the theoretically calculated  $D_{\max}$ .



*Figure 4.10 Representative graphical illustration of concentration-dependent effect on  $R_g$ . Radius of gyration was calculated for each concentration and was plotted as a function of concentration of protein. blue curve represents the  $R_g$  values of mouse P-glycoprotein in apo state and orange curve represent the  $R_g$ s' of ADP trapped state.*

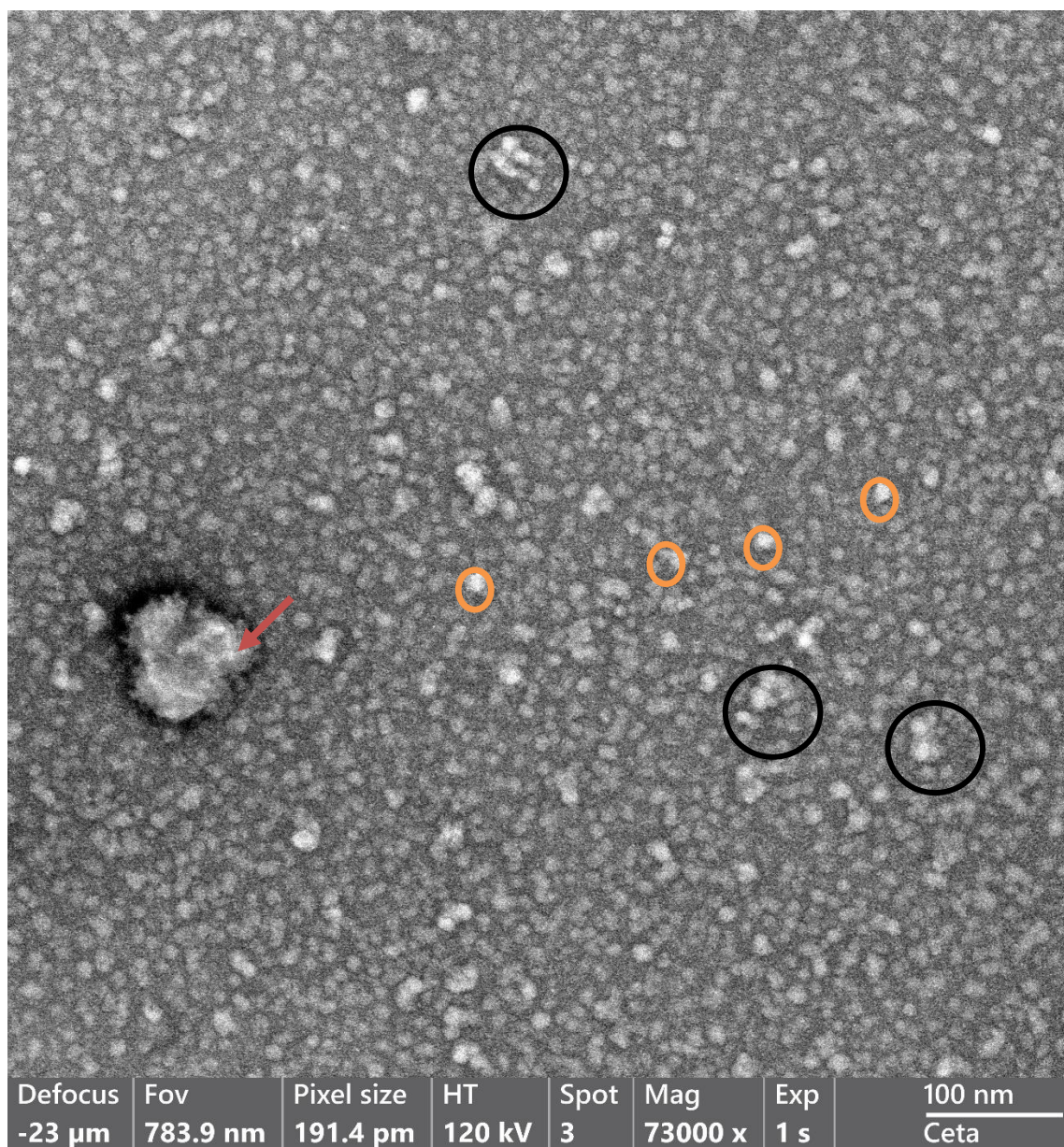
*Table 4.3 Structural parameter of extrapolated profiles from the reduced dataset.*

		<b>Rg (Å)</b>	<b>Dmax (Å)</b>
Mouse	Apo	70.05	421.31
	ATP vanadate	70.63	388.46
	Verapamil	65.79	346.38
	ATP vanadate verapam	65.81	353.74

#### 4.5 Negative stain EM and D<sub>max</sub> analysis:

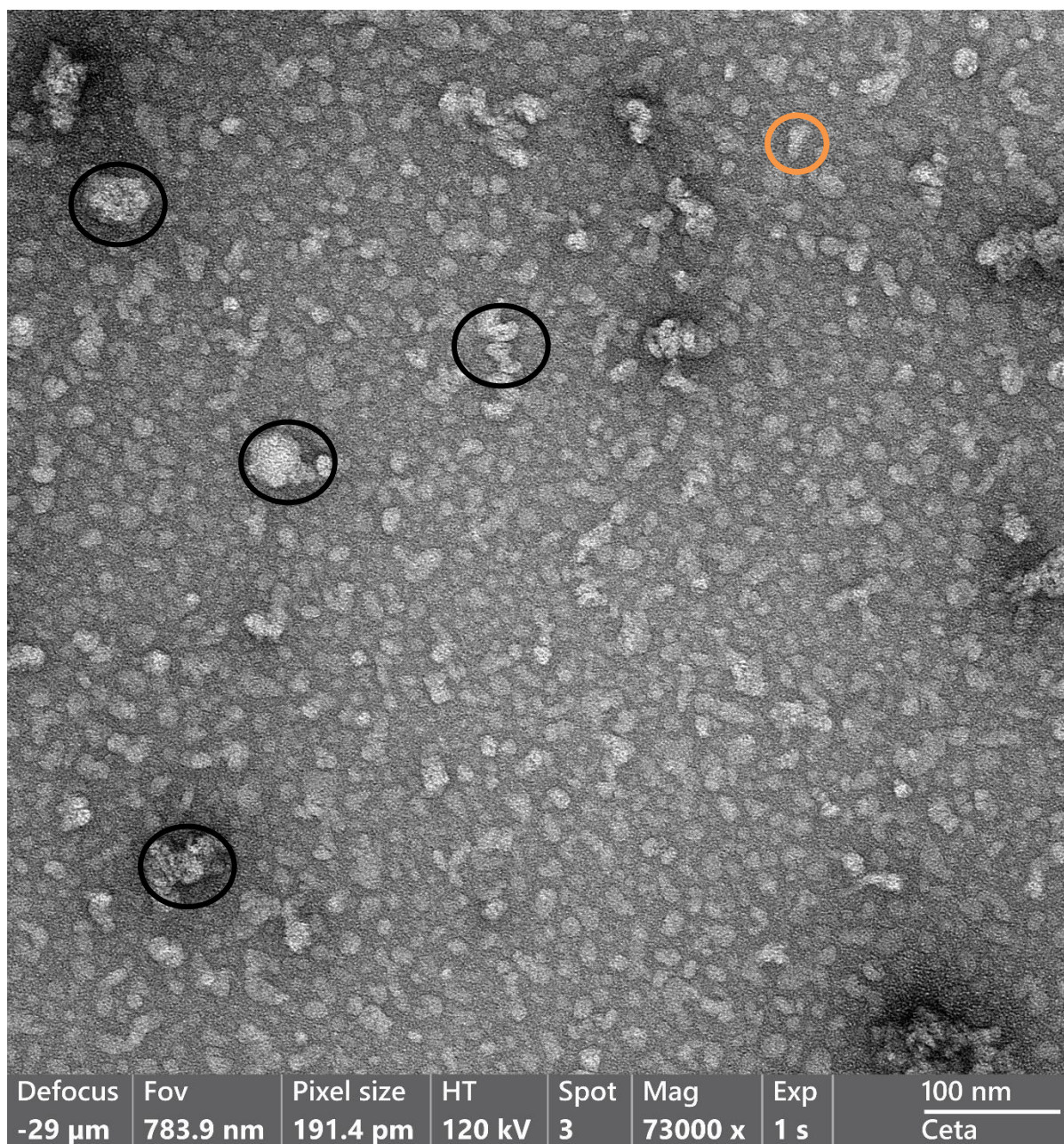
Sample homogeneity of mouse and human P-glycoprotein was further investigated using negative stain microscopy. The protein in the apo state was diluted to a concentration of 10µg/ml and stained using uranyl acetate as explained in the methodology section 2.3.3.1. The images were recorded with a Talos L120C transmission electron microscope with a Ceta CMOS camera. The images were recorded at a magnification of 73,000X with a pixel size of 1.91Å. Figures 4.11 and 4.12 shows images of mouse and human P-glycoprotein in the apo state. Both mouse and human P-glycoprotein showed heterogeneity of the samples. Mouse P-glycoprotein showed more uniformity in terms of sample quality as compared to the human P-glycoprotein. When particle size was calculated using ImageJ mouse P-glycoprotein showed a range from 130Å to 300Å and monomer showed an average particle size of 164Å and aggregates showed average particle size of 256Å which is similar to GNOM calculated D<sub>max</sub> values in the extrapolated dataset. However, human P-glycoprotein showed aggregation values with some particle sizes reaching 897Å (see Figure 4.12). These results from negative stain electron microscopy are consistent with the D<sub>max</sub> obtained from the SAXS profiles.





*Figure 4.11 Representative negative stain images of the mouse P-glycoprotein. The black circle show aggregated proteins whereas orange circles represent probable monomeric proteins. The red arrow represents contamination.*





*Figure 4.12. Representative negative stain images of the human P-glycoprotein. The black circle show aggregated proteins. Orange circles represent probable monomeric proteins.*

## 4.6 Discussion and Conclusions

Small-angle X-ray scattering is a relatively easy and robust method to study the structural properties of protein samples at low resolution (Svergun, Koch et al. 2013). These results do not give atomic resolution information but can estimate the basic structural constraints like shape, size and volume (Svergun, Koch et al. 2013). It can also provide information on the disordered volumes (Bernadó and Svergun 2012). Several studies have successfully used SAXS in determining the low-resolution quaternary structures of multimeric proteins (Korasick and Tanner 2018, Hura, Hodge et al. 2019) as well as to study the enzyme kinetics (Rivas-Pardo, Herrera-Morande et al. 2013). Other than determining the structural parameters, these results can also be used in understanding the sample poly-diversity and for searching for optimal conditions required for protein stability and avoiding interparticle interactions e.g. aggregation and repulsion.

P-glycoprotein is a very flexible protein that demonstrates structurally different states in the transport cycle and all these states show different thermodynamic propensities (Jardetzky 1966, Rosenstein, Sanjurjo et al. 1989, Li, Guttman et al. 2018). Outward-facing conformations are hard to obtain during crystallographic and cryo-EM studies. The recent structure which delineated an outward conformation had to render protein inactive in-order to trap the protein in an outward occluded state (Kim and Chen 2018). Therefore, utilizing a dynamic method like SAXS can be handy in providing structural insights into molecular dynamics. Previously only one study from (McDevitt, Shintre et al. 2008) utilized SAXS to study structural parameters of P-glycoprotein. They tried different ATP analogues to trap P-glycoprotein in outward-facing states. In our study, we tried to trap protein in different catalytic states i.e. apo, drug bound, occluded and outward-facing states. In both studies aggregation of protein

was a major problem. Even in serially diluted protein concentration, the presence of aggregation did not solve the problem. The presence of aggregates even at low concentrations such as 0.5mg/ml suggests that the aggregation is not easily reversible and can not be reversed by dilutions. This aggregation problem is greater in human P-glycoprotein samples this could imply that human P-glycoprotein is more aggregation-prone than mouse P-glycoprotein. Negative stain electron microscopy also shows more aggregation in human P-glycoprotein. This might be the reason why successful structural studies on human P-glycoprotein have tended to be with lipid nanodiscs (Alam, Kowal et al. 2019, Nosol, Romane et al. 2020). Interestingly, our results suggest that the presence of substrates can stabilize the protein thus reducing the aggregation. This is in line with the findings of (Lusvarghi and Ambudkar 2019) which suggest that the presence of ATP analogues leads to NBD dimerization thus stabilizing the P-glycoprotein.

SAXS analysis for structural characterization of membrane protein is standard procedure now. It can provide information about the system which can be later used in high-resolution studies. Purification of membrane proteins requires solubilizing agents to keep their structure intact in the aqueous solutions. Detergent molecules above the CMC are the easiest and widely used method of membrane protein solubilization (Seddon, Curnow et al. 2004). These detergent molecules develop a corona around the lipid embedded region (Pérez and Koutsioubas 2015). These regions can scatter the X-rays strongly influencing the final scattering profiles of integral membrane proteins (Koch, Vachette et al. 2003, Koutsioubas 2017). Moreover, quantification of micelle concentrations in the samples and buffer control is a challenging problem. In ideal conditions, the concentration of buffers and samples are similar. However, in the presence of detergents with concentrations above the CMC, this becomes very complex with the

heterogeneity of micelles sizes, consequently systematic errors in buffer subtractions (Chen and Hub 2015). With the recent developments in SEC-SAXS, the issues of protein aggregation and detergent micelles can be minimized (Brennich, Round et al. 2017). The aggregated protein and detergent micelle have different elution rate as compared to the monomeric protein.

In recent times several atomistic models of P-glycoprotein have been reported. However, these models have provided limited information on the conformational heterogeneity of the protein. In this study, we used an alternative low-resolution small-angle scattering method to study the structural parameters of P-glycoprotein in different catalytic states. Disappointingly our results suggested aggregation as a confounding factor. However, the scattering profiles of P-glycoprotein in presence of substrates showed improved profiles. This information can be useful in further high-resolution studies on P-glycoprotein.

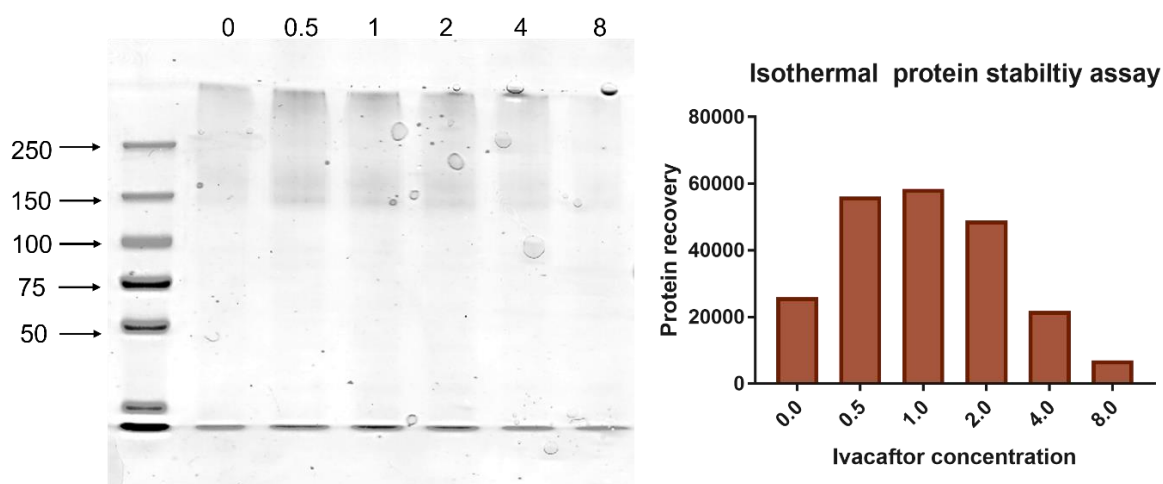
## Chapter 5: Structural studies on ligand-protein interaction of ivacaftor with P-glycoprotein

Ivacaftor is an FDA approved drug for potentiation of Cystic Fibrosis transmembrane conductance regulator (CFTR) to facilitate the increased chloride transport (Csanády and Töröcsik 2019, FDA 2019). In a preliminary study, reported by Dr Swathi Lingham ivacaftor could strongly induce ATPase activity in human P-glycoprotein in the sub-micromolar range (Lingam, Thonghin et al. 2017). Moreover, the drug was able to compete with Hoechst 33342 in an in-vivo drug transport experiment. These results suggested the binding of ivacaftor in the conventional P-glycoprotein binding site. However, The recent structure of CFTR bound with ivacaftor delineates binding site between TM4, 5 and 8 with drug-exposed to the lipid bilayer (Liu, Zhang et al. 2019). This site is contrary to the central, large, water-filled cavity, delineated from structural and biochemical studies on MDR proteins. (Ward, Szewczyk et al. 2013, Li, Jaimes et al. 2014, Nicklisch, Rees et al. 2016, Thonghin, Collins et al. 2018). Interestingly, ivacaftor demonstrate a highly lipophilic property with water solubility <0.05µg/ml and a calculated logP value of 5.76 (Chin, Hung et al. 2018), suggesting partitioning of ivacaftor to a hydrophobic environment. Therefore, it would be interesting to probe the ligand-protein interaction of ivacaftor with P-glycoprotein. In this study, we tried to demonstrate ligand-protein interaction of ivacaftor with mouse and human P-glycoprotein. For this, we employed different methods such as static light scattering, thermal unfolding, and tryptophan fluorescence to understand the binding affinity of ivacaftor with mouse P-glycoprotein.

### 5.1 Preliminary analysis on ivacaftor binding with Mouse P-glycoprotein

Preliminary analysis of ivacaftor binding with the mouse P-glycoprotein was studied using the protein thermal unfolding method i.e. when transmembrane proteins are treated at higher

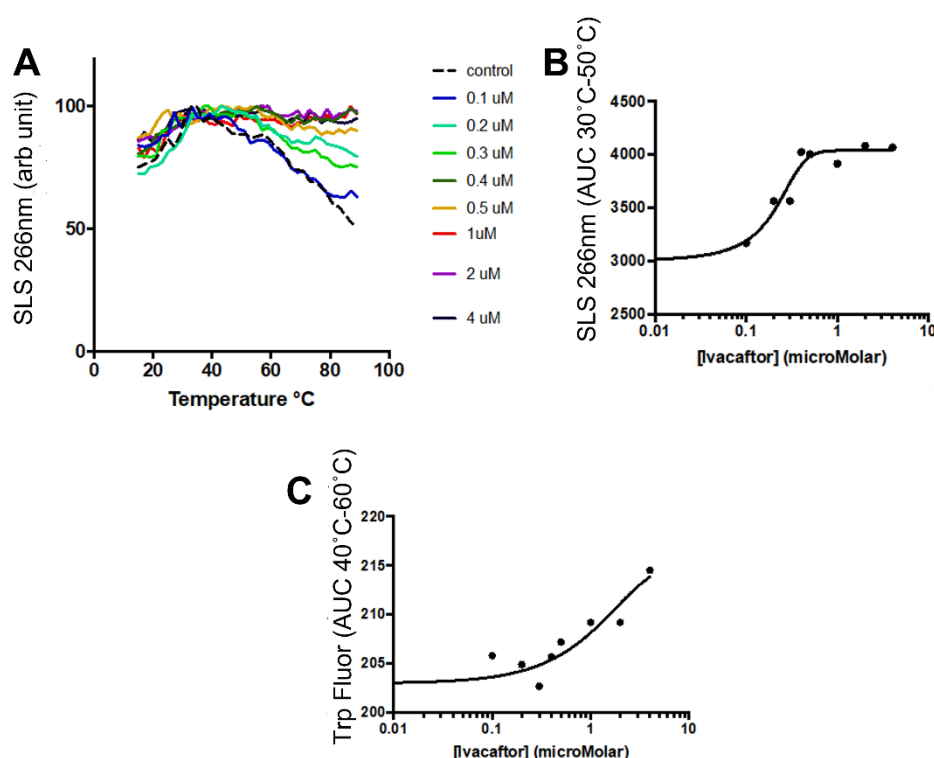
temperatures the protein unfolds and aggregates thus reducing the total recoverable protein on SDS-PAGE (Lysko, Carlson et al. 1981). In the presence of a substrate, the protein may change its thermostability therefore more or less protein is recovered. Figure 5.1 shows protein recovery of P-glycoprotein in presence of different concentrations of ivacaftor. The results suggest stabilization of protein at a sub-micromolar range of drug which reaches a maximum protein recovery at 0.5 $\mu$ M ivacaftor this recovery is not significantly changed at 1 $\mu$ M and 2 $\mu$ M concentrations. However, at higher concentrations protein recovery fall below the levels of control. This can be explained based on DMSO levels at higher concentrations of Ivacaftor. Ivacaftor is practically insoluble in water at room temperature. Therefore, working concentrations were made in DMSO. At 4 and 8  $\mu$ M concentrations, the concentration of DMSO was 4% and 8% in the final volume of reactions. This concentration is well beyond the experimentally recommended concentrations of DMSO on proteins stability (Tjernberg, Markova et al. 2006, Arakawa, Kita et al. 2007).



**Figure 5.1** Thermal unfolding assay for ivacaftor-mP-gp interaction. (Left panel) SDS-PAGE of protein recovery during the thermal unfolding assay. Each lane represents a different concentration of ivacaftor in the  $\mu$ M range. (Right panel) Quantification of protein recovery after thermal unfolding assay by the densitometric method. Y-axes represent protein recovery in arbitrary units.



The binding of ivacaftor with mouse P-glycoprotein was further studied using the static light scattering at 266nm and tryptophan fluorescence. Serial dilutions of ivacaftor dissolved in DDM containing buffer were added to 1 $\mu$ g protein in a capillary cassette of an UNCLE fluorometer (Unchained Lab) (see section 2.3.2.4). Static light scattering (SLS) and tryptophan fluorescence were measured at periodic (2 $^{\circ}$ C intervals) temperature points from 15 $^{\circ}$ C to 90 $^{\circ}$ C. The static light scattering shows the linear effect of ivacaftor binding on mouse P-glycoprotein stability till 0.5 $\mu$ M (see figure 5.2). The ivacaftor induced SLS effect is saturated after 0.5 $\mu$ M. This result is also consistent with the isothermal unfolding assay where maximum recovery concentration was between 0.5 $\mu$ M to 2 $\mu$ M ivacaftor (Figure 5.1). When tryptophan fluorescence was observed between 40 $^{\circ}$ C to 60 $^{\circ}$ C as a function of ivacaftor concentration it showed an ivacaftor induced increase with a  $k_D$  of 1 $\mu$ M.



*Figure 5.2 Mouse P-glycoprotein affinity for ivacaftor measured by thermostability changes: (a) Static light scattering (SLS), as a function of temperature. (B) Static light scattering as a function of Ivacaftor concentrations. (C) Tryptophan fluorescence as a function of ivacaftor concentrations.*



## 5.2 Docking of ivacaftor with mouse P-glycoprotein (4KSB)

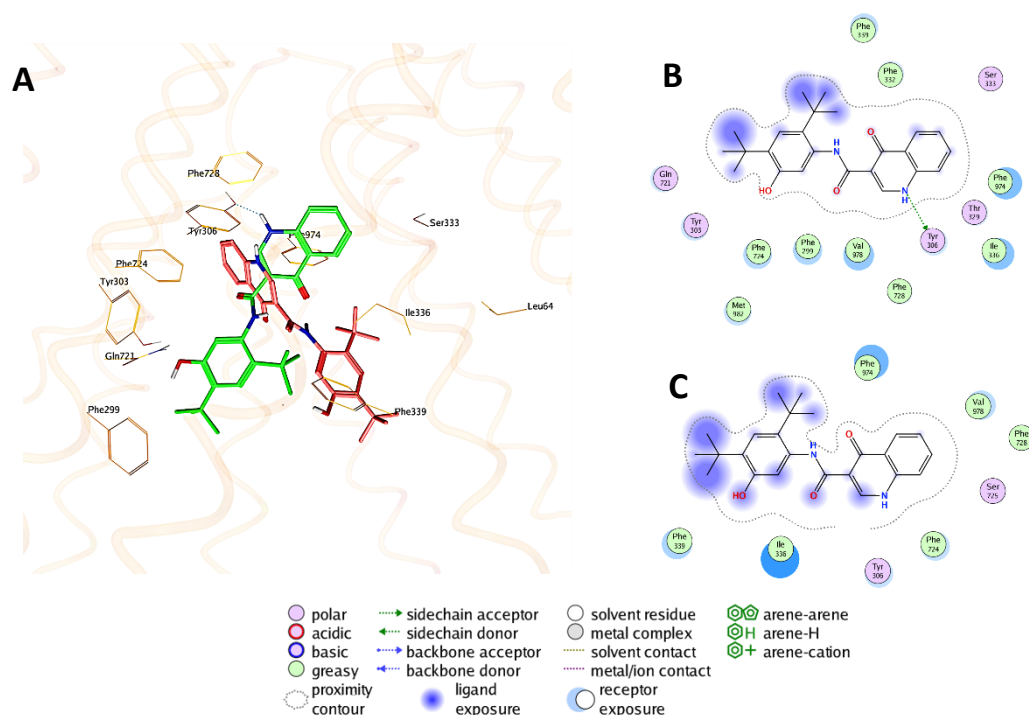
To explore the ligand-protein interaction of ivacaftor, a series of experimental possibilities were explored. For preliminary analysis ivacaftor was docked with the crystal structure of mouse P-glycoprotein in the apo state (4KSB) (Ward, Szewczyk et al. 2013). The two-dimensional structure of ivacaftor and Dodecyl- $\beta$ -D-maltoside (DDM) were obtained from Pubchem. The ligand was calculated for charges using the MMFF94x algorithm provided in MOE (Molecular Operating Environment 2021). The 2D structure was energy minimized to obtain the local minima. 4KSB was imported in the GOLD interface and protonated. Both ligand and protein were considered flexible. A total of 100 conformations were generated and ranked according to the scoring functions. For the control experiment, the already resolved ligand was re-docked in the mouse and human P-glycoprotein structures (see Appendix Figure 5.1). The protocol was able to find the same binding cavity within the top 10 ranked conformations. Furthermore, conformations with RMSD of less than 1Å were also present within the top ten conformations. Therefore, for this exploration, the top ten ranked conformations from both scoring functions were used in the consensus scoring method. For predictive flexibility in conformation exploration, 'rank by number' and 'rank by vote' method of consensus scoring method was used. 'Rank by number' is the simplest method where top 'N' scoring values from the scoring functions are used as they are. Whereas in 'Rank by vote' only those conformations are considered which are the highest vote in top 'N' conformations ranked by the scoring function. Table 5.1 shows the top ten conformations ranked according to the scoring functions i.e. 'rank by number' (see Appendix Fig 5.2 for predicted binding cavity).

Table 5.1 Shows the solution number and their corresponding scoring values of the top ten ivacaftor conformers with mouse P-glycoprotein (PDBID: 4KSB).

docking conformation	ChemPLP	docking conformations	GOLDscore
dock3	52.6138	dock92	16.9808
dock11	52.139	dock79	14.6454
dock28	52.138	dock84	8.4602
dock34	51.4108	dock69	7.8295
dock65	51.2955	dock85	7.3898
dock1	50.6304	dock44	5.4518
dock43	49.9635	dock52	3.5953
dock46	49.9178	dock17	1.0607
dock80	49.8643	dock34	0.6625
dock25	49.3007	dock72	-0.7754

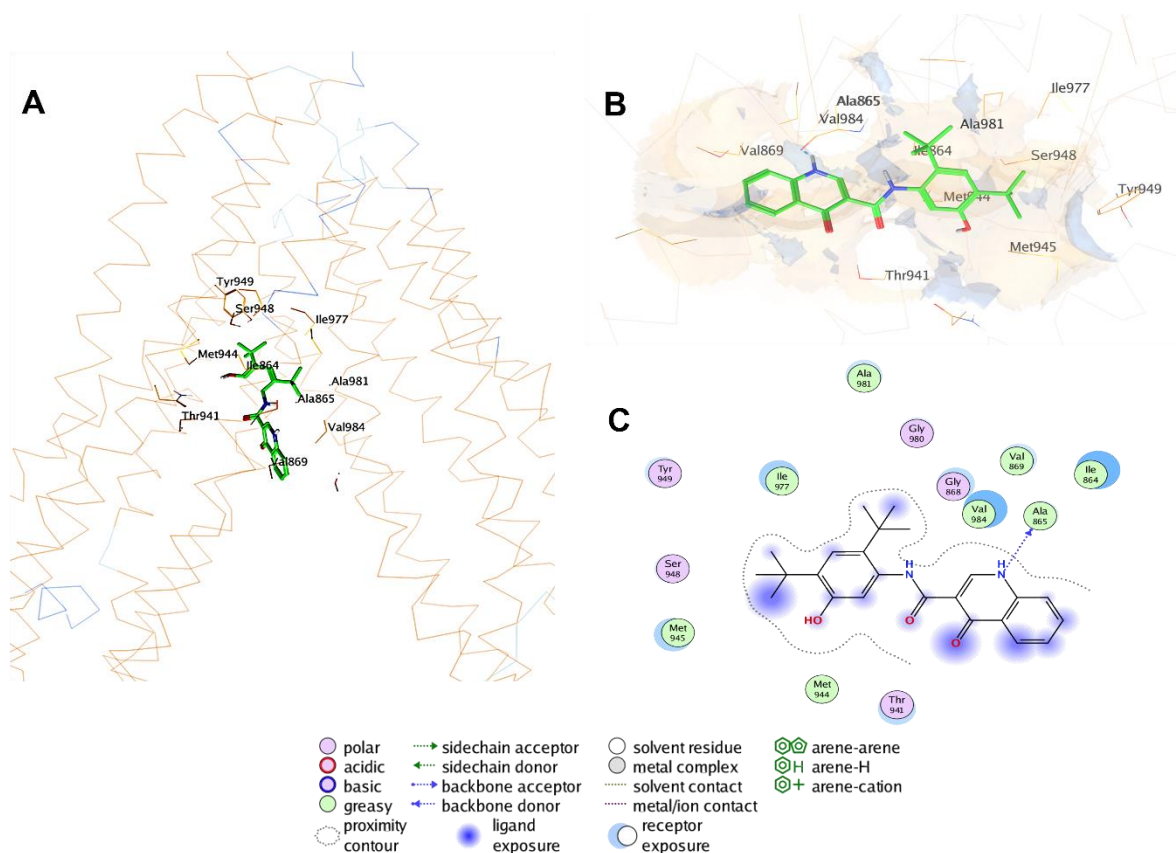
All the top 10 ranked conformations from ChemPLP showed binding of ivacaftor similar to the binding site identified by crystallographic structures of P-glycoprotein with cyclic peptides (Li, Jaimes et al. 2014, Szewczyk, Tao et al. 2015). The docked conformation is extended along the TM6, and TM7 (see Figure 5.3A). The two tertiary-butyl moieties are surrounded by hydrophobic residues like Phe299, and Phe339 whereas the hydroxyl group on hydroxyphenyl moiety is surrounded by polar residues Tyr303 and Gln721 within 3Å. The amino group on quinoline moiety forms a hydrogen bond with Tyr306 (see Figure 5.3B). In comparison, the top-ranked conformation of GOLDscore also shows a similar binding site as of ChemPLP conformation. However contrary to the conformation from ChemPLP, it is almost a mirror orientation with other half of the pseudo-symmetry, with ligand extending between TM7 and TM6 between the two transmembrane domains (Figure 5.3C). The quinoline substructure is embedded in

aromatic amino acids such as Phe728, Tyr306 and Phe974 at TM7, TM6 and TM12, whereas the tertiary-butyl-phenol moiety is surrounded with hydrophobic residues Ile336 and Phe339.



**Figure 5.3** Ligand-protein interaction of ivacaftor with mP-gp (4KSB). (A) 3D representation of ivacaftor interaction with mP-gp. green conformation represents the top-ranked conformation from ChemPLP. Pink conformation represents the top-ranked conformation from GOLDScore. (B) 2D representative of ChemPLP conformation with mP-gp. (C) 2D ligand-protein interaction map of GOLDScore conformation with mP-gp

Interestingly, docking conformation 34 is present in the top ten of both the scoring functions. This conformation shows a novel binding site for ivacaftor close to the entry portal between TM12 and TM10 (see figure 5.4). The conformation is extended along with the Transmembrane Domain (TMD) 2 with an amine group on quinoline forming hydrogen bonding with the backbone of Ala865. The tertiary butyl moiety is surrounded by lipophilic residues like Ile977, Tyr949, Ala981 and Met945 (Figure 5.4B). Figure 5.4C shows that Overall the drug is surrounded with hydrophobic residue with the inferior part exposed to the large solvent-exposed cavity. This binding site is lower towards the inner membrane than the central binding cavity identified from the crystallography and biochemical data (Li, Jaimes et al. 2014, Szewczyk, Tao et al. 2015). This site is close to the R site (Rhodamine 123 Binding site) identified by cross-linking data (Loo and Clarke 2002).



*Figure 5.4 Ligand-protein interaction analysis of consensus conformation 34. A) 3D representations of the location of ivacaftor in the mP-gp binding cavity. B) 3D representation of ligand-protein interaction of ivacaftor. The hydrophobic cavity is represented with a molecular surface (brown). The hydrophilic regions are represented with blue contours. C) 2D representation of ivacaftor interaction in the binding cavity. The dotted line represents the cavity boundary of the protein. Blue spheres represent the solvent exposed part of the ligand.*

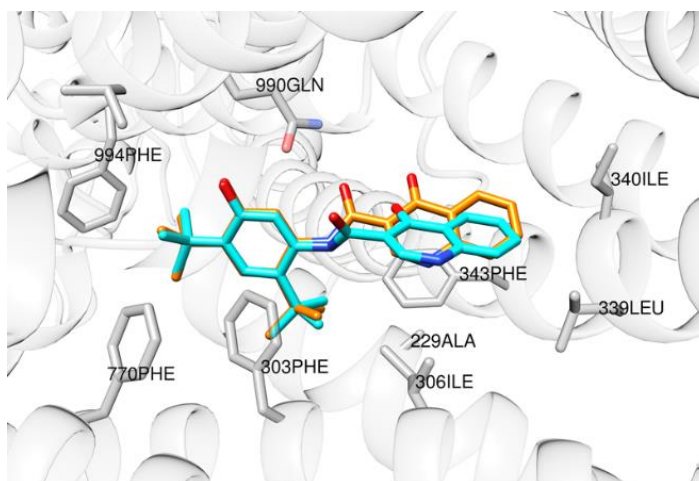
### 5.3 Docking of ivacaftor with human P-glycoprotein

The same protocol used for docking of ivacaftor with mouse P-glycoprotein (PDBID: 4KSB) was used to study the interaction of ivacaftor with human P-glycoprotein (see section 2.3.4.1.1). The structure of human P-glycoprotein bound with vincristine (PDBID: 7A69) was used for the ivacaftor-human P-glycoprotein interactions (Nosol, Romane et al. 2020). Table 5.2 shows the top 10 conformations from both scoring functions along with their respective ranking scores (see Appendix Figure 5.3 for predicted binding cavity).

*Table 5.2 Top ten solutions of ivacaftor with human P-glycoprotein (7A69)*

solutions	ChemPLP	solutions	GOLDscore
dock98	78.6147	dock41	23.9027
dock34	76.1383	dock48	23.1528
dock11	72.3228	dock36	20.3058
dock37	67.0149	dock51	17.1422
dock86	64.8806	dock17	14.8963
dock2	64.2575	dock14	14.8152
dock8	61.3114	dock21	13.872
dock45	60.0321	dock37	13.6726
dock6	59.9571	dock81	13.4725
dock25	59.1194	dock84	13.2079

Docking solution 37 is present in the top ten solutions of both the scoring functions. Figure 5.5 shows the interaction of solution 37 in the binding cavity. Overall the ligand shows a good fit with hydrophobic parts surrounded by aromatic and hydrophobic residues and the central amine group in the vicinity of residues Gln990, Gln347 and Met876. Interestingly, this conformation is similar to the conformation of the top-ranked pose from ChemPLP (solution 98) (see figure 5.5). These conformations show an RMSD of 4.3Å in the binding cavity. This site is similar to the drug-binding site identified in the recently published cryo-EM structure of human P-glycoprotein with inhibitor tariquidar (PDBID: 7A6E) (see Appendix Figure 5.3C) (Nosol, Romane et al. 2020).



*Figure 5.5 ligand-protein interaction of ivacaftor with human P-glycoprotein. orange top-ranked pose from ChemPLP. Cyan pose from consensus analysis*

#### 5.4 Structural studies on ivacaftor binding

Mouse P-glycoprotein incubated with 2 $\mu$ M ivacaftor was vitrified on 1.2/1.3 Quantifoil grids (as described in section 2.3.3.3). The data was collected using a Volta phase plate method on Titan Krios microscope at Diamond Light Source (Harwell, UK). The volta phase plate method helps to increase the image contrast, therefore, increase the signal to noise ratio. The image processing and single-particle analysis were performed by Alexasadio Barbieri (a PhD student in the Ford Lab). For summary, 2400 images were collected and corrected for CTF in cisTEM. The dataset was manually filtered for bad CTF fitting and defocus. The filtered dataset was used for further processing. Particles were picked using the ab-initio method with a maximum particle radius of 72Å. Initially, 0.4 million particles were picked, which were filtered for bad particles in successive 2D classification (particles on carbon supports, micelles and contamination). The final classification consisted of approximately 105,000 particles classified into 50 classes. These classes were further used to generate five ab-initio 3D classes. These 3D maps were further manually refined in the cisTEM package. The final 3D maps were able to show differences in distances between the two NBDs suggesting different catalytic states in the inward-facing conformation (see figure 5.6).

(for more information on the methodology see our paper, 'Structure of ABCB1/P-glycoprotein bound to the CFTR potentiator ivacaftor' in appendix).



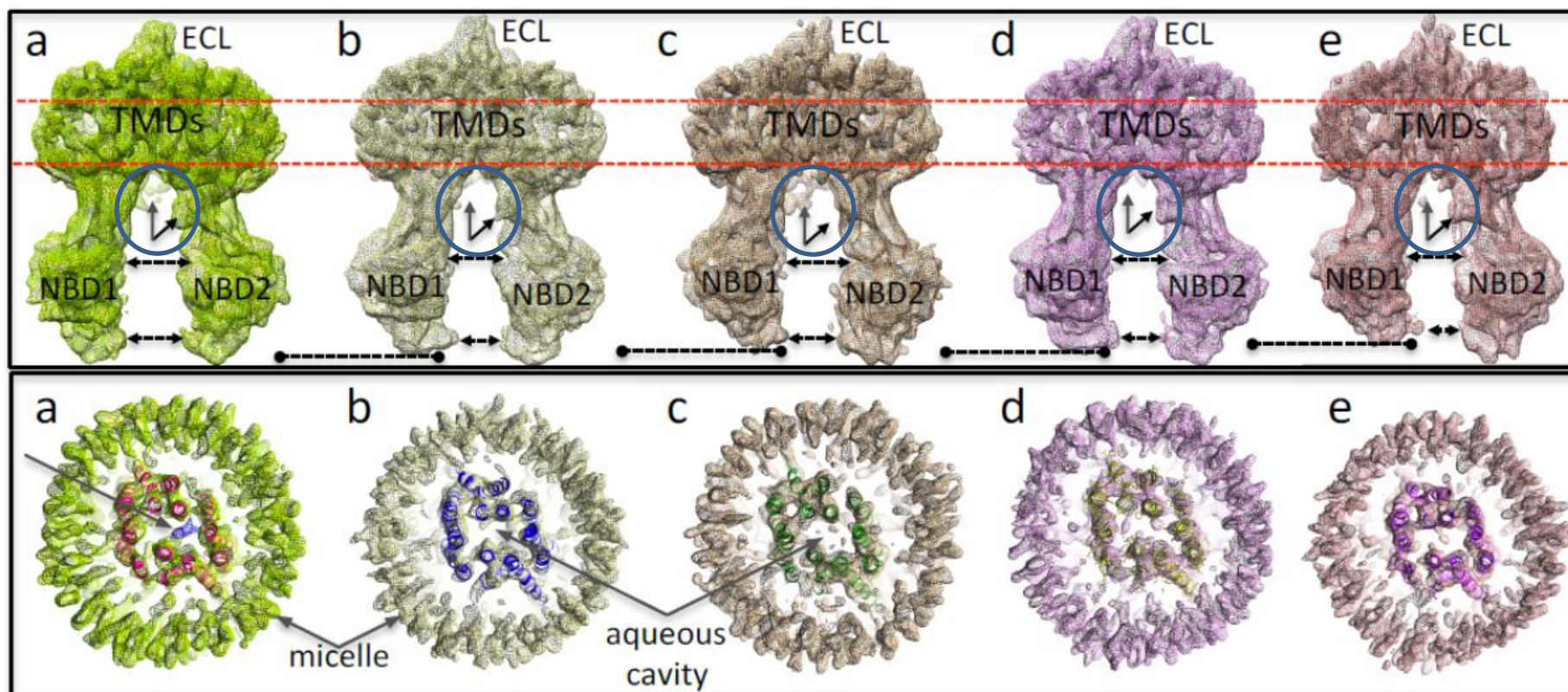


Figure 5.6 Differentiation of five 3D classes. The maps showed an inverted V shaped characteristic ABC transporters inward facing conformation with a large central cavity and a doughnut-like micelle around. The first large density at the bottom represents the Nucleotide Binding Domains (NBDs). Each class showed the difference in distances and arrangement of the NBDs. The pointy edge on the top is the extracellular loop. The blue circle represents the additional densities for N terminus and linker region. The vertical arrow represents the additional density of N-terminus region towards the back, whereas the tilted arrow represents the density for C-terminal end of the linker region between the NBD1 and TMD2.

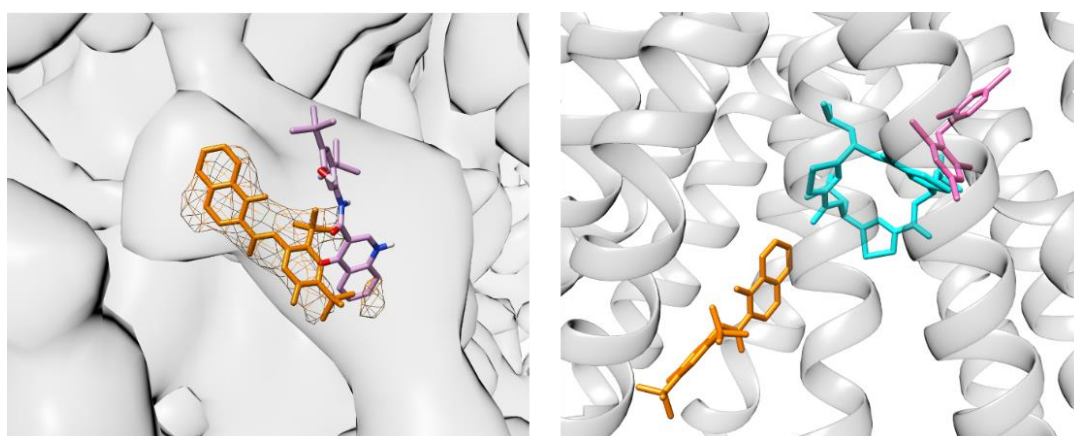
Each state showed different NBD distances demonstrating structural flexibility (see table 5.3). class a, c, and d show larger distances between the NBD1 and NBD2. Class b shows an intermediate distance between the two domains. Whereas, this distance is the smallest in class e. Moreover, the NBD1 is rotated upwards along the hinge-like intracellular loop. This hinge rotation increases from class a to class e. When molecular dynamics force field MDFF fitted real-space structures were calculated for NBD-NBD distance (C $\alpha$  Thr625 - C $\alpha$  Met1266) (see Table 5.3), class a shows the largest distance of 35.65Å whereas, class e showed an NBD distance of 24.4Å. This conformational flexibility is intrinsic to the inverted 'V' like structure with large central solvent-accessible cavities. This flexibility of mouse P-glycoprotein is also evident from previously reported structural and biochemical data (Li, Jaimes et al. 2014, Szewczyk, Tao et al. 2015, Verhalen, Dastvan et al. 2017, Kim and Chen 2018). The 3D classes were also able to resolve additional densities for N terminal loops in class a and class c however, class e shows weak densities for the N terminal region and this region is absent in class b and d. Additional densities of the linker region at TMD2 junctions were also visible in class a, b, d and e, but absent in class c (see figure 5.6). These densities were also recorded in the previous in house dataset of post-hydrolytic ADP-vanadate trapped P-glycoprotein maps (Thonghin, Collins et al. 2018). These regions were previously not resolved in P-glycoprotein structures generated using crystallography (Ward, Szewczyk et al. 2013, Li, Jaimes et al. 2014, Szewczyk, Tao et al. 2015).

*Table 5.3 Distances between two NBD domains measure between C $\alpha$  of Thr625 and C $\alpha$  of Met1266*

<b>Class</b>	<b>Distance in Å</b>
Class1	35.65
Class2	26.33
Class3	29.01
Class4	29.52
Class5	24.40



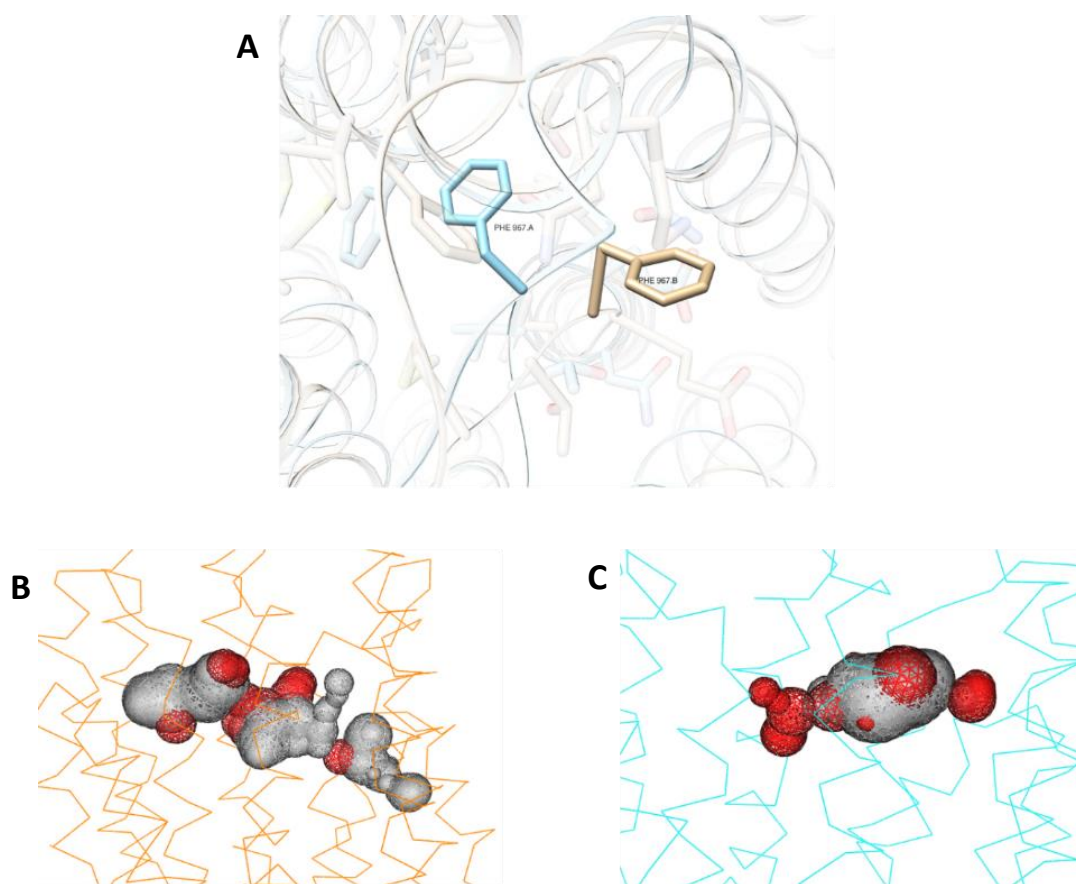
Interestingly, class1 shows additional rocket-shaped density near the entry portal between the TM6 and TM12 (see figure 5.7). The density shows a similar shape like ivacaftor with the two fin like structures similar to the tertiary butyl moieties on the phenyl sub-structure. The remaining rigid structure of the drug shows an elongated density. When compared with docking results of ivacaftor with 4KSB (see the previous section) this density showed a similar location with the consensus ranked conformation (see figure 5.7a). However, the orientation of the ivacaftor molecules is in the opposite direction with the tertiary butyl end towards the inner cavity. This binding site is close to the inner interface of the lipid bilayer, lower than the ligand-binding site identified in crystallographic experiments (Figure 5.7). This site is close to the Rhodamine 123 binding site (R-site) identified by cross-linking assay (Loo and Clarke 2002). However, this site is not well characterized for P-gp-substrate interaction in structural experiments. Therefore, it was interesting to elucidate this density.



*Figure 5.7 3D representation of Ivacaftor binding site. (Left panel) Fitting of ivacaftor in the density (contour levels: 0.67) using Phenix (orange). Ivacaftor in Pink is the predicted conformation from docking studies. (Right panel) Location of the ivacaftor binding site (brown) as compared to the binding of cyclic compounds (cyan) and BDE-100 (pink)*

Furthermore, the real space refined model of mouse P-glycoprotein shows the rearrangement of residues near the extracellular domain. Bulky residues like Phe967 and Phe78 shows rotamers conformation different from the conformations observed in the previous structures. Figure 5.8a shows the rearrangement of Phe976 at TM12 with respect to the same residue in crystal structure 4XWK. This rearrangement extends the binding cavity towards the

extracellular loops. When this was further investigated using the LIGSITE algorithm (Hendlich, Rippmann et al. 1997) implemented in MOE, the new structure showed an increased solvent-accessible cavity near the extracellular loops (Figure 5.8B and 5.8C).



*Figure 5.8 Re-arrangement of side chains in mP-gp. (A) Shows the rearrangement of Phe976 in in-house structure (brown) and 4XWK (cyan). (B and C) Solvent accessible area identified using LIGSITE implemented in MOE. The new structure (B) shows increase in the volume of binding cavity as compared solvent accessible area identified from the crystal structure of mP-gp (C).*

## 5.5 Docking of Ivacaftor with in-house mouse P-glycoprotein

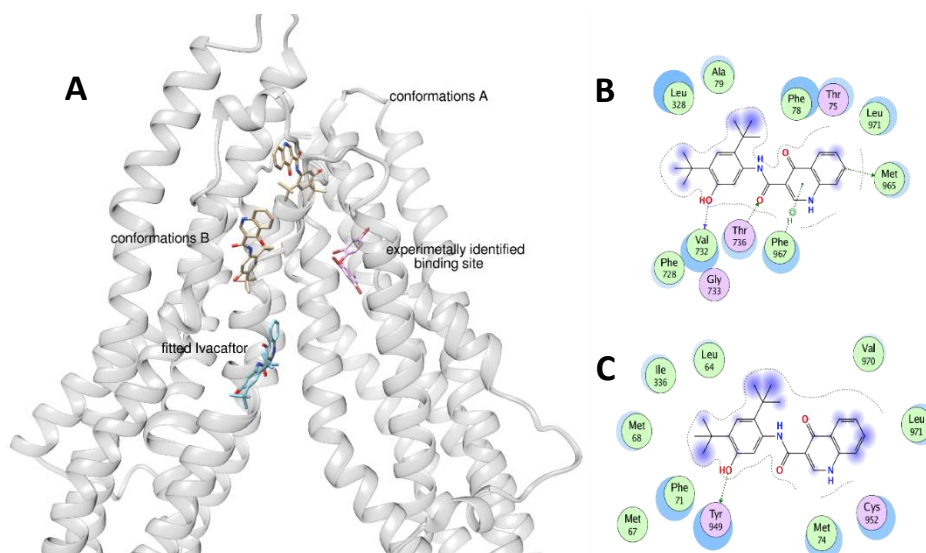
The fitted model of mouse P-glycoprotein with ivacaftor was loaded into GOLD. A total of 100 conformations were generated and ranked according to the scoring function. The conformations were additionally scored according to the GOLD scoring functions. Table 5.4 shows the scores of the top ten conformations from both scoring functions.

*Table 5.4. Shows the solution number and their corresponding scoring values of the top ten ivacaftor conformers with the in-house mouse P-glycoprotein model.*

Solution no	ChemPLP score	Solution no	GOLD score
dock20	72.9709	dock56	6.7068
dock59	63.7511	dock47	0.1461
dock22	58.5087	dock17	-5.1317
dock67	57.7778	dock30	-6.6722
dock7	57.6987	dock90	-7.2753
dock3	56.8013	dock61	-8.5881
dock25	56.6496	dock60	-9.8146
dock47	56.5193	dock10	-10.5955
dock72	56.367	dock20	-10.7902
dock28	56.176	dock95	-11.0999

Docking solution 20 and docking solution 47 were present in the top 10 ranked solutions in both the docking scores (hereafter represented as conformation A and conformation B respectively). The docking conformation A (dock20: table 5.2) showed the highest fitness scores with a fitness value of 72.97 according to the ChemPLP scoring function. The predicted conformation is buried deep in the binding cavity with 1,2 tertiary butyl benzene moiety surrounded by hydrophobic residues Ala79, Val317, Leu328 and Phe728. The central amide group is hydrogen bonding with Thr736 (see Figure 5.9). Interestingly, this site is similar to the expanded cavity identified in the structure (see Figure 5.8). In comparison, docking conformation B is predicted to be in the lower cavity as compared to docking conformation A. This site is identical to the binding cavity reported in crystallographic and cryo-EM structures with other ligands. These results suggest that ivacaftor could show poly-specificity in interaction with mouse P-glycoprotein. Furthermore, rearrangement of rotamers in the upper region

of TMDs also suggests multiple possibilities *i.e.* (i) induced fit mechanism of ivacaftor binding (II) Limitation of MDFF fit due to low resolution or poly-specificity of ivacaftor which is not resolved due to sub-stoichiometric concentration (4:1 protein: drug). The concept of poly specificity is not new for P-glycoprotein. Previously, crystal structures with macrocyclic compounds QZ-RRR and QZ-SSS showed poly-specificity of mP-gp in the binding cavity (Li, Jaimes et al. 2014). Even the new structures of human P-glycoprotein with third-generation inhibitors also showed poly-specificity of substrates with the binding cavity (Nosol, Romane et al. 2020). Therefore, it was interesting to further explore the binding of ivacaftor in excess of the drug.



**Figure 5.9.** Ligand protein interaction analysis of ivacaftor with in-house P-glycoprotein. (A) 3D representation of conformation A, conformation B along with the Phenix fitted ivacaftor in the upper binding cavity. The binding site of BDE-100 (pink) in mouse P-glycoprotein. (B) 2D interaction map of conformation A in the central binding cavity. (C) 2D interaction map of conformation B in the central binding cavity.

## 5.6 Structural studies with a molar excess of ivacaftor

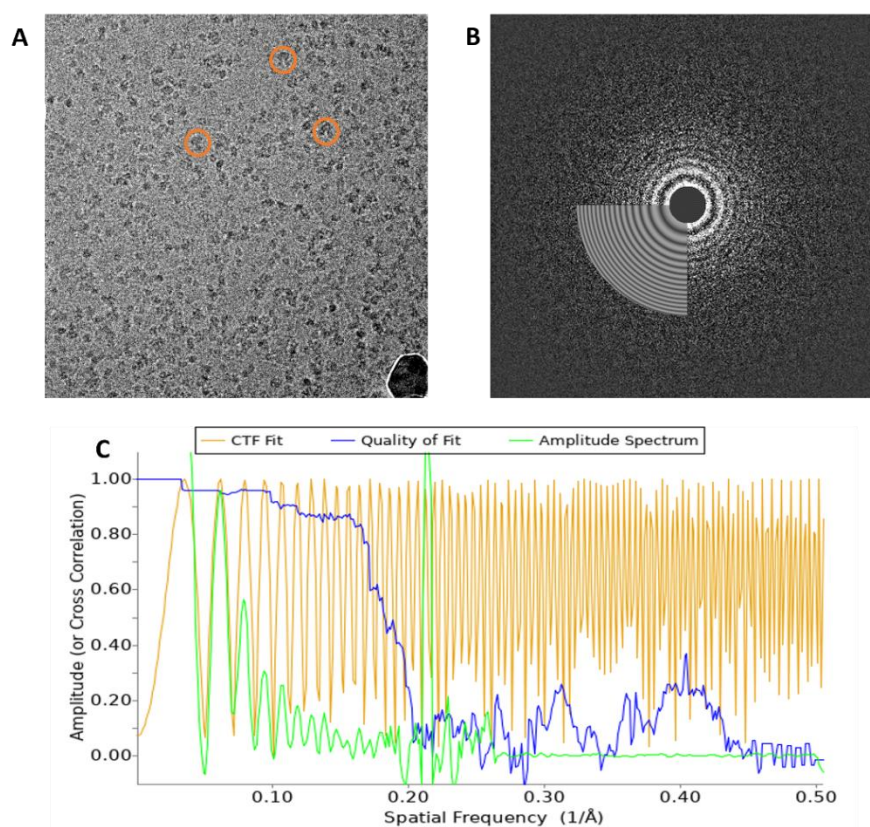
The docking of ivacaftor at two different places predicted the poly-specific binding of ivacaftor with the mouse P-glycoprotein. Therefore, it was interesting to see if ivacaftor interact with mouse P-glycoprotein in a poly-specific manner. The protein at a concentration of 1.5mg/ml was buffer exchanged in glycerol and detergent-free buffer. The buffer exchanged sample was incubated with

20 $\mu$ M ivacaftor (approximately 1:2 ratio of protein to the drug) for 30 minutes on ice before applying on to the grid. The sample was vitrified on Quntifoil 1.2/1.3 copper grids as explained previously (see section 2.3.3.3.1).

Cryo-EM data were collected on a FEI Glacios Cryo-EM with a 200kV electron gun and with a Gatan 3 direct electron detector using an automated EPU software (ThermoFisher). For this dataset 3205 images were recorded and motion-corrected on the fly with motioncorr2 (Zheng, Palovcak et al. 2017). The motion-corrected data was used for single particle analysis (Grant, Rohou et al. 2018). The corrected images were imported in CistEM (Grant, Rohou et al. 2018) and calculated for CTF (Contrast Transfer Functions) corrections using CTFFIND4 (Rohou and Grigorieff 2015) methods implemented in cisTEM (see figure 5.10 B and C). The CTF corrected images showed severe detergent effects (see Appendix Fig 5.4). 'Meniscus effect' was one of the most common detergent-based effects that were observed in our dataset. Detergents reduce the surface tension of the sample causing the ice to become very thin in the centre and thick towards the edges of the holes, termed as meniscus effect (Kampjut, Steiner et al. 2021). The other artefact included free micelle or long threads of detergents and the concentration of protein along the walls of the holes. Despite replacing the sample in the detergent-free buffer, it seems possible that ultrafiltration using 100kDa cutoff filters concentrated the detergents micelles in the samples rather than filtering out. The use of ultrafiltration columns for buffer replacement has been used successfully used in our previous studies (Thonghin, Collins et al. 2018, Barbieri, Thonghin et al. 2021).

The dataset was manually screened for bad CTF fitting, carbon support and blurred images (see Appendix Figure 5.4). The remaining images showed estimated CTF fit resolutions between 3-6Å. The final set of 2195 images after CTF filtering was used for the ab-initio particle picking method implemented in cisTEM. Characteristic particle radius was set at 68Å (estimated from the atomic model) with a maximum radius of 72Å. The threshold for particle picking was set at 1 for maximum particle extraction. 440,137 particles were picked and were classified into 50 two dimensional classes

masked at a diameter of 180Å (see figure 5.11). Initial 2D classes showed particles in lateral and anterior orientations, demonstrating a clear distinction between the micelle and the TMDs and NBDs (see Figure 5.11A). Further, chameleon eyes like two blobs, characteristic of micelles and carbon supports were also observed. Classes with probable P-glycoprotein particles were selected from the 2D class averages and were subjected to a series of 2D re-classifications to remove the bad particles. Figure 5.11B shows the number of particles extracted after each classification cycle.



*Figure 5.10 Single particle analysis of mouse P-glycoprotein with ivacaftor. (A) Representative micrograph after motion correction with orange circles showing P-gp particles (orange). (B) Power spectrum (Thon ring) of the raw image fitted with theoretical thon rings calculated from CTF calculations. (C) 2D representation CTF fit of experimental power spectrums (green) with theoretical spectrum (yellow). Blue line represent the quality of the CTF fit used to estimate the theoretical resolution.*



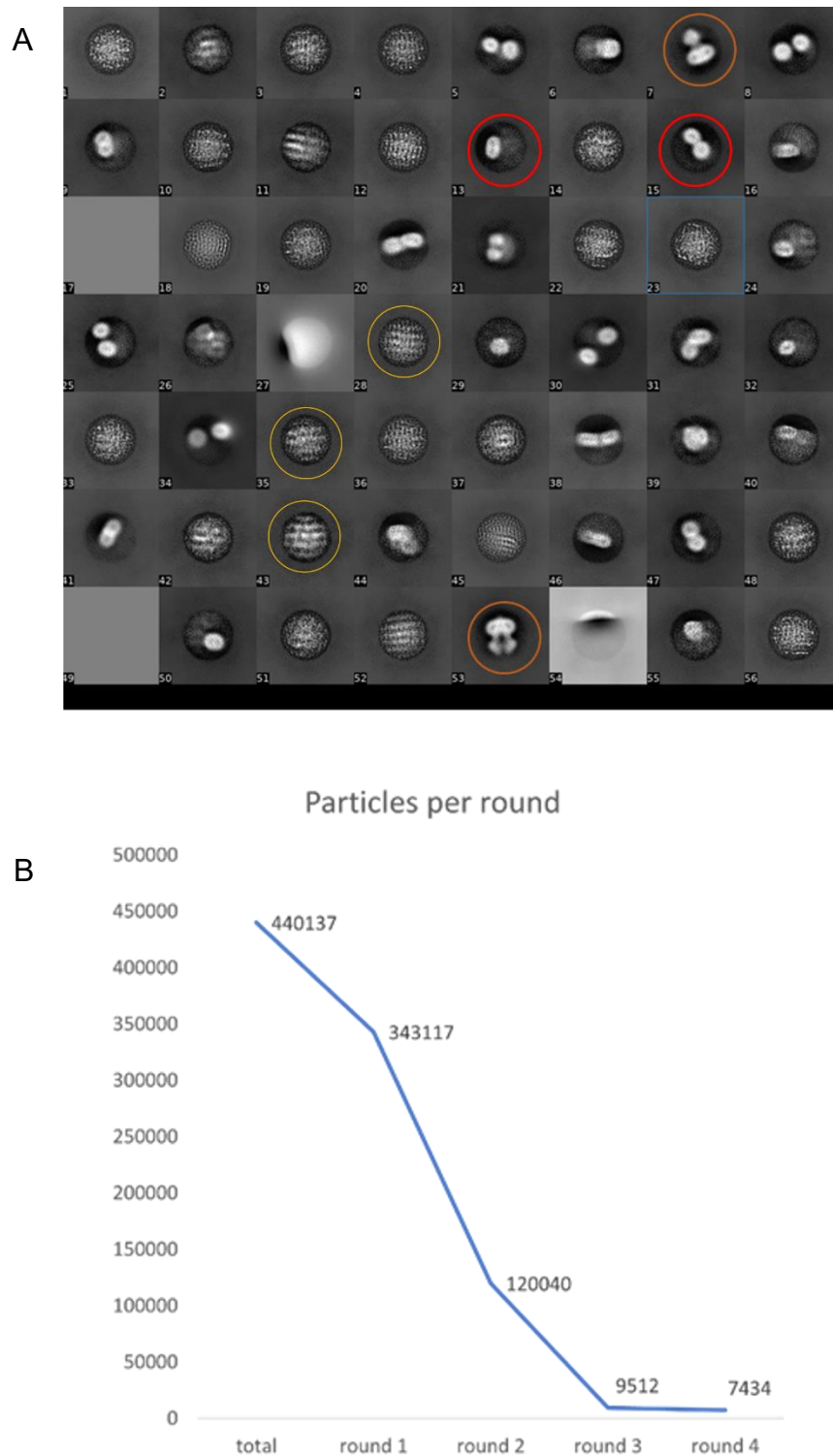


Figure 5.11 2D classification and particle selection. (A) 2D class averages after first round of 2D classification. The orange class averages show a distinct P-glycoprotein structures i.e the TMDs surrounded by micelle and fist like NBDs. The red circles show micelles whereas yellow circles depict carbon support. (B) Number of particles left in the dataset after each round of 2D classification. Initially 440137 particles were picked after 4 rounds of 2D classification we had 7434 particles

The final dataset consisted of 7434 particles. The 2D classes showed P-glycoprotein in an inward-facing conformation. The micelle showed a characteristic disc-like shape with sharp edges and relatively weaker densities in the middle (Thonghin, Collins et al. 2018). This can be explained based on the high-density head group which scatters more electrons as compared to the tail region. The final set of particles was used in *ab initio* 3D classification and model generation in cisTEM. Three *ab initio* models were generated (see figure 5.12). ClassA showed micelle like structure and partially NBD like densities. No densities for TMD like helical structures were resolved. ClassB and classC showed characteristic P-glycoprotein like density structures with disc-like micelle and two bell-like structures connected to the tubular structure. However, due to the small set of particle TMDs were not properly resolved furthermore, densities for TMDs inside the micelles were also missing. Overall models showed approximately 14.79Å resolution at an FSC cutoff of 0.143 (see Figure 5.13).



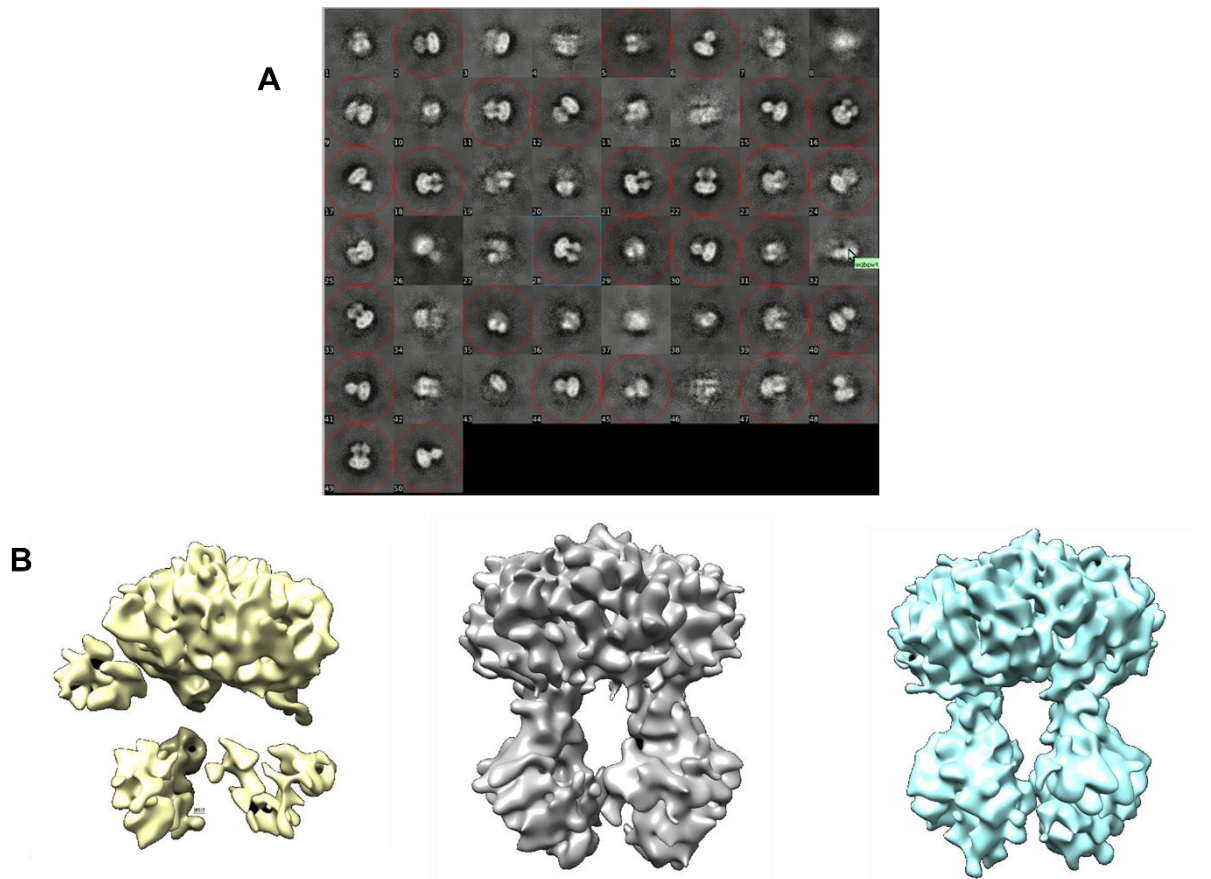


Figure 5.12 2D and 3D classification of mouse P-glycoprotein from 7434 particles. (A) 2D classification of mouse P-glycoprotein demonstrating inward-facing conformations (red). (B) Density maps obtained from 3D classification. ClassA (yellow) classB (grey) and ClassC (cyan).

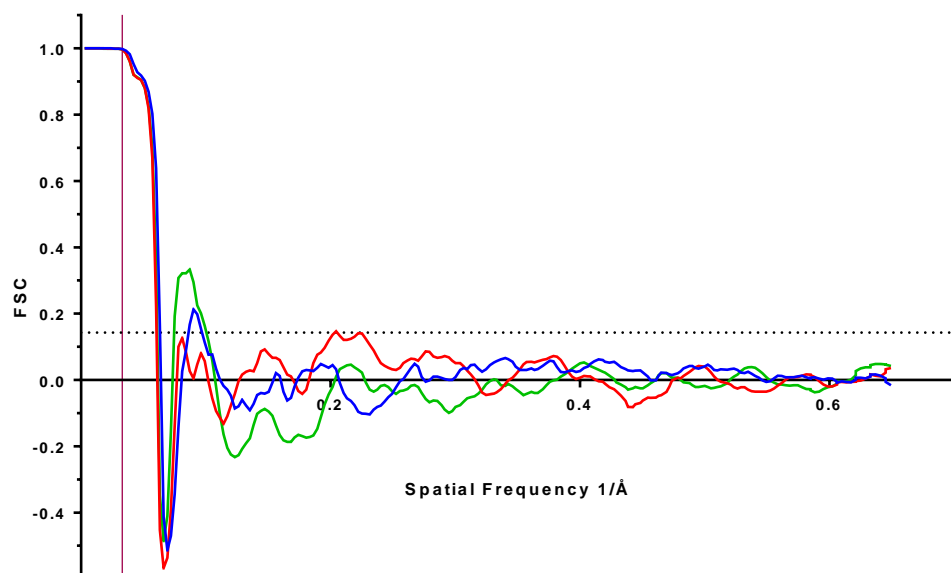
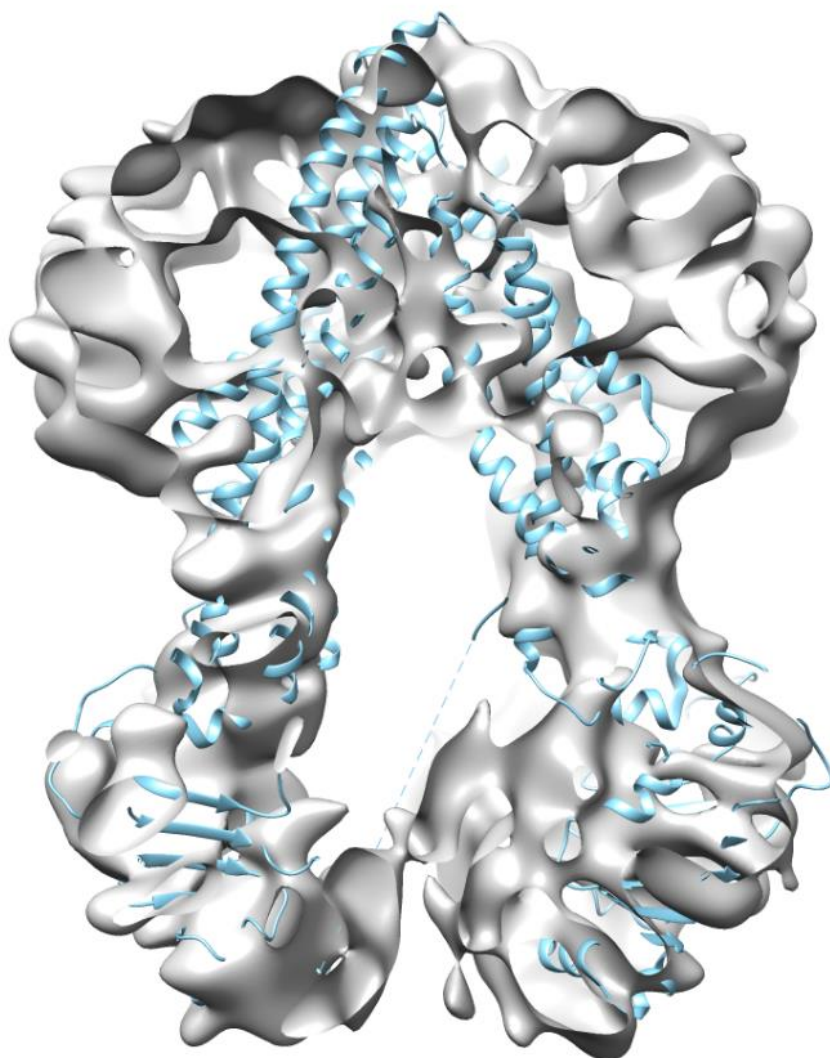


Figure 5.13 Resolution estimation from Fourier shell correlation. The resolution of each structure was calculated using the FSC value of 0.143.

In order to improve the data, another attempt with mouse P-glycoprotein and ivacaftor was made. 1390 images under automated data collection with EPU software were obtained in a small screening experiment. Detergent micelles were once again prevalent components observed in the images. After repeating the aforementioned process of CTF correction, filtering and 2D classification. Additional 6537 particles were collected from the new dataset. Both sets of particles from the first data set and this data set were combined and used for 2D Classifications and 3D classifications.

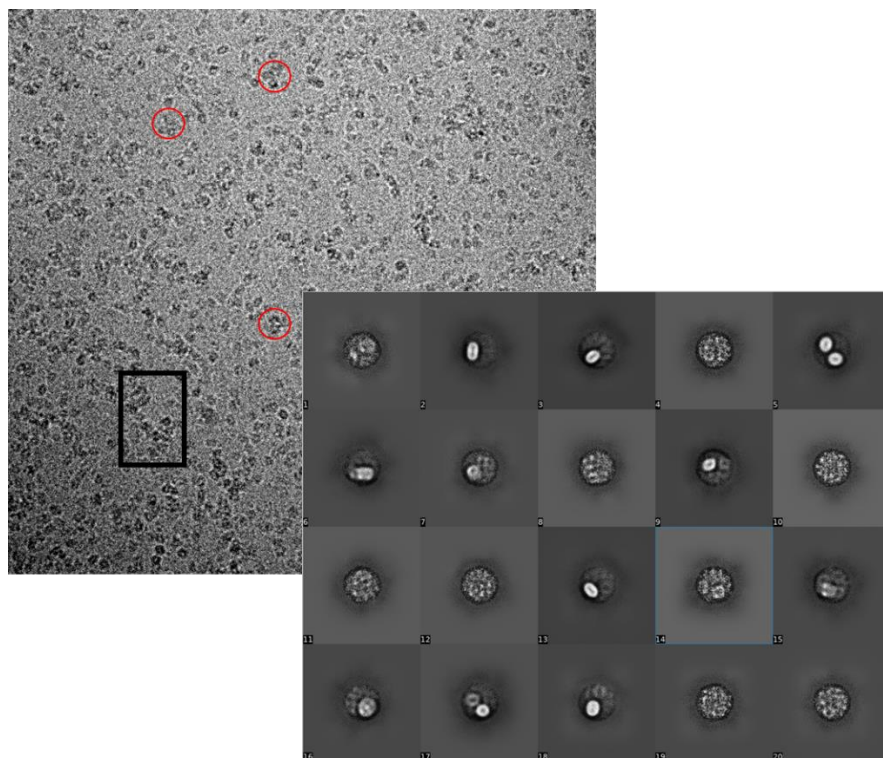
Figure 5.14 shows the electron density of mouse P-glycoprotein generated from the combined dataset. The 3D density map shows a blob-like structure with a disc-like micelle at the top and dumbbell like NBDs at the bottom. Contrary to the 2D classes where we could observe the TMDs present (see figure 5.11A) the density cloud was empty from inside suggesting inability to resolve the TMDs at the 3D classification level (see figure 5.14). This can be due to multiple reasons: (i) skewed CTF estimations because of lensing effects (ii) disappointingly low number particles (iii) presence of detergent micelle around the TMDs. When a similar dataset of approximately 13,000 particles from pre-resolved ABCB6 at 3.2Å was used, it showed better resolved TMDs (see Appendix Figure 5.5) however, it also showed missing regions in TMDs inside the micelles. The refined mouse P-glycoprotein model showed an estimated resolution of 10.8Å.



*Figure 5.14 Electron density map of mouse P-glycoprotein with ivacaftor (grey) fitted with atomic resolution mouse P-glycoprotein (4KSB: cyan).*

## 5.7 Structural studies on ivacaftor binding with human P-glycoprotein

Human P-glycoprotein at a concentration of 1.5mg/ml was replaced in glycerol and detergent free buffer and loaded onto the grids and vitrified using vitrobot Mark IV (as explained in section 2.3.3.1). The data was screened as mentioned in the methodology section. The screening images were imported into CisTEM and calculated for CTF correction and particle picking. 36967 particles were picked using an ab initio algorithm with a particle radius of 80Å (see figure 5.14). Unfortunately, the 2D classification does not show any P-glycoprotein like class. All classes were either micelles or carbon support (see Figure 5.15).



*Figure 5.15. Single-particle analysis of human P-glycoprotein. The red circles represent the probable protein-like particles. The black square can be protein aggregated or contaminated. Approx. 37000 particles were picked and 2D classified into 20 classes.*

It has been reported previously that hydrophobic protein and membrane proteins tend to adhere to the carbon support. Therefore, it was hypothesized that protein might have embedded itself into the carbon support. (Cheung, Kajimura et al. 2013) suggested an alternative method to evade this carbon

support film problem. They proposed that pre-treatment of grids with detergents will mask the hydrophobic environment of grids allowing the protein to go in holes. Subsequently, grids were pre-treated with 0.1% DDM solutions and vitrified samples. Screening of grids showed an uneven distribution of samples on the grid. Furthermore, when holes were screened the ice showed ripple-like effects (Appendix Figure 5.6). Further graphene oxide coated lacey carbon grids were also tested. however, no P-glycoprotein-like particles were observed on the grids. The possible explanation for this can be that human P-glycoprotein partition itself into the air/water interface and aggregated or is blotted off the grid during the vitrification process. Noble et al in a systematic investigation on 31 proteins in Cryo-EM grid preparation have shown that all proteins show more or less tendency of partitioning into an air/water interface (Noble, Dandey et al. 2018). 90% of the proteins at the air/water interface is partially denatured (D'Imprima, Floris et al. 2019). This might be a reason that all the recently published human P-glycoprotein structures are trapped with soluble Fab which reduces the surface hydrophobicity (Alam, Kowal et al. 2019, Nosol, Romane et al. 2020).

## 5.8 Virtual screening

Interestingly, our docking protocol was able to locate the substrate binding cavities within the top ten conformations. Therefore, we used the protocol to screen the library of FDA approved drugs to identify novel classes of drugs. A dataset of known substrates and non-substrates was extracted from the literature and used for optimizing the cut-off threshold.

### 5.8.1 Optimization of cut-off threshold

Virtual screening was performed using CCDC GOLDSuite version 2021. Initial optimization was done using a known data set from (Shukla, Kouanda et al. 2014) and (Marighetti, Steggemann et al. 2013). A total of 100 conformations were generated and were calculated for binding scores. Top scored conformations were exported and analyzed for the cut-off threshold. substrate compounds showed ChemPLP scores ranging between 84.95-66.90. Figure 5.16 shows a comparison between the substrate and non-substrate compounds in ChemPLP and GOLDScores. Chemplp showed mean values of 76.05

with a 25% percentile value of 70.4. whereas non-substrate compounds showed a mean of 62.5 and 75% values were below 66.3. In comparison, GOLDScore was not able to classify substrate and non-substrate compounds. Both classes showed mean score values of 44.12 and 39.19. With standard deviations of approximately 6 scoring values demonstrate a strong overlap between substrates and non-substrates. Therefore, ChemPLP with the cut-off threshold of 70 (from 75% percentile) was used for final library screening.

## Optimization of ABCB1 virtual screening

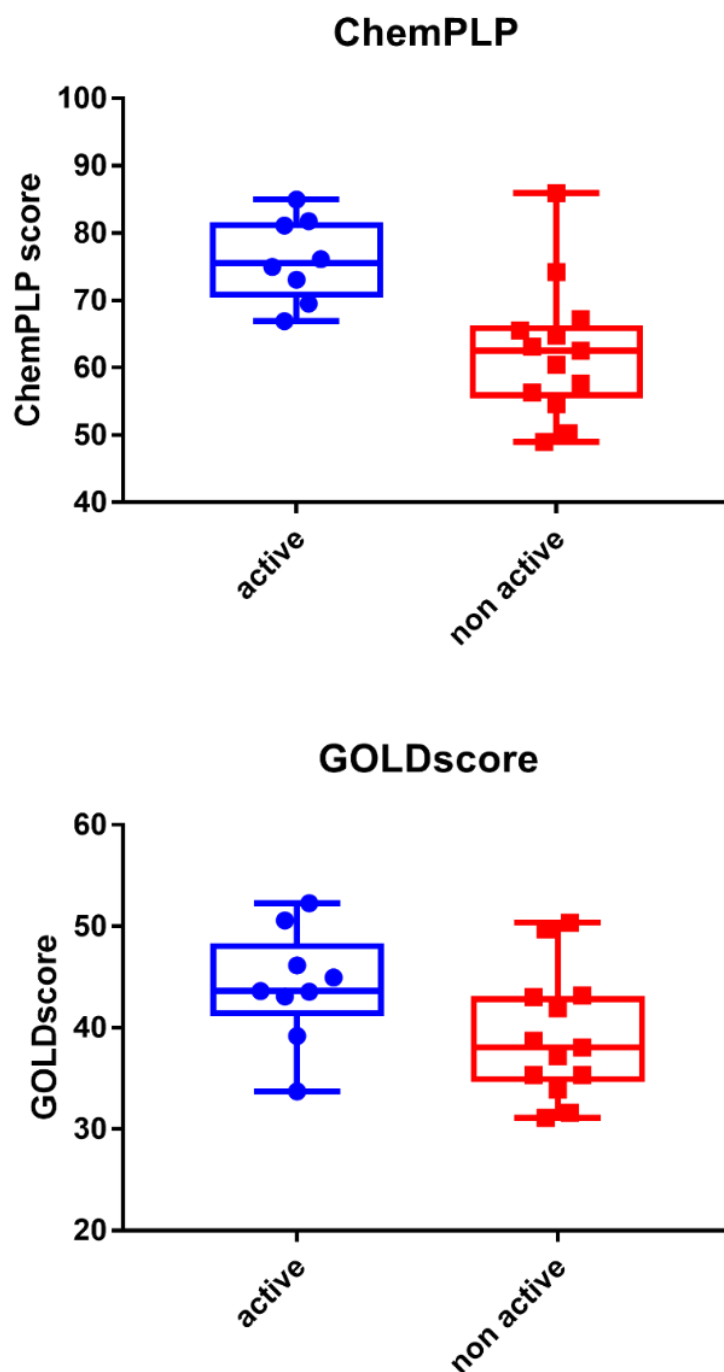


Figure 5.16 Graphical representation of Optimisation of cut off threshold for ABCB1 virtual screening. The scoring values of top conformation of each molecule from the active and the non active was used in plotting a box and whisker plot (top panel) Distribution of ChemPLP scores in active and non-active compounds. (bottom panel) Distributions of GOLDscore

### 5.8.2 Library screening

The Library of FDA approved drugs (L1021) from APEXBio was downloaded and screened using the optimized protocol. A total of 154 drugs were identified during the virtual screening (see appendix table for complete list). Including 36 drugs from the kinase class of the drugs and 118 drugs from non-kinase class. Because kinase inhibitors are the most studied modulators for MDR proteins and they were the largest class of drug identified in the hits. Therefore, hits were classified into two classes: kinase and non-kinase. Figure 5.17 shows some of the major classes of hits identified. Several studies have been reported modulations of ABCB1 via kinase inhibitors. Our protocol also suggests kinase targeting drugs interact with ABCB1 with 16/36 kinase inhibitors showed top ten conformations above the threshold. EGFR inhibitors showed the strongest hits with 4/5 drugs showing top ten conformations above the threshold. Antivirals protease was the biggest class of inhibitors identified from the non-kinase category. 8/13 antiviral hits observed were protease inhibitors from HCV and HIV. CFTR correctors were also an interesting class identified from the screening experiment.



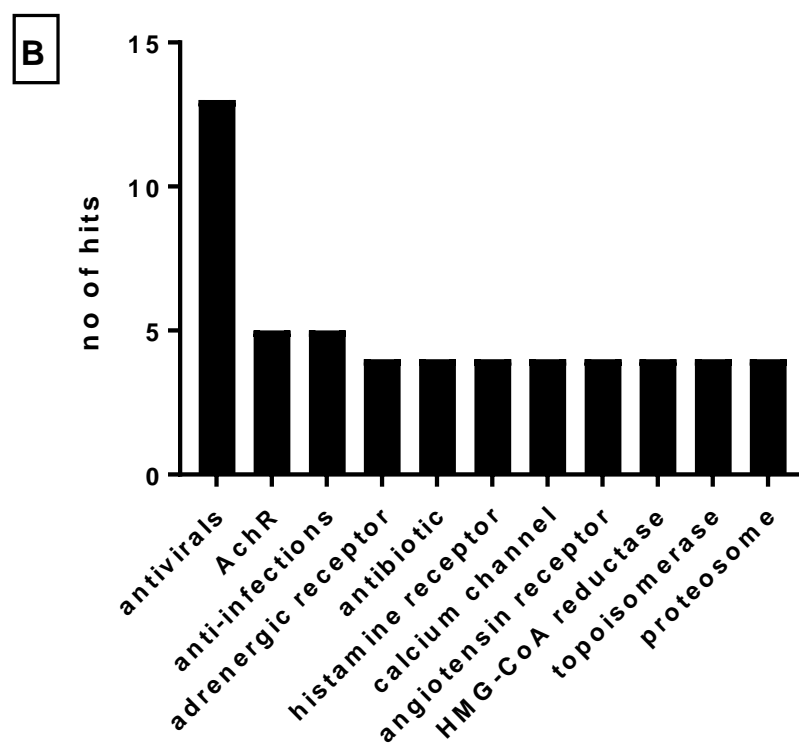
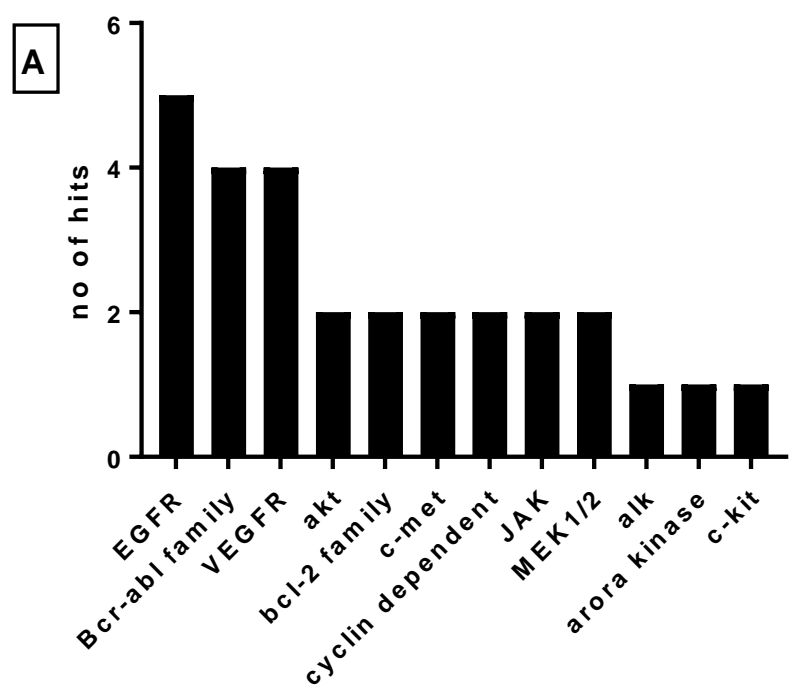


Figure 5.17 Histogram representation of major drug hits according to their targets. (A) Hits from kinase class. (B) Hits from non-kinase class

## 5.9 Discussion and conclusions

ABCB1 is one of the most studied membrane transporter proteins due to its role in substrate translocation and drug-protein interaction. Furthermore, due to its role in chemotherapeutic resistance, it has also been a key target for modulating multidrug resistance. In the last two decades, a plethora of studies has been reported to decode the underlying molecular mechanisms and ligand-protein interactions. Recently our lab demonstrated that ivacaftor a CFTR corrector can simulate human P-glycoprotein two times the simulated activity of verapamil, a gold standard molecule for P-glycoprotein simulations studies (Lingam, Thonghin et al. 2017). Here in this study we used thermal stability assay to demonstrate interactions of ivacaftor with mouse P-glycoprotein. Thermal unfold assays have been used to study ligand-protein interactions (Senisterra, Chau et al. 2011, Gose, Shafi et al. 2020). The protein is heated in presence and in absence of a ligand. The ligand stabilizes the protein therefore more protein is recovered, as compared to the sample where ligand was absent. Ivacaftor showed a  $K_d$  value of 1 $\mu$ M with mouse P-glycoprotein. This was similar to the  $IC_{50}$  value observed in Hoechst 33342 transport assay with human P-glycoprotein (Lingam, Thonghin et al. 2017). It was important to note that ivacaftor is a highly lipophilic compound and practically insoluble in water. Further, the ligand protein interaction of ivacaftor with mouse P-glycoprotein was studied using the molecular docking and Cryo-EM

In recent times cryo-electron microscopy has been effective in structural studies for membrane transporters that are not compatible with alternative methods like crystallography (Goldie, Abeyrathne et al. 2014). However, oftentimes these structures limit themselves to mid resolutions to provide atomic models through chain tracing. This becomes more challenging when working with subtle conformational dynamics studies or ligand-protein interactions. Below 5Å resolutions ligands look like blobs where fixing ligand orientations becomes challenging (Chakraborti, Hatti et al. 2021, Zhao and Tajkhorshid 2021). Our cryo-EM results from mouse P-glycoprotein with ivacaftor also showed mid resolution structures with one of the five structures showing additional density for ligands. However, due to low resolutions and sub-stoichiometric concentrations, the density was not very well

resolved. This density was close to the entry portal and was unique as no other mouse or human P-glycoprotein structure has demonstrated a lower binding cavity close to the entry portals. Thus elucidating this cavity was very important.

Docking algorithms have progressed leaps and bounds in exploring ligand-protein interactions. These algorithms can be vital in exploring the binding site as well as ligand-protein interactions. Therefore, we employed docking to explore the unbiased ligand-protein interactions profile. The results showed two probable binding sites from top-ranked conformations and consensus scoring conformation (Figure 5.3A and Figure 5.4A). The consensus scoring solution showed a similar binding site for ivacaftor as found in cryo-EM structure (figure 5.7A). However, the orientation of ivacaftor in docking conformations and Phenix (an automated ligand fit method for structural studies) fixed conformation were inverted. One possible explanation to the fact that in docking protein is treated in isolation whereas in cryo-EM protein was in solvent environment with water and ionic molecules able to access the binding cavity. Often at mid to low resolutions, these molecules are not resolved. These docking results also suggested polyspecificity of P-glycoprotein the top score conformations showed a central binding cavity which was consistent with the binding cavity identified in structural studies (Li, Jaimes et al. 2014, Nosol, Romane et al. 2020). Other binding site was close to the entry portal which is consistent with cysteine mutagenesis studies (Loo and Clarke 2002). Furthermore, our ligand-bound structure also showed re-arrangement of helix near the extracellular loop giving an insight to a unique inward state observed for the first time. Further, experiments to obtain higher resolution structures are required on this helical rearrangement to elaborate whether there is an induced fit model or these features arose from imprecision of molecular fitting. Furthermore, successful structural studies with molar excess ivacaftor can also elucidate the polyspecificity.

Molecular docking has been one of the most used methodologies for computer-assisted drug design. Several studies have used molecular docking approaches in understanding ligand-protein interactions and virtual screening. Structure-based drug design has been used successfully in identifying novel

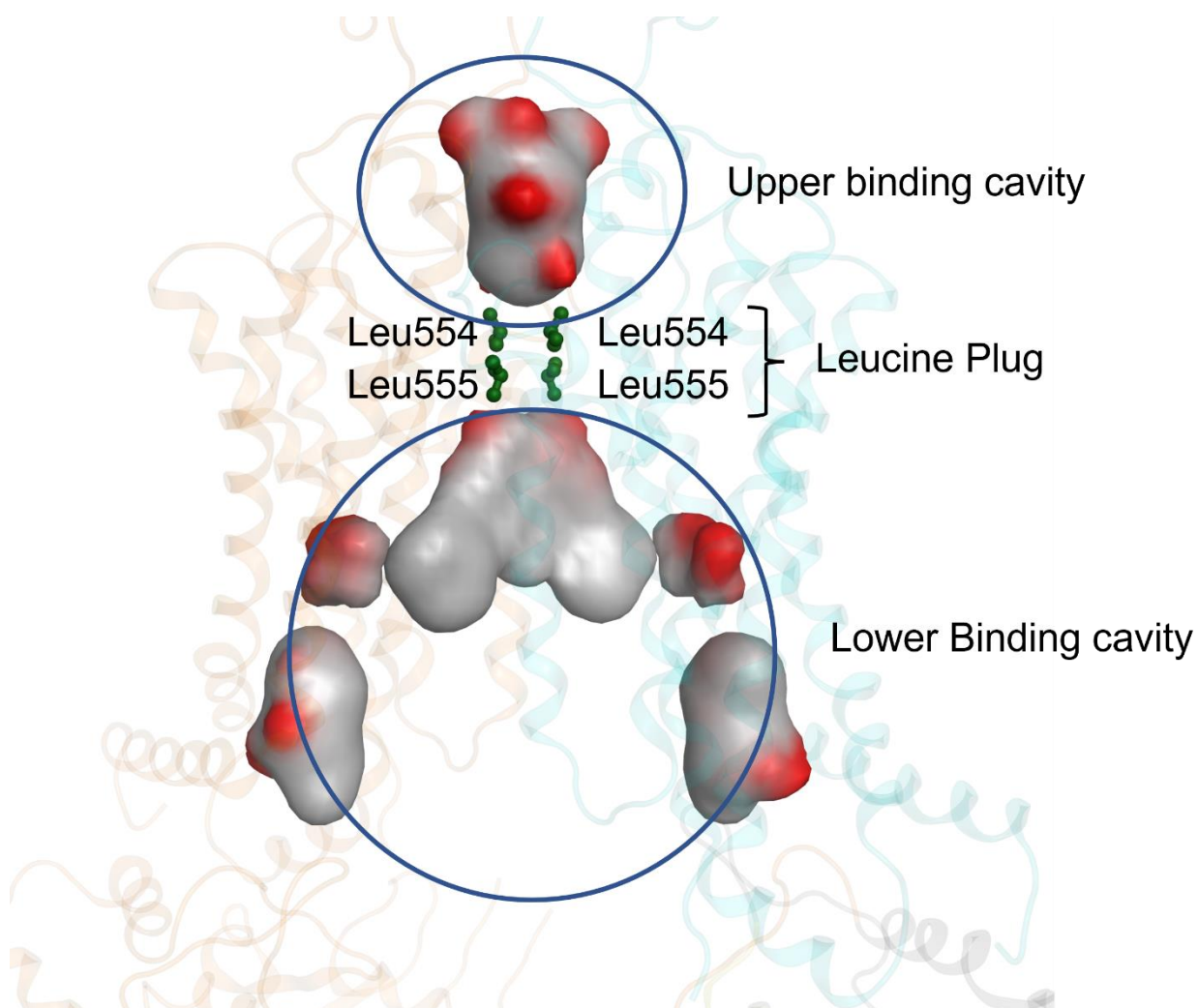
compounds. Recently, (Marques, Šupolíková et al. 2021) used knowledge-based virtual screening to identify novel flavonoids targeting P-glycoprotein. We also used the empirical method to optimize our docking protocol. Our virtual screening protocol suggested HCV and HIV antivirals were the biggest class identified from non-kinase hits. A recent study has also identified that HIV and HCV antivirals modulate rhodamine transport via ABCB1 (Martinec, Huliciak et al. 2019). CFTR correctors, ivacaftor and VX-661 (tezacaftor) were also an interesting class of drugs identified from hits. Ivacaftor has been reported to simulate ABCB1 activity previously in our lab, furthermore, in the aforementioned study, we have also tried to elucidate the ligand-protein interaction of ivacaftor with P-glycoprotein.

## Chapter 6, ABCG2: mechanistic in-sights to ligand-protein interaction

The structural studies on ABCG2 have delineated ABCG2 as a homo-dimeric structure with two large central cavities divided by leucine plugs (see Figure 6.1) (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018, Orlando and Liao 2020). These structures of ABCG2 have shown binding of drugs in the lower binding cavity between TMH1, 2, and 5 (see Figure 6.1). Furthermore, these structures also have suggested multiple binding modes of substrates with the binding cavity. ABCG2 selective inhibitor Ko143 derivative shows binding of two molecules with each molecule bound along with TMH2 and 5 on each monomer (Jackson, Manolaridis et al. 2018) whereas other compounds bind in a 1:2 molar ratio of drug to protein demonstrating an arch-like conformation between the two monomers (Orlando and Liao 2020) (see Figure 1.10, page 32). Despite much work done on the structural aspect in recent times, at present very little is known about the governing factors that contribute to these ligand-protein interactions. This information is vital in understanding the underlying mechanism of drug-protein interaction and translocation pathways (Li, Koh et al. 2019).

ABCG2 is one of the prime culprits for causing multidrug resistance (MDR) in multiple cancers (Doyle, Yang et al. 1998, Miyake, Mickley et al. 1999). Since the discovery of the first ABCG2 inhibitor Fumitremorgin (FTC) (Rabindran, Ross et al. 2000) several other classes of inhibitors have been reported. In the past decade, a plethora of work on the modulation of ABCG2 has been reported (Cramer, Kopp et al. 2007, Marighetti, Steggemann et al. 2013, Shukla, Kouanda et al. 2014, Jackson, Manolaridis et al. 2018). These modulators included natural flavonoids (Pick, Müller et al. 2011) antivirals/antibiotics (Shiozawa, Oka et al. 2004, Weiss, Rose et al. 2007) and chemotherapeutics (Mazard, Causse et al. 2013). In recent times, tyrosine kinase

inhibitors have been the biggest class of ABCG2 modulators under consideration (Krchniakova, Skoda et al. 2020). Lapatinib and erlotinib have been reported to inhibit



*Figure 6.1 Diagrammatic representation of ABCG2 binding cavity. The cavities were calculated with LIGSITE implemented in MOE. ABCG2 demonstrate a large central binding cavity divided into two cavities via a leucine plug (green). The two cavities are represented as the upper binding cavity and lower binding cavity here.*

ABCG2 in sub-micromolar concentrations in vitro (Shi, Parmar et al. 2009, Shukla, Kouanda et al. 2014). Despite a lot done on identifying new modulators for ABCG2, very little information is available on mechanisms of drug binding and substrate translocation. This emphasizes the pressing need of understanding the underlying molecular mechanism. This information can lead to a successful drug discovery process. Therefore, here in this project, we used different *in-silico* and wet lab approaches (The later in collaboration with Tomoka Gose at Schuetz lab, St Jude Children Hospital, Memphis) to elucidate the key features involved in molecular mechanisms of drug transport.

## 6.1 Preliminary analysis with Kinase inhibitors

Preliminary cellular thermal unfolding experiments on ABCG2 expressing MEK microsomes were conducted in Schuetz's Lab. The sample was heated in the presence and absence of various kinase inhibitors and quantified for protein stability/recovery on a western blot. The untreated sample was used as a control for comparing the protein recovery. Higher protein recovery suggests higher stabilization of ABCG2 implying stronger ligand-protein interactions. Lapatinib and nilotinib showed maximum stability of protein suggesting strong interactions with ABCG2. Further, when lapatinib along with its analogues was analysed (Figure 6.2B), all analogues were able to recover protein significantly as compared to the untreated control, except for arry380 which showed minimal recovery of ABCG2. Similar results were observed when these analogues were analysed with pheophorbide-A accumulation for transport inhibition using hABCG2 expressed mouse embryonic fibroblast (MEK) cells (Figure 6.2A). Lapatinib showed an  $IC_{50}$  of  $0.085\mu M$  and its thiazole analogue (GW583340) showed  $IC_{50}$  values of  $0.096\mu M$  (Figure 6.1A). Four out of five analogues showed  $IC_{50}$  values in the sub-micromolar range. These results proposed lapatinib and its analogues as interesting candidates to study the ligand-protein interaction with ABCG2. A series of docking

experiments were performed to optimize a docking protocol which was later used to dock these compounds.

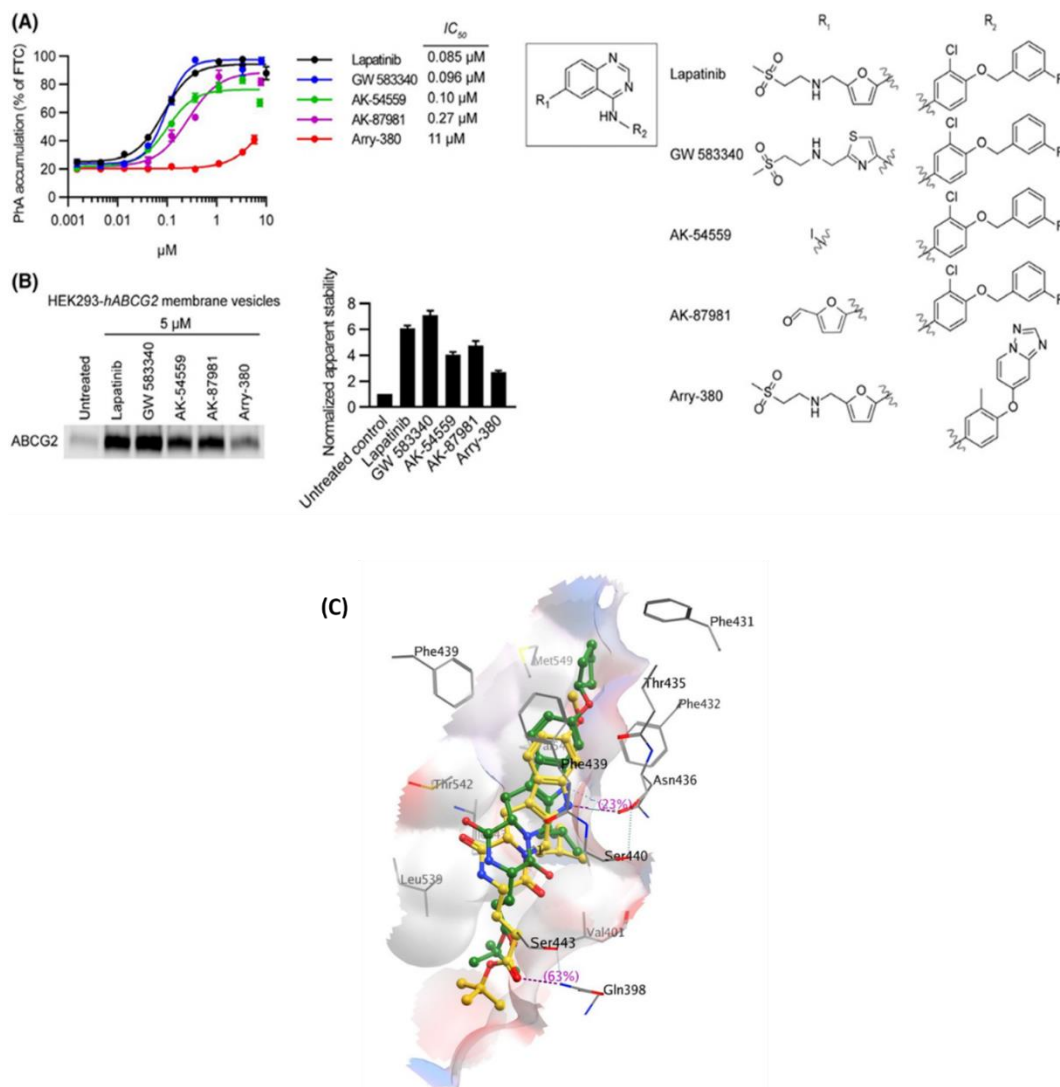


Figure 6.2. Ligand protein interaction of lapatinib analogues with ABCG2. A) Concentration dependent analysis of Pheophorbide A accumulation due to G2 inhibition. Lapatinib shows strong inhibition whereas arry-380 shows the weakest inhibition (performed by Schuetz's Lab). B) Ligand protein interaction analysis using cellular thermal shift assay. Protein was heated at 62°C for three minutes in the absence or presence of 5 $\mu\text{M}$  drug. The protein recovery was analysed with Western blotting using anti-ABCG2 (BXP-53) antibody (performed by Schuetz's Lab). C) Diagrammatic representation of substrate protein interactions from docking studies. binding of ko143 (yellow) as compared to experimentally determined MZ29 (ko143 analogue, green) in optimized protocol.



## 6.2 Docking of Lapatinib derivatives

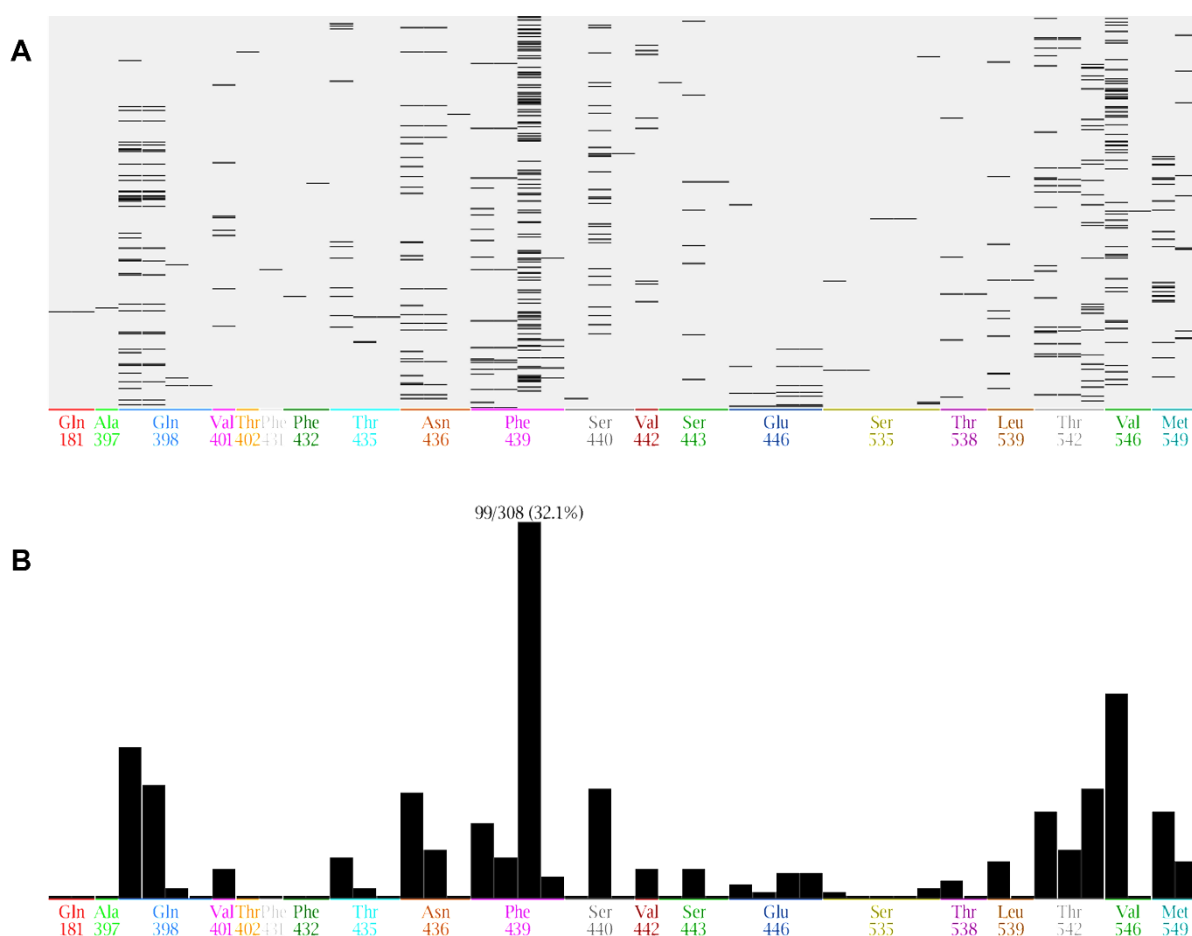
The docking protocol was optimized via re-docking the Ko143 analogue (MZ29) along with Ko143 with the inward-facing ABCG2 structure bound with MZ29 (PDBID: 6ETI) (Jackson, Manolaridis et al. 2018) using different combinations of placement and scoring methods in MOEv2019 (Molecular Operating Environment 2021) as explained in the methodology section 2.3.4.2.1. The binding site was generated using the site finder algorithm and manually reduced to the large central cavity identified from cryo-EM excluding the surface pockets (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018, Orlando and Liao 2020). A combination of four different placement methods along with two different scoring methods was used for optimization. Alpha-sphere and Triangle-matcher showed docking of MZ29 near the intracellular loops (see Appendix Figure 6.1), as compared to the experimentally resolved binding cavity (Jackson, Manolaridis et al. 2018) thus were excluded. The proximity-triangle method and alpha-triangle were able to find the same binding cavity (Appendix Figure 6.1). Further, when conformers were ranked against the scoring functions, affinityDG along with the alpha-triangle methods was able to identify the conformation similar to the conformation identified during the cryo-EM experiments in the top conformation (see figure 6.2C). Proximity-triangle along with affinityDG were able to find the experimentally determined conformation within the top 10 confirmations. LondonDG scoring function showed conformation extending between the two monomers. Therefore, alpha-triangle along with the scoring function of affinityDG were used for probing the ligand-protein interactions of Lapatinib analogues.

Ligand-protein interaction of docked conformations were analysed using Protein-ligand interaction fingerprint (PLIF). The fingerprint profile was generated using the PLIF algorithm

implemented in MOEv2019 (Molecular Operating Environment 2021). The prepared ABCG2 protein (6ETI) was imported into the MOE interface and the docked conformation was individually assessed for interaction with binding cavity residues. These interactions were represented as barcode maps with each mark representing the interaction with the corresponding residue. Figure 6.3 shows the interaction fingerprint of lapatinib analogues in the ABCG2 binding cavity. The results showed that Phe439 was the most common interactor for lapatinib-like analogues. When the frequency of Phe439 interaction was calculated approximately 1/3 conformations showed direct interaction (see Figure 6.3b) with Phe439.  $\pi$ - $\pi$  and  $\pi$ -proton interaction was the most common interaction observed with Phe439 side chain. Whereas a small proportion of conformations also showed backbone interactions with Phe439. Further, when each conformation was manually analysed, 90% conformations showed Phe439 present within 5Å distance suggesting probable interaction with the drug. In addition to Phe439, Val546, Ser440, Asn436, and Gln398 were also shown to be prevalent interactors. These results implies the pivotal role of Phe439 in ligand-ABCG2 interaction where ligands anchor at Phe439 and the remaining of the ligand arrange itself accordingly in the binding cavity.

To further analyse the specific ligand-protein interactions, Algorithmic hierarchical clustering analysis based on the common scaffold was employed. The common scaffold clustering showed a total of 145 clusters at cut-off levels of 1Å RMSD in the common scaffold. For cluster to be considered the most significant cluster it had to fulfil these two basic criteria (i) contains maximum number of docked drugs, (ii) comprises top-ranked conformations from scoring functions. A total of 12 clusters showed 4/5 or 5/5 docked drugs. One cluster contained 5/5 drugs whereas, 11 clusters showed 4/5 drugs docked. The clusters were further reduced based

on scoring functions. Only three clusters showed conformations that were represented by the top 10 conformations from the scoring functions. Cluster2 contained 5/5 docked drugs plus top-ranked conformations from lapatinib and arry380 were also present in this cluster (see Figure 6.4). The closest cluster was cluster1 which contained 4/5 drugs with top-ranked conformations of arry380 and gw583340. However, both clusters showed a similar binding cavity around Ph439.



*Figure 6.3. Protein ligand interactions fingerprint of lapatinib analogues. A) Barcode representations of PLIF where the x axis shows residues and the y axis represent each conformation. Each bar represents interaction of that residue in that conformation. B) Frequency histogram of PLIF interactions. Representing number of conformations that interacts with the residue shown at the bottom.*

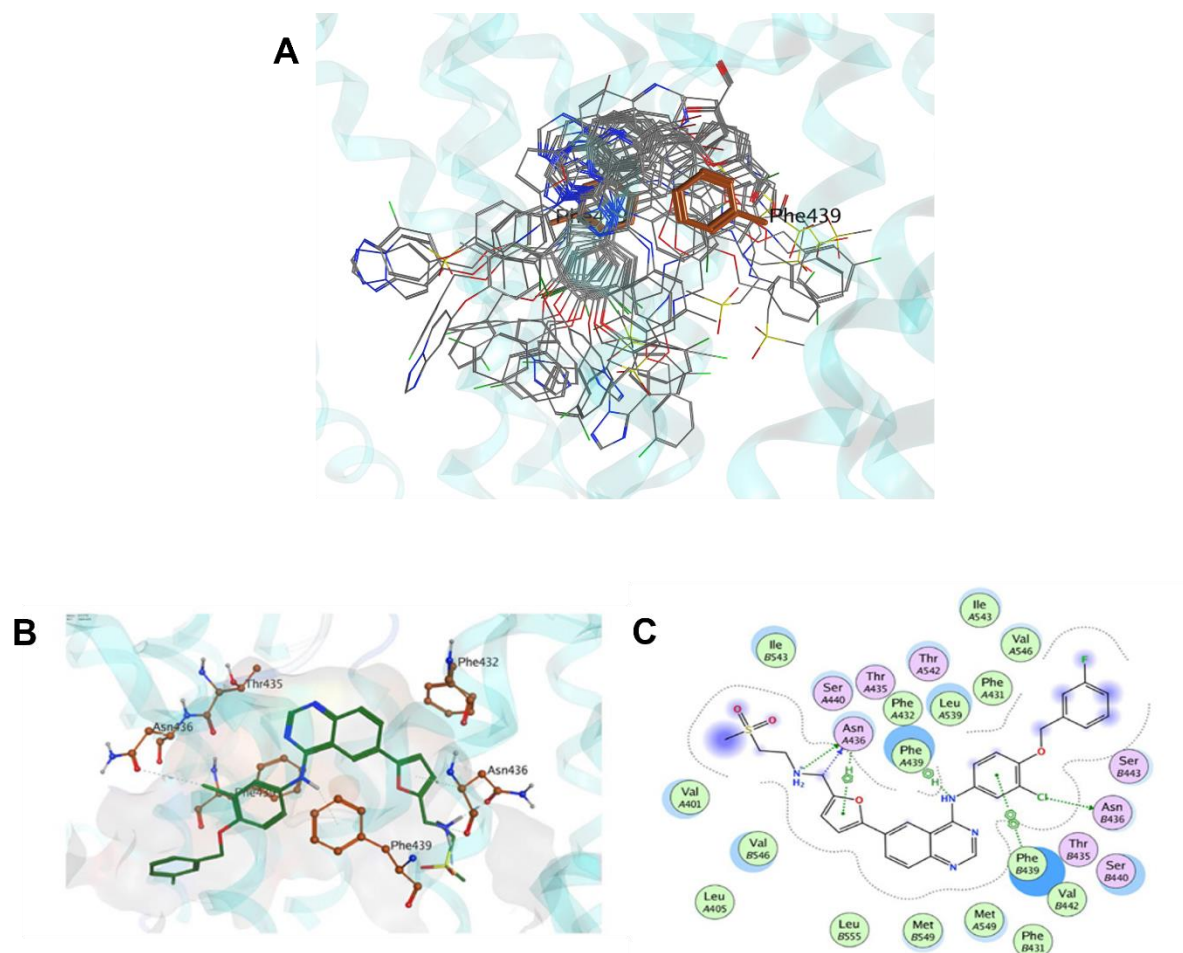


Figure 6.4 Graphical representation of ligand-protein interaction. A) Conformation of cluster2 in the ABCG2 binding cavity. B) 3D representation of lapatinib in ABCG2 binding cavity. C) 2D representation of lapatinib fit in ABCG2 cavity. The dotted line demonstrates the envelope of the drug binding cavity.

When ligand-protein interaction of lapatinib analogues in cluster2 was analysed. It showed an arched-like conformation between the two monomers with the central quinazolinylamine sandwiched between the two Phe439 residues, contributed from each monomer (see Figure 6.3 B, C). The top-ranked lapatinib conformation also showed the same binding cavity between the Phe439. However, its orientation was flipped due to the symmetrical structure of the binding cavity. Overall, the lapatinib showed good fit in the binding cavity with hydrophobic parts buried near hydrophobic residues like Val546, Ile543, and Leu539 on TM5, whereas hydrophilic moieties in ligands were surrounded by hydrophilic residues like Thr435, Asn436, Ser440, and Ser443 on TM2 (see Figure 6.4C).

### 6.3 Evolutionary conservation of Phe439

Aligned sequences of ABCG2 from different species were downloaded from the online database Aminod (Chang, Guo et al. 2018). Paralogues of the human ABCG family were downloaded from UniProt and were aligned using Clustal-Omega (Sievers, Wilm et al. 2011). The aligned sequences were imported into the sequence annotation software Jalview (Waterhouse, Procter et al. 2009). Figure 6.5 shows the multiple sequence alignment of ABCG2 among species and within the human ABCG family. Interestingly, Phe439 showed evolutionary conservation amongst species from frogs to mammals. Similarly, the presence of an aromatic residue at position 439 in the ABCG family was also conserved, where all ABCG family proteins showed the presence of Phenylalanine except ABCG5 where it was replaced with another aromatic residue, tyrosine. This evolutionary conservation of an aromatic residue at position 439 is consistent with its probable role in the substrate translocation pathway.

			439F	
<i>Homo_sapiens_human</i>	L F F L T T N Q C F	S S V S A V		
<i>Pan_troglodytes_champenzee</i>	L F F L T T N Q C F	S S V S A V		
<i>Nomascus_leucogenys_gibbon</i>	L F F L T T N Q C F	S S V S A V		
<i>Gorilla_gorilla</i>	L F F L T T N Q C F	S S V S A V		
<i>Papio_anubis_baboon</i>	L F F L T T N Q C F	S S V S A V		
<i>Chlorocebus_sabaeus_greenmonkey</i>	L F F L T T N Q C F	S S V S A V		
<i>Callithrix_jacchus_marmoset</i>	L F F L T T N Q C F	S S V S A V		
<i>Otolemur_garnettii_galago</i>	L F F L T T N Q C F	S S V S A V		
<i>Dasypus_novemcinctus_armidello</i>	L F F L T T N Q C F	S S M S S M		
<i>Loxodonta_africana_elephant</i>	L F F L T T N Q C F	S S V S A V		
<i>Bos_taurus_cow</i>	L F F L T T N Q C F	S S V S A V		
<i>Ovis_aries_sheep</i>	L F F L T T N Q C F	S S V S A V		
<i>Sus_scrofa_pig</i>	L F F L T T N Q C F	S S V S A V		
<i>Canis_familiaris_dog</i>	L F F L T T N Q C F	S S V S A V		
<i>Mustela_putorius_furo_ferret</i>	L F F L T T N Q C F	S S V S A V		
<i>Ailuropoda_melanoleuca_gaint_panda</i>	L F F L T T N Q C F	S S I S A V		
<i>Felis_catus_cat</i>	L F F L T T N Q C F	S S V S A V		
<i>Erinaceus_europaeus_europeanhedgehog</i>	L F F L T T N Q C F	S S V S A V		
<i>Rattus_norvegicus_brownrat</i>	F F F L T T N Q C F	T S V S A V		
<i>Mus_musculus_mouse</i>	L F F L T T N Q C F	S S V S A V		
<i>Tursiops_truncatus_dolphin</i>	L F F L T T N Q C F	S S V S A V		
<i>Sarcophilus_harrisii_tesmaniandevil</i>	L F F L T T N Q C F	S S V S A V		
<i>Monodelphis_domestica_opposum</i>	L F F L T T N Q C F	S S V S A V		
<i>Choloepus_hoffmanni_sloth</i>	L F F L T V D Q C F	S S V W A V		
<i>Ornithorhynchus_anatinus_platypus</i>	M F F L T T N Q C F	S S I S A I		
<i>Tarsius_syrichta_tarsiur</i>	L F F L T T N Q C F	S S V S A V		
<i>Anas_platyrhynchos_mallard</i>	M F F L T T N Q C F	S S I S A I		
<i>Ficedula_albicollis_flycatcher</i>	M F F L T T N Q C F	S S V S A I		
<i>Taeniopygia_guttata_zebrafinch</i>	M F F L T T N Q C F	S S V S A L		
<i>Xenopus_tropicalis_clawedfrog</i>	L F F V T T N Q C F	S S V S A I		
<i>Procapra_capensis_hyrex</i>	L F F F T T N Q C F	T S V S A V		
<i>Meleagris_gallopavo_turkey</i>	L F F V T T N Q C F	S S V S A I		
<i>Gallus_gallus_junglefowl</i>	L F F V T T N Q C F	S S V S A I		
<i>Echinops_telfairi_hedgehog</i>	L F F L V T N Q C F	S S V S A V		
				439F
		<i>hABCG1</i>	L F F S M L F L M F	A A L M
		<i>hABCG2</i>	L F F L T T N Q C F	S S V S
		<i>hABCG4</i>	L F F S M L F L M F	A A L M
		<i>hABCG5</i>	L Y Q F V G A T P Y	T G M L
		<i>hABCG8</i>	L F M I G A L I P E	N V I L

Figure 6.5. Sequence analysis ABCG2. (Left) <Multiple sequence analysis of ABCG2 orthologues from different species. F439 was conserved throughout the evolution from frogs to mammals. (Right) Sequence analysis of human ABCG2 with other members of human ABCG family.

## 6.4 Docking of lapatinib with $\Delta$ F439

It was evident from PLIF and cluster analysis that Phe439 plays an important role in ligand-protein interaction in ABCG2. Further, it was interesting to test whether Phe439 makes hydrophobic interactions or if the aromatic ring is involved in  $\pi - \pi$  interactions. Therefore, we modelled energy minimized F439A, and F439Y mutant models using structural constraints implemented in Modeller (Sali and Blundell 1993). Alanine mutant was selected because it is a hydrophobic amino acid with an aliphatic side chain therefore can not contribute for  $\pi - \pi$

whereas, tyrosine belongs to the aromatic class of amino acids and also the sequence of analysis of human ABCG family showed the presence of tyrosine (Figure 6.5). A total of 100 structures were generated and ranked according to the DOPE score (Shen and Sali 2006) and GA341 score (Melo, Sánchez et al. 2002). Models with the lowest DOPE score and highest GA341 scores were used for docked lapatinib. Figure 6.6 shows a box and whisker plot of binding scores of lapatinib with different Phe439 models. Docking of lapatinib with F439A shows a decrease in the docking scores whereas, this decrease was restored when Phe439 was replaced with tyrosine (an aromatic amino acid) these results are in consistent with the proposition that  $\pi - \pi$  interactions play a vital role in substrate interactions and when mutated can lead to decrease interactions of substrates.

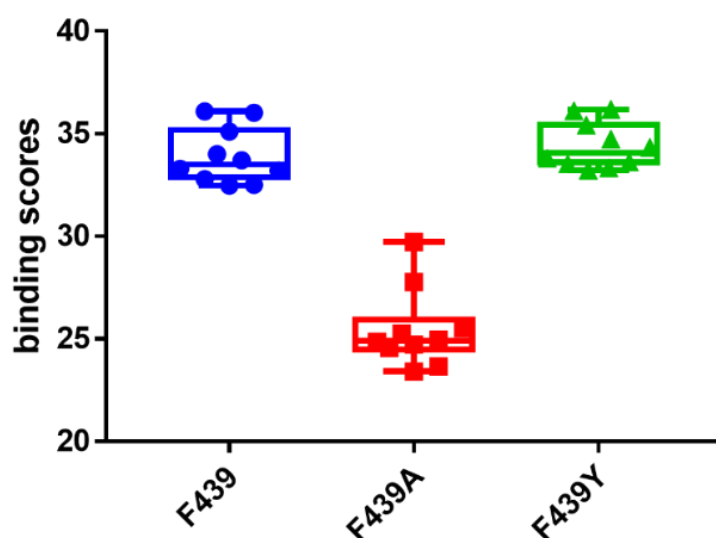


Figure 6.6. Box and whisker plot representation of binding scores of top ten conformations of lapatinib with different Phe439 mutant models. Both F439 WT (blue) and mutant F439Y (green) showed similar scoring values for lapatinib. The binding scores drops when lapatinib was docked with F439A mutated (red).

## 6.5 Docking of structurally different substrates:

To evaluate the role of Phe439, structurally diverse ABCG2 substrates (see appendix) were docked using the optimized protocol and then a ligand-protein interaction profile was generated using PLIF (see Appendix Figure 6.2). Once again arene interaction with Phe439 was the dominant interaction observed. When top-ranked conformations were analysed. All conformations were sandwiched between the two Phe439 residues with the remaining molecule adjusted in the large binding cavity (see Figure 6.7). Other than the Phe439, Val546 showed  $\pi$ -proton interaction. Thr435, Asn436, and Thr542 provided hydrogen bonding capacity. These results suggested a common mechanism of ligand-protein interaction in ABCG2 where Phe439 provided the anchor and remaining ligand arrange themselves in the binding cavity. Interestingly, PLIF fingerprints showed hydrogen bond donor capacity on ABCG2 ligands as a dominant interactions force. However, no hydrogen bond acceptor capacity was identified.

Figure 6.7 shows the ligand-protein interaction of structurally diverse compounds in the binding cavity. All compounds showed a sandwich conformation between the two Phe439 residues from each monomer. Mitoxantrone also showed hydrogen bonding with Asn436. Whereas topotecan showed interaction with Thr542. These residues were also identified from PLIF analysis of Lapatinib analogues with ABCG2 (see figure 6.2).



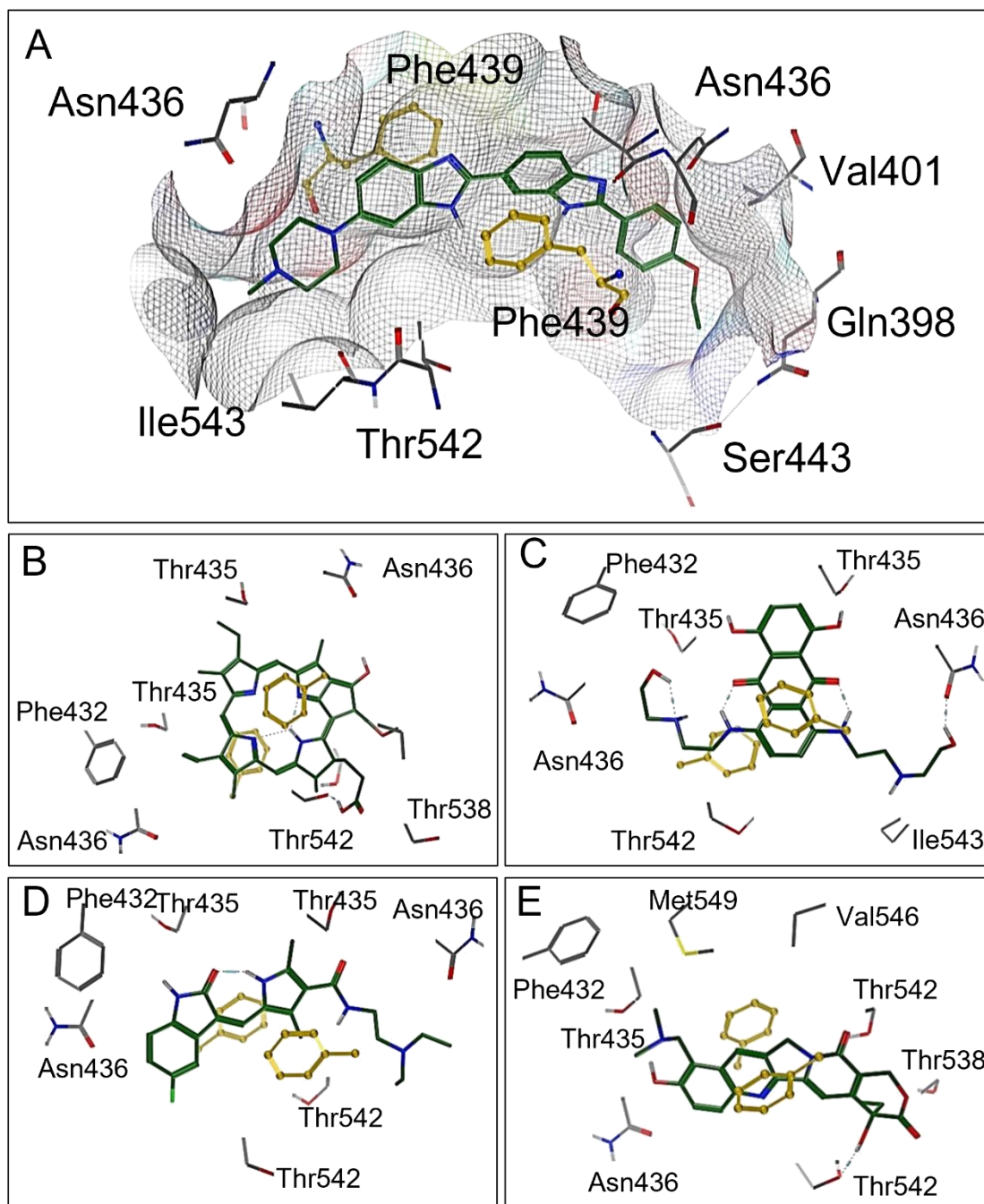


Figure 6.7. Diagrammatic representation of structurally different ligands in the ABCG2 binding cavity and a depiction of the F439 interaction. (A) Hoechst 33342. (B) Pheophorbide A. (C) Mitoxantrone. (D) Topotecan. (E) Sunitinib. F439 (yellow) provides an anchoring point for ABCG2 ligands (A-E).

## 6.6 CeTSA analysis and the ABCG2 'clamp' model

The aforementioned *in-silico* studies have proposed a key role of Phe439 in the ligand-protein interaction of ABCG2 substrates. To validate these results, F439 in ABCG2 expressed in MEK cells was mutated to alanine, tyrosine, and tryptophan. These mutant orthologues were used in a cellular thermal shift assay (CeTSA) to study the ligand-protein interactions and drug transport assays (see paper 2 in the Appendix: Gose *et al*, ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket.). In brief, the microsomes expressing the WT ABCG2, and mutant ABCG2 were heated in the presence and absence of drugs and total recovery of protein was measured using western blotting. The recovery of the protein was measured in comparison to the untreated sample. Interestingly, F439A mutants showed significantly reduced recovery of protein indicating loss of stability and assumed ligand-protein interaction. F439A also showed loss of substrate translocation in pheophorbide accumulations and Hoechst 33344 transport assays (Figure 3, Gose *et al*, 2020). Furthermore, when F439 was mutated to either tyrosine or tryptophan an aromatic residue, no loss of function was observed in ABCG2 (see figure 4, Gose *et al*, 2020).

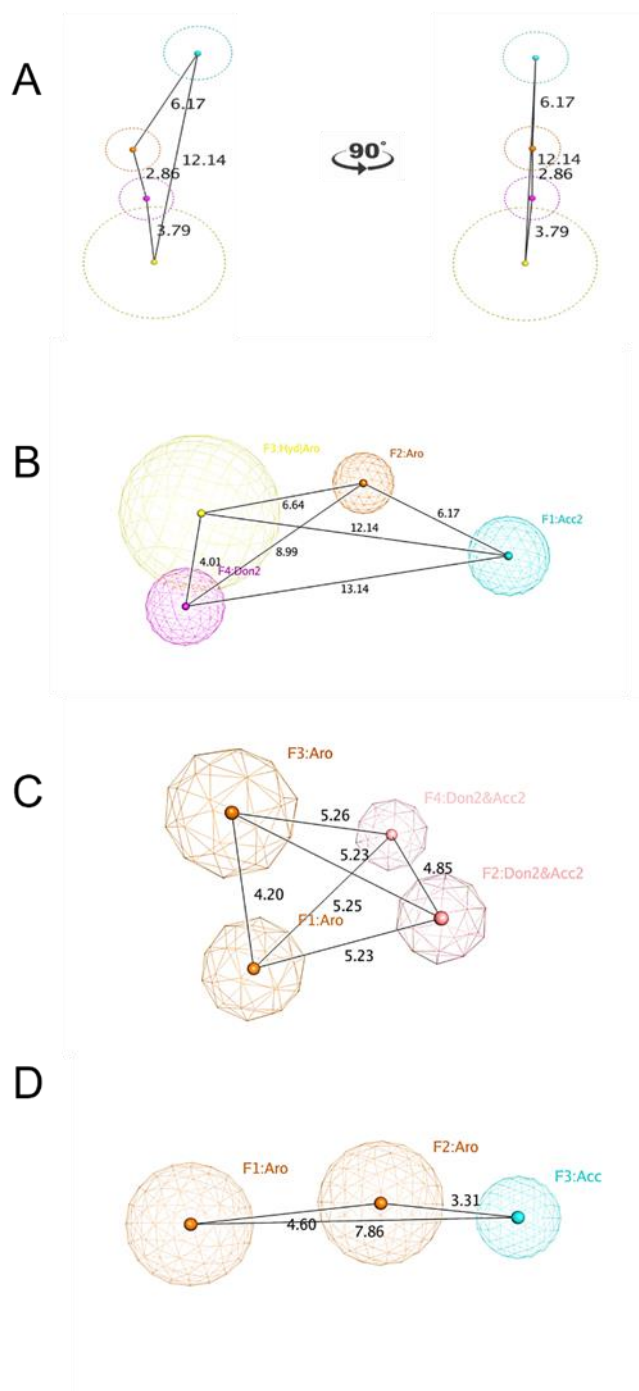
## 6.7 Pharmacophore modelling

Coordinates of two of the ABCG2 substrates imatinib (which was also under investigation in aforementioned docking and cellular thermal shift assays) and topotecan analogue (SN38) were extracted from cryo-EM structures and were used to generate 4 pharmacophore models (2 from imatinib and 2 from topotecan analogue). Table 6.1 shows features present in the pharmacophore models.

*Table 6.1 Features present in pharmacophore models*

		<b>Aromatic</b>	<b>Hydrophobic</b>	<b>Acceptor</b>	<b>Donor</b>
Imatinib	Model 1	1	1	1	1
	Model 2	1	1	1	1
Topoteka	Model 1	2		1	
	Model 2	2		1	1

Interestingly, all models showed common features with 2 aromatic/hydrophobic features and one hydrogen bond acceptor. Imatinib also showed the presence of a hydrogen bond donor feature along with aromatic and acceptor features. All models showed two aromatic/hydrophobic features which corresponded to aromatic moieties of Phe439 in ABCG2. The 3D pharmacophore models are represented as maps of above mentioned chemical features as spheres (see Figure 6.8). The distance between these spheres represent the shape of the molecules as well as the ligand-protein interaction of the molecules inside the binding cavity (Yang 2010). Interestingly, except from topotecan model 1, where aromatic moieties are in 3D conformation (see Appendix Figure 6.3 in Appendix), all other models demonstrated 2D planar conformation (see Figure 6.8A). This 2D planar orientation of drugs was also observed in the above-mentioned docking studies. This 2D orientation of molecules was also consistent with drug conformations observed in structural studies (Jackson, Manolaridis et al. 2018, Kowal, Ni et al. 2021, Yu, Ni et al. 2021). Furthermore, all models showed an average  $6.36 \pm 1.5$  Å distance between the aromatic and hydrogen acceptor moiety.



**Figure 6.8** Spatial orientation of pharmacophore models. (A) Represents imatinib model 1 consisting of two aromatic features and two Donor/acceptor features. the models shows a 2D planar conformations (B) Three-dimensional orientation of imatinib model 2 consisting of two aromatic features and hydrophobic features and two donor/acceptor feature. (C) Correspond to topotecan model 1 consisting of 2 aromatic and 2 acceptor/donor features and (D) Spatial representation of topotecan model 2.

### 6.7.1 Model evaluations

Models were evaluated for their predictive efficacy using a training data set as reported by Hazai *et. al.*, (2013). The Spatial Data Files (SDF) were downloaded from supplementary materials in (Hazai, Hazai et al. 2013) and imported into a MOE database. The 3D structures were energy minimized and stochastically searched for 3D conformational space which was subsequently used in pharmacophore mapping and hit predictions. A final training data set consisted of 148 compounds out of which 98 were ABCG2 substrates and 50 were non-substrates. All models were effectively able to classify substrates and non-substrates. However, Topotecan model 2 showed the highest hit rate with 95 drugs identified as positive (see table 6.2). Moreover, the model also showed a sensitivity of 76.53% and precision of 79%, suggesting the models' accuracy in predicting true positives. The closest model to topotecan model 2 was imatinib model 2 which showed a hit rate of 61.5% and precision of 80%. In comparison, both model 1 from topotecan and imatinib demonstrated better efficiency in identifying true positive and true negative. However, both model 1 showed a catch 22 paradox of low hit rate. Table 6.2 shows the predictive abilities of the topotecan and imatinib models.

Table 6.2 Prediction efficiency of pharmacophore models against the known dataset.

		Hits &	True positive *	False- positive **	True negative #	False- negative ##
Topoteca	Model 1	52 (35%)	42 (81%)	10 (19%)	40	56
	Model 2	95 (64.1)	75 (79%)	20 (21%)	30	23
Imatinib	Model 1	49 (33%)	45 (92%)	4 (8%)	46	53
	Model 2	91 (61.5%)	73 (80%)	18 (20%)	32	25

		Sensitivity %	Specificity %	Precision %	Accuracy %
Topoteca	Model 1	42.86	80	80.77	55.4
	Model 2	76.53	60	78.95	70.95
Imatinib	Model 1	45.92	92	91.84	61.49
	Model 2	74.49	64	80.22	70.95

&Hits = number of compounds predicted substrates from the total dataset

\*True-Positive = TP = compounds that were positive and predicted as positive

\*\*False-Positive = FP = compounds that were negative but predicted positive

#True\_Negative =TN= compounds which were negative and predicted negative

##False-Negative +FN= compounds which were positive but predicted negative

*Sensitivity* =  $TP/(TP + FN)$  = true positive rate

*Specificity* =  $TN/(TN + FP)$  = true negative rate

*Precision* =  $TP/(TP + FP)$  = positive prediction value

*Accuracy* =  $(TP + TN)/(TP + FP + TN + FN)$  = closeness of prediction with true values

## 6.8 Virtual screening

*In-silico* screening study was performed on CCDC GOLDSuite version 2021.1.0 (Verdonk, Cole et al. 2003) as explained in section 2.3.4.2.4. Initial optimization of the cut-off threshold was done on a known training data set of substrate and non-substrate (Shukla, Kouanda et al. 2014). A total of 100 conformations for each molecule were generated and binding scores were calculated. The top ten scores of each ligand were exported and analysed for threshold cut-offs. Figure 6.9 shows that ChemPLP performs better in differentiating between the substrates and non-substrates. ChemPLP showed a mean of 81.2 for non-substrate and 94.7 for the substrates. Moreover, percentile analysis showed that 75% of non-substrate scores were below 85.4 scoring value with an upper 95% CI of 82.94. Whereas, in the GOLD score the mean score of substrate vs non-substrate was 49.6 and 42.9 and median of 52.5 and 49.82 respectively suggesting an inability of GOLDScore to distinguish between substrate from non-substrate.

The final protocol used ChemPLP as a scoring function with a threshold of 82.5. The compounds with scoring values higher than the threshold value were considered the probable hits. The threshold was intentionally kept two units below at 82.5 from the 75% percentile value to accommodate flexibility in hit predictions. However, the scoring values between 82.5 to 85.5 was considered a grey region. The hit compounds with scoring value in this region were considered potential candidate only if top 10 conformations have scoring values above the threshold value. The protocol was set such that only the top ten conformations above the threshold were kept as hits.

## Optimization fo ABCG2 cut-off threshold

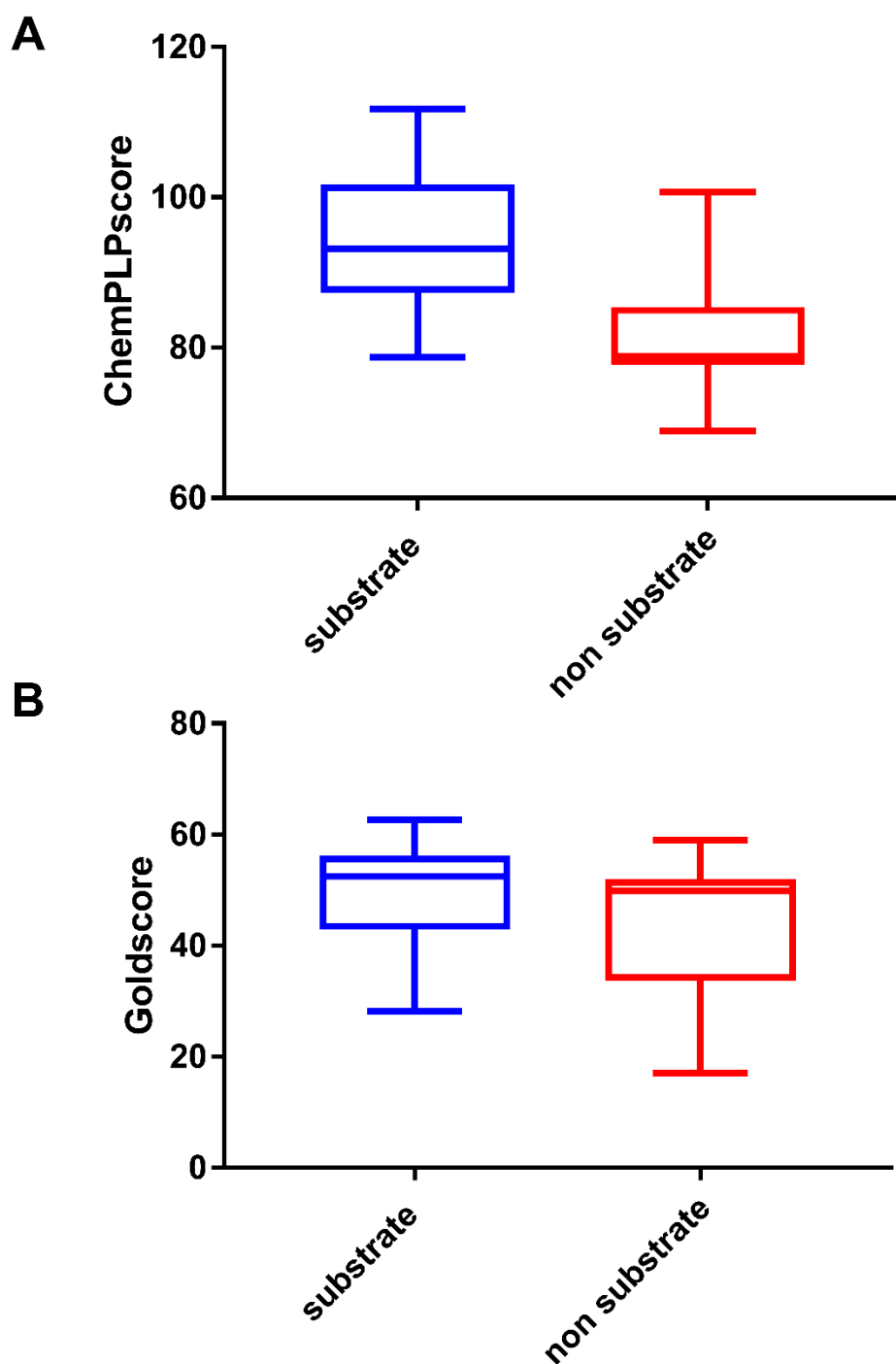


Figure 6.9 Graphical representation of scoring distribution in substrates and non-substrate with ChemPLP (A) and GOLD scoring (B) functions.



FDA approved drug library (L1021) from APExBIO was prepared and screened using the optimized protocol. A total of 278 hits were identified from 1198 compounds. These hits were classified into two classes the kinase class the non-kinase class.

#### 6.8.1 Kinase class

Kinase class of inhibitors were the biggest class identified from the docking based virtual screening experiments. A total of 71 kinase inhibitors out of 121 kinase inhibitors against different targets were identified as candidate hits (see Figure 6.9). Table 6.3 shows the list of all the hits from the kinase class of inhibitors.

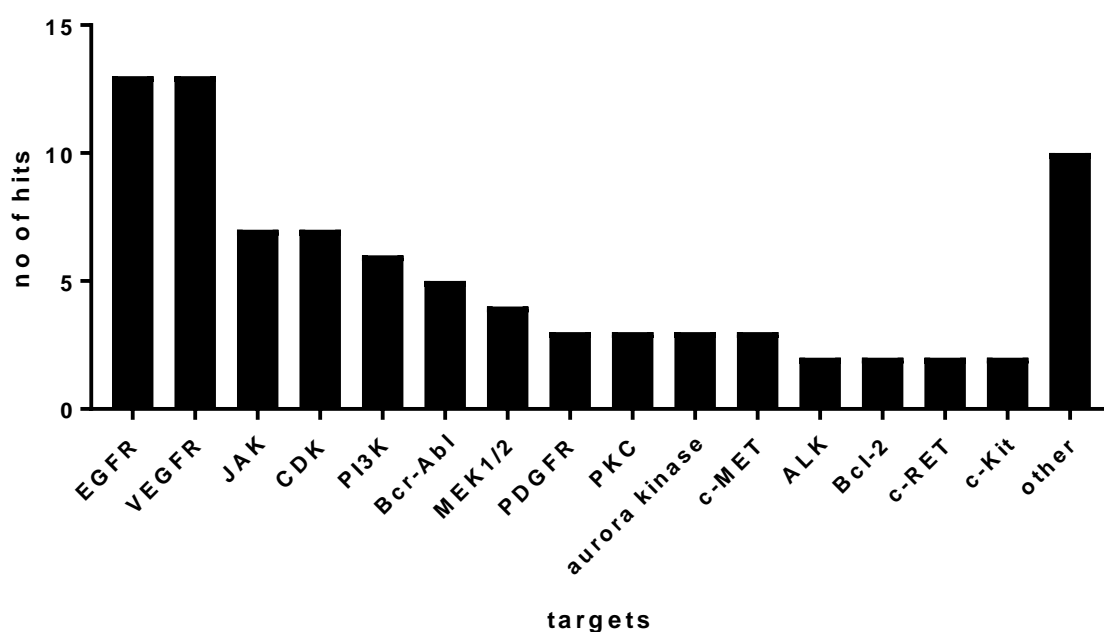


Figure 6.10 Histogram representation of hits from the kinase class according to their targets

Kinase inhibitors targeted against EGFR and VEGFR was the biggest class identified from the virtual screening protocol. EGFR inhibitors were also validated from cellular thermal shift assay. (Gose, T. Shafi, T *et al*). ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket. 2020. FASEB J). lapatinib showed the highest interaction during the virtual screening with scores of 103. It also showed the highest inhibitory activity in pheophorbide A transport in mABCG2 knockout hABCG2 expressing MEK cells. However, a relation between IC<sub>50</sub> and ChemPLP scores was not observed (data not shown).

VEGFR inhibitors are the second largest class identified from the screening. Interestingly most of the VEGFR inhibitors showed value in the grey area as defined during the protocol. Axitinib a VEGFR inhibitor showed a clear non-substrate during the screening process. Our data also showed that when microsomes containing hABCG2 were heated in the presence of axitinib no protein was recovered suggesting axitinib is a non-substrate for ABCG2 (Gose, Shafi et al. 2020). Further, when sunitinib was tested it showed transport inhibition in the sub-micromolar range. Our screening results also showed sunitinib as a hit.

#### 6.8.2 Non-kinase class

187 hits were identified out of 1077 compounds from the non-kinase class. Figure 6.10 shows hits according to the class of those drugs. Anti-infection drugs (i.e., antibiotics, antivirals, and antifungals) were the biggest hits in the non-kinase class. 35 drug hits were from these anti-infections class.

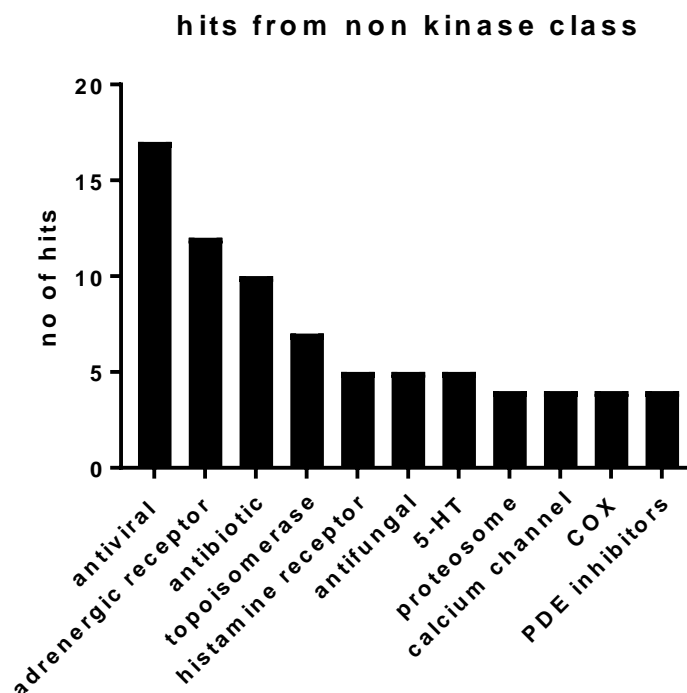


Figure 6.11. Histogram representation of major drug hits from the non-kinase class of inhibitors.

Antivirals are a promising class of drugs tested for ABCG2 modulations. Several studies have reported the interaction of HIV direct-acting antivirals (Wang, Furukawa et al. 2003, Weiss, Rose et al. 2007). However, HCV antivirals are not well studied for ligand-protein interaction with ABCG2. Another interesting class identified was CFTR correcting drugs lumacaftor and tezacaftor. Previously our lab reported the interaction of ivacaftor a CFTR corrector with P-glycoprotein (Lingam, Thonghin et al. 2017). Woodward et al have reported the use of molecular correctors for stabilizing the Q141K mutants of ABCG2 (Woodward, Tukaye et al. 2013). However, no evidence on the interaction of molecular correctors with wild type ABCG2 is available. (see Supp Table 6.1 for complete list of hits)

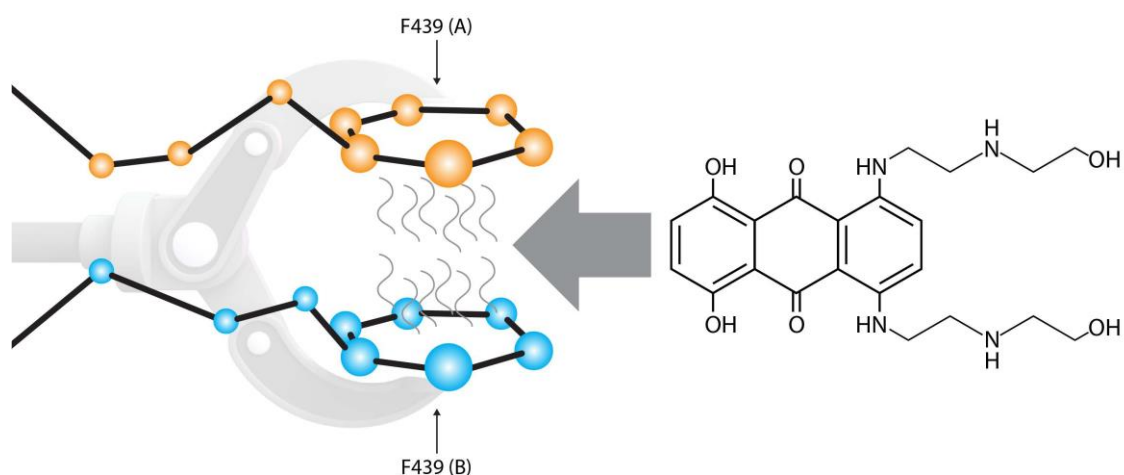
## 6.9 Discussion and conclusions

Multidrug resistance is a burning clinical challenge and ABCG2 is one of the prime culprits. In recent times, several efforts have been made to understand the underlying molecular factors and identify modulators to regulate ABCG2 mediated drug resistance. Recent snapshots of ABCG2 in substrate and drug bound states have delineated a large central binding cavity where all substrates and inhibitors were reported to be bound in the vicinity of the same residues. However, these structures were not able to elucidate the importance of these residues in ligand-protein interactions and drug translocations pathways. Therefore, during this study, we employed in-silico approaches along with thermal shift and drug transport assays to elucidate keys residues involved in the translocation pathway and to identify new classes of candidate drugs.

Our result demonstrated that a single evolutionary conserved residue Phe439 plays a critical role in ligand-protein interaction. Even docking of structurally diverse compounds showed that all compounds are sandwiched between the two phenylalanine residues contributed by each monomer. These observations were consistent with recently published ABCG2 structures bound with topotecan, imatinib, and mitoxantrone (Orlando and Liao 2020) where drugs were anchored between the two Phe439 residues. When this residue was mutated to a non-aromatic residue, alanine the protein showed complete loss of function suggesting its critical role in substrate interaction and drug translocation. This observation was consistent with (Manolaridis, Jackson et al. 2018) where F439A showed loss of estrone-3-sulfate transport. Importantly, this loss of function was not altered when this phenylalanine was replaced with tyrosine which is also found in ABCG5 (see Figures 6.4 and 6.5). Moreover, when Phe439 was

replaced with tryptophan (bulky aromatic residue), a slight gain of function was observed in ABCG2.

These results suggested a critical role of the aromatic residue Phe439 in ligand-protein interaction and the substrate translocation pathway in ABCG2. Phylogenetic analysis of ABCG2 analogues across species, and across the human ABCG family, also supports these results where this residue remained conserved over the course of evolution. The identification of key phylogenetically conserved residue has led to the proposal of single amino acid 'clamp' model of ligand-protein interaction (Figure 6.12), where structurally diverse compounds anchor at the pivot residue and re-arrange themselves in the binding cavity.



*Figure 6.12 Graphical representation of 'clamp' model. "A" and "B" indicate F439 has been donated from separate ABCG2 monomers. The molecule that will interdigitate between the two F439 is the ABCG2 substrate and chemotherapeutic agent, mitoxantrone*

This clamp model was also consistent with our pharmacophore models where aromatic features were common in all models. When these models were placed in the ABCG2 binding cavity these features corresponded to the Phe439. Whereas the acceptor and donor features represented Asn436 whereas in some cases represented by Thr435 and Thr542. These features are similar to the model reported by (Le, Hoang et al. 2021). Except they reported an additional hydrophobic feature.

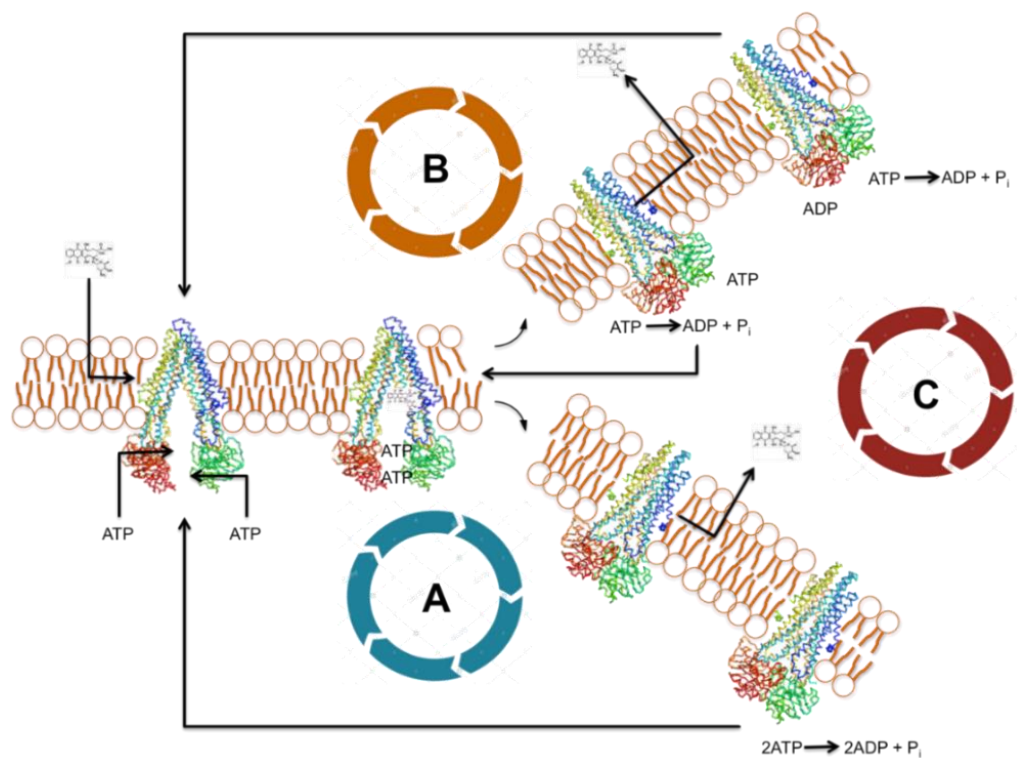
Our model only proposes the role of phe439 in the drug translocation pathway it does not exclude the possibilities of alternative substrate/drug binding sites. The earliest residue reported for drug interaction Arg482 at TM3 has been extensively studied for its role in drug translocation (Ejendal, Diop et al. 2006, Ni, Bikadi et al. 2011, Haider, Cox et al. 2015). Recent structures of ABCG2 in outwards and inward states also suggest two drug binding cavities separated by leucine residues (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018, Orlando and Liao 2020). Furthermore, Biochemical and mutational analysis on mitoxantrone binding has also supported the notion of multiple binding sites (Cox, Kapoor et al. 2018). However, the cryo-EM structure of ABCG2 bound with mitoxantrone showed binding of mitoxantrone at only one site near Phe439 which suggesting a high-affinity state. Our study also does not exclude the possibility of poly-specificity but only suggests that binding of the drug at Phe439 (probably a high-affinity site) is crucial for the drug translocation pathway for several structurally unrelated compounds.

## Chapter 7 General Conclusions and Future Prospects

Integral membrane proteins are a large family of proteins that demonstrate extraordinary diversity in structure and functions. Due to this structural and functional diversity, they are clinically very important and are the target for about 50% of the drugs (Terstappen and Reggiani 2001, Arinaminpathy, Khurana et al. 2009). Therefore, structural studies on membrane proteins can be vital. However, the expression and purification of membrane protein is a challenging task and needs optimizations (Tzvi Tzfira 2005, Wagner, Bader et al. 2006). Here in this study, we have expressed mouse and human analogues of P-glycoprotein in yeast expression systems. Yeast has been used for the expression of membrane proteins in the past (Vieira Gomes, Souza Carmo et al. 2018). It can show a controlled regulation of post-translational modifications on membrane proteins. Glycosylation is an important post-translation modification for the trafficking of P-glycoprotein to the membrane surfaces (Schinkel, Kemp et al. 1993). Both *Pichia pastoris* and *Saccharomyces cerevisiae* were able to traffic the P-glycoprotein to the membrane surfaces. Furthermore, F-PNGase also suggests yeast can process the glycosylation on P-glycoprotein. This is consistent with the finding of (Cant 2013) where she demonstrated glycosylation on CFTR analogues in yeast. She also demonstrated phosphorylation of CFTR in yeast. Therefore, we can imply that yeast can be used as an expression system for membrane proteins with post-translational modifications.

P-glycoprotein is an energy driven transporter that regulates the movement of substrate across the membrane. Several models of drug translocation pathways have been proposed in the past. One of the initial models was the Alternative site model which proposed that two ATP molecules bind to the NBDs and in the sequential events of ATP hydrolysis the transporter goes to the outward-facing conformation where the substrate is effluxed and with another event of second

ATP hydrolysis it restores the transporter in the initial state (Senior, al-Shawi et al. 1995). However, recent structural and biochemical studies have shown that the binding of ATP is sufficient to switch between the inward and outward-facing conformations (Kim and Chen 2018). This led to a new translocation model referred to as the ATP switch model that says binding of ATP is sufficient to induce the conformational changes that lead to an outward-facing state. Furthermore, several studies have demonstrated a basal state hydrolysis activity in P-glycoprotein. This led to the modification ATP switch model to a constant contact model where the two NBDs are in a constant basal hydrolysis state. Figure 7.1 shows the diagrammatic representation of P-glycoprotein translocation models.

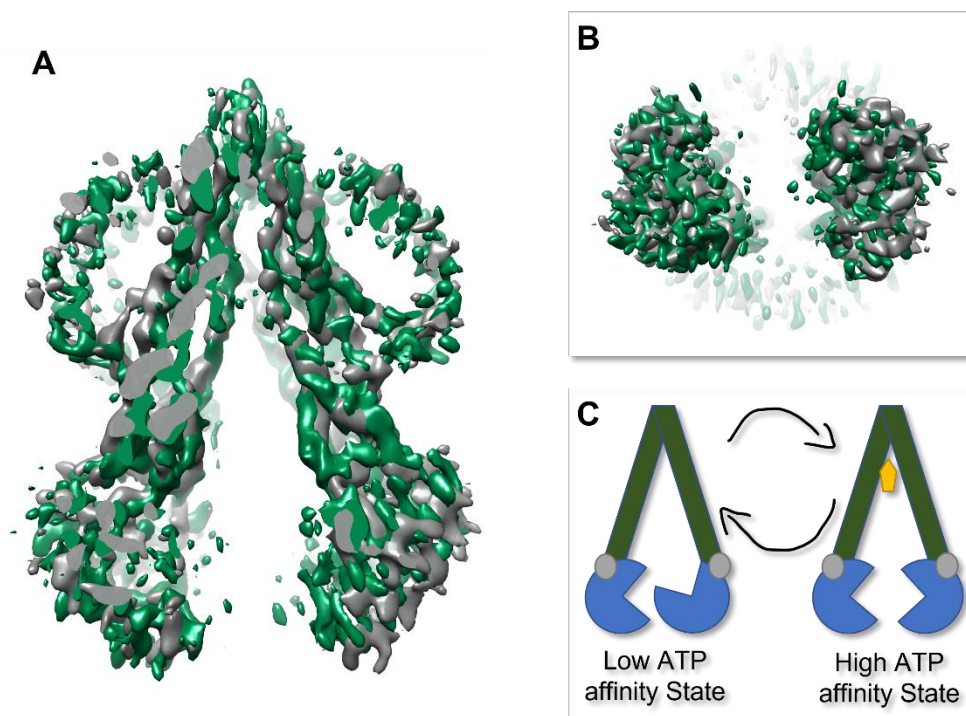


*Figure 7.1. Diagrammatic representations of putative ABC substrate translocations models. A) Represents the ATP switch model where binding of ATP to the NBDs provides the energy to go into the outward open configuration. Hydrolysis of ATP reinstates the P-glycoprotein into its basal state. B) Represents the alternating site model where first ATP hydrolyse to provide power stroke that leads to open conformation. Hydrolysis of second ATP restores P-glycoprotein into closed conformation. C) In constant contact model the two NBDs remain in close conformation and repeat the efflux cycle.*



This continuous hydrolysis of ATP is energetically not favourable for cells and there must a regulatory mechanism which regulates this ATP hydrolysis. One putative mechanism can be explained on the basis of TMD rearrangement after the substrate-binding (Szewczyk, Tao et al. 2015). Our results of ivacaftor binding with mouse P-glycoprotein also sheds light on similar kind of drug-induced structural rearrangements between the TMDs and NBDs. 3D class, bound with the ivacaftor density shows flex of the NBDs as compared to the classes which represent the putative apo state (see Figure 7.2). This flex like structural re-arrangement was not observed in the cryo-EM maps with ADP and Vanadate (Thonghin, Collins et al. 2018). This suggests a regulatory role of this flex like re-arrangement of NBD upon drug binding. Which can be one of the possible mechanisms of ATP hydrolysis regulation. This hypothesis is also supported by the biochemical evidence of long-range conformational coupling between TMDs and NBDs that come from several studies where drug binding can stimulate ATP hydrolysis (Scarborough 1995, Lingam, Thonghin et al. 2017).

Further, the cryo-EM structure has shown binding of ivacaftor near the entry portal. This binding site is consistent with the residues identified in cross-linking studies to study the ligand-protein interaction of verapamil and rhodamine. This site also overlaps with Ile336, Phe339 and Gln343 on the TMH6 identified from crystallographic structure mouse P-glycoprotein bound with peptide-mimetic compound QZ-Phe (PDB ID: 4Q9L) (Szewczyk, Tao et al. 2015) and zosuquidar (PDB ID: 6FN1) (Alam, Küng et al. 2018). However, both QZ-Phe and zosuquidar binds higher in the binding cavity away from the entry portals.



*Figure 7.2 Diagrammatic representation of NBD re-arrangement in drug binding state. (A) Anterior view of P-glycoprotein in drug bound state (grey) and apo state (green). The NBD2 in drug bound state showed extended length. (B) NBDs re-arrangement from the bottom. (C) Cartoon representation of proposed ATP regulation method. The binding of the drug induces conformational changes which increase the affinity of NBDs for ATP.*

The catalytic cycle of P-glycoprotein shows diverse catalytic states. EPR studies on mouse P-glycoprotein have shown that the inward-facing conformations are more dominant in the majority of the catalytic states (Verhalen, Dastvan et al. 2017). However, when the protein was trapped using the ATP and vanadate ion, an outward conformation was observed. This was in contrast with our findings from cryo-EM studies, where we trapped the protein with ATP and vanadate (Thonghin, Collins et al. 2018). The single-particle analysis showed that the protein was predominantly in an inward-facing conformation. Further, in the recent structure of P-glycoprotein in the outward occluded state, the protein was forced to fix in the outward occluded state by mutating the residues in the NBDs (Kim and Chen 2018). Even with the forced constraints, the single-particle analysis showed only 15% particles in the outward conformations remaining particles were either in the inward states or showed the transition

between the inward and outward states. All this evidence imply that the outward state is highly unstable and tend to go in an outward state transiently. Capturing a transient state in cryo-EM is not trivial (Lyumkis 2019). Small-angle X-ray scattering is a low-resolution technique that can not only tell about the size, oligomeric states but also can elucidate the structural properties. This information can be used to generate the *ab-initio* envelope models that can provide the complete landscape of molecular dynamics in an aqueous environment. Our preliminary analysis has shown the robustness of SAXS to identify different states. However, our results showed aggregation as a confounding factor in scattering profiles. Even dilutions to a factor of 10 fold did not help reduce the aggregation effect in the final scattering profiles. However, it was observed that the presence of P-glycoprotein substrates ATP and or drugs stabilize the P-glycoprotein. This was consistent with the findings of Lusvarghi et al where ATP binding thermostabalize P-glycoprotein (Lusvarghi and Ambudkar 2019). Analysis of SAXS samples with negative stain electron microscopy showed structural parameter comparable to the values obtained from scattering profiles. Therefore negative stain electron microscopy can be a good initial sample check for SAXS.

The work on ABCG2 was done in collaboration with John Scheutz lab. The project aimed to explore the ligand-protein interaction of drugs with ABCG2 using the in-silico approaches. The predictions were validated with wet-lab studies performed by Tomoka Gose. Initially, it was observed that Lapatanib showed a strong affinity towards ABCG2. While its analogues show different affinities towards ABCG2. So, it was interesting to understand the ligand-protein at molecular levels. Therefore, we proposed to analyze the ligand-protein interaction via molecular docking. In recent times, molecular docking approaches have advanced leaps and bounds for the study of ligand behaviour in the protein binding cavity (Fu, Zhao et al. 2018).

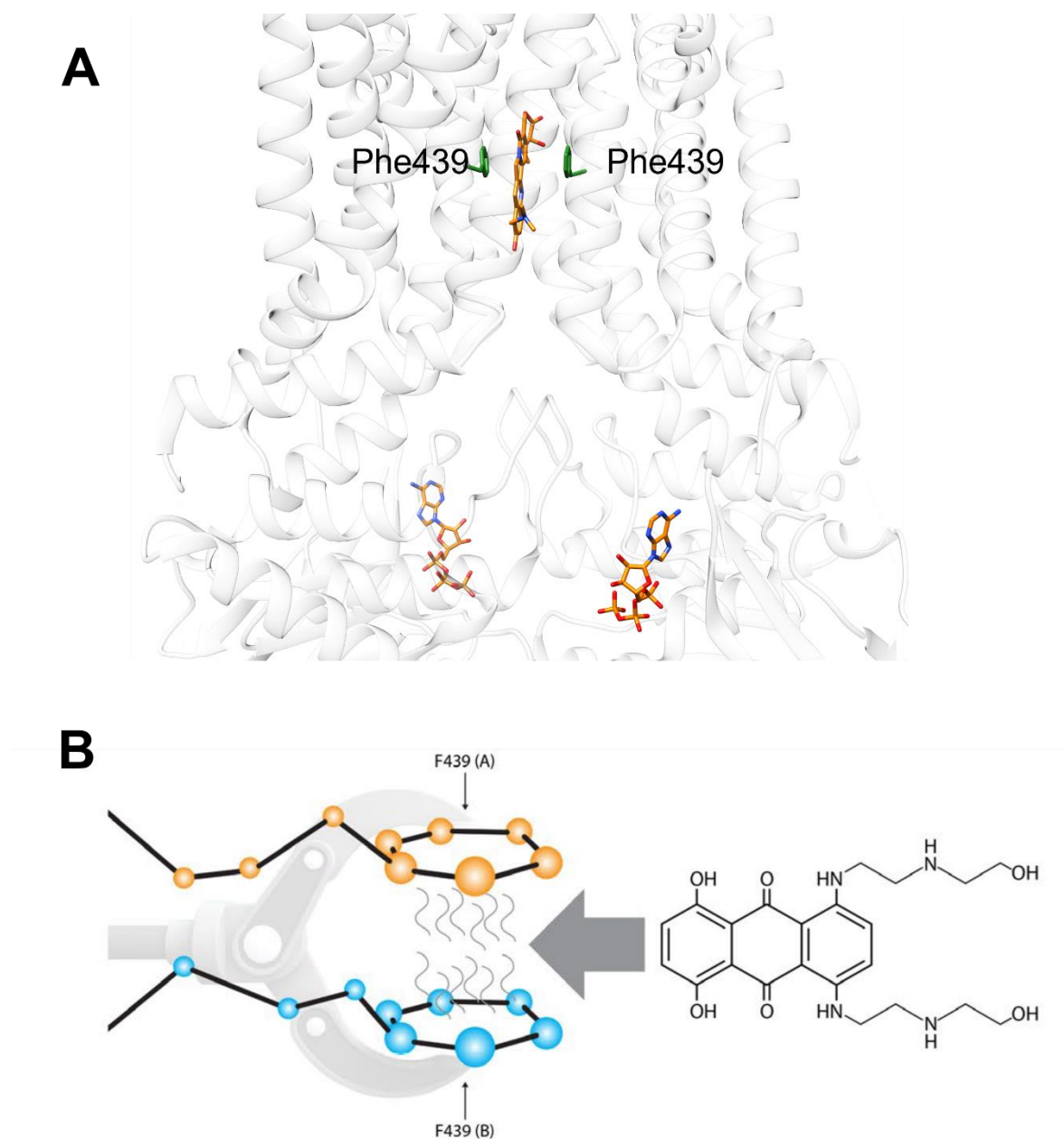
Several examples of successfully using molecular docking to understand the ligand-protein interactions and structure-based drug design have been reported (Boehm, Boehringer et al. 2000, Kirton, Murray et al. 2005). In general, the docking process involves two practical steps, first is conformational sampling where it explores the different binding sites for probable conformations and interactions, the second step is scoring the conformation according to the interactions (Saikia and Bordoloi 2019). Ideally, the conformational sampling should produce a conformation similar to the conformations that can be obtained from structural or biochemical studies. Moreover, scoring functions should be able to rank these predicted conformations according to the probability of obtaining them in the experimental studies. However, in practice, this needs to be optimized. Therefore, we tested several combinations of placement and scoring functions that were implemented in the software package MOE to optimize the predictions based on the known conformations of drugs from structural studies.

The ABCG2 structure bound with a Ko143 analogue (PDBID: 6ETI) was used for the docking study because (i) it represented the drug bound state, (ii) it demonstrated the highest resolution in available structures. The final combination of the placement method and scoring functions showed a similar interaction of Ko143 with ABCG2 as reported in structural models reported by Locher's lab (PDBID: 6ETI) (Jackson, Manolaridis et al. 2018). When lapatinib derivatives were docked the top-ranked conformations and the algorithmic hierarchical clustering conformations showed the same binding cavity between the two Phe439 residues. Interestingly, 90% percent of the conformations showed Phe439 present within the 4Å vicinity. This site is also consistent with the residues identified from ABCG2 structures. All structures of ABCG2 in the PDB database with a drug bound showed interaction of drug with Phe439. Furthermore, this Phe439 was conserved throughout the evolution of ABCG2 from amphibians to humans. This led to a

question that whether this Phenylalanine contributes to the overall lipophilicity of the binding cavity or  $\pi$ - $\pi$  interactions are crucial for ABCG2 interactions. Therefore, when Phe439 was mutated to alanine (a lipophilic residue) and tyrosine: an aromatic residue (can contribute the  $\pi$ - $\pi$  interaction) in *in-silico* docking as well as wet-lab experiments. the alanine residues showed a decrease in the affinity for drug binding. However, when mutated to tyrosine, no change in the affinity was observed as compared to wild type ABCG2. Further mutation of Phe439 to alanine diminished the transport ability of ABCG2 for Hoechst33342. This was also observed by Manolaridis et al for estrone 3-sulphate (Manolaridis, Jackson et al. 2018). All this evidence, imply the critical role of  $\pi$ - $\pi$  interaction in drug binding and drug translocation. The contribution of aromatic moiety was also reflected from the pharmacophore models based on ABCG2 ligands.

Here in this study proposed the single residue 'Clamp' like model of ABCG2 function (see Figure 7.3). Where ABCG2 substrate gets clamped between the two Phe439 residues contributed from each monomer of ABCG2. This clamp model can rationalize that ABCG2 substrates are generally flat and polycyclic (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018). Our model does not exclude the possibility of multiple binding cavities. It only proposes that Phe439 is required for structurally different compounds to optimally interact with ABCG2. This argument can be supported by the fact that all the structures of ABCG2 with drug bound to them showed drug clamped between the two residues (Jackson, Manolaridis et al. 2018, Manolaridis, Jackson et al. 2018, Yu, Ni et al. 2021). Furthermore, structural and biochemical data implies that the binding of substrate to Phe439 can be the first step of the drug translocation pathway (Gose, Shafi et al. 2020, Yu, Ni et al. 2021). Subsequently, the hydrolysis

of ATP leads to the opening of 'leucine plug' and the drug is translocated to the upper cavity from where it induces the opening of extracellular roof (Khunweeraphong, Szöllősi et al. 2019).



*Figure 7.3 Diagrammatic representation of single residue clamp models. (A) Snapshot of topotecan trapped between the two Phe439 residues (PDBID: 7OJI). (B) Cartoonic representation of the ABCG2 clamp model where the two Phe439 residues from each chain (represented as A and B here) provide the anchoring point for the drug (mitoxantrone)*

## Future Prospect

Removal of micelles and detergent was the limiting factor in good quality grid preparation. Cut-off filters are usually good at removing detergents during buffer replacements, however, we observed a concentration of micelles during cryo-EM. Presumably, the presence of a highly lipophilic drug, like ivacaftor altered the dynamics of the detergent micelle, therefore, concentrating the micelles in the cryo-EM samples. One approach that was under consideration was the use of size exclusion chromatography (SEC) to remove micelles. Theoretically, micelles have different elution rates as compared to the protein, therefore, separating the protein from the empty micelles. Miss Iqra Younus is currently working on optimizing the sample preparation via SEC. The use of glycerol gradient centrifugation can be an alternative method that has been reported to remove the detergent micelles from the cryo-EM samples, especially for detergents like DDM which have very low CMCs (Hauer, Gerle et al. 2015).

In addition, to improve the quality of cryo-EM data and electron density maps for mouse P-glycoprotein with molar excess of ivacaftor, capturing the catalytic state with both drug and ATP bound can be vital in understanding the translocation pathway. Previously, Dr Swathi Lingham has shown that ivacaftor can stimulate human P-glycoprotein twofold than the gold standard substrate verapamil (Lingam, Thonghin et al. 2017). Here in this study, we have demonstrated that ivacaftor can also interact with mouse P-glycoprotein. Theoretically, if we have strong hydrolysis simulating substrate in addition to the ATP molecules the probability of having outward-facing conformations increases. Furthermore, time-resolved cryo-EM analysis can be another avenue that can be explored to study the translocation pathway in high resolution and to capture protein in outward-facing conformations.

Small-angle scattering results were dominated by aggregation and needs improvements for generating the ab-initio models. One of the proposed ways was to use the SEC-SAXS to separate the aggregates from the sample. Unfortunately due to covid lockdown and post covid limited access to the synchrotron facility could not proceed with the data collections. However, Miss Iqra Younus is working on generating the SEC-SAXS profiles of P-glycoprotein as well as other ABC transporters. Further, It will also be an interesting analysis to see the effect of a serially concentrating sample on protein aggregation in scattering profiles which is the opposite of serial dilutions we used in our study such kind of experiment can help in understanding the behaviour of membrane proteins in detergents.

Structure-based virtual screening of the FDA drug library has proposed some interesting classes of the drug as potential candidates for both P-glycoprotein and ABCG2. Validation of those drugs will be an interesting project that can not only validate the reliability of our model but also help in finding novel drugs from large databases.



## References

- Alam, A., J. Kowal, E. Broude, I. Roninson and K. P. Locher (2019). "Structural insight into substrate and inhibitor discrimination by human P-glycoprotein." *Science* 363(6428): 753-756.
- Alam, A., R. Küng, J. Kowal, R. A. McLeod, N. Tremp, E. V. Broude, I. B. Roninson, H. Stahlberg and K. P. Locher (2018). "Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1." *Proc Natl Acad Sci U S A* 115(9): E1973-e1982.
- Arakawa, T., Y. Kita and S. N. Timasheff (2007). "Protein precipitation and denaturation by dimethyl sulfoxide." *Biophys Chem* 131(1-3): 62-70.
- Arinaminpathy, Y., E. Khurana, D. M. Engelman and M. B. Gerstein (2009). "Computational analysis of membrane proteins: the largest class of drug targets." *Drug Discov Today* 14(23-24): 1130-1135.
- Austin Doyle, L. and D. D. Ross (2003). "Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2)." *Oncogene* 22(47): 7340-7358.
- Azzaria, M., E. Schurr and P. Gros (1989). "Discrete mutations introduced in the predicted nucleotide-binding sites of the *mdr1* gene abolish its ability to confer multidrug resistance." *Mol Cell Biol* 9(12): 5289-5297.
- Bai, J., D. J. Swartz, I. I. Protasevich, C. G. Brouillette, P. M. Harrell, E. Hildebrandt, B. Gasser, D. Mattanovich, A. Ward, G. Chang and I. L. Urbatsch (2011). "A Gene Optimization Strategy that Enhances Production of Fully Functional P-Glycoprotein in *Pichia pastoris*." *PLOS ONE* 6(8): e22577.
- Barbieri, A., N. Thonghin, T. Shafi, S. M. Prince, R. F. Collins and R. C. Ford (2021). "Structure of ABCB1/P-glycoprotein bound to the CFTR potentiator ivacaftor." *bioRxiv*: 2021.2006.2011.448073.
- Barbieri, A., N. Thonghin, T. Shafi, S. M. Prince, R. F. Collins and R. C. Ford (2021). "Structure of ABCB1/P-Glycoprotein in the Presence of the CFTR Potentiator Ivacaftor." *Membranes* 11(12): 923.

- Bell, J. M., M. Chen, T. Durmaz, A. C. Fluty and S. J. Ludtke (2018). "New software tools in EMAN2 inspired by EMDatabank map challenge." *Journal of structural biology* 204(2): 283-290.
- Benjin, X. and L. Ling (2020). "Developments, applications, and prospects of cryo-electron microscopy." *Protein science : a publication of the Protein Society* 29(4): 872-882.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." *Nucleic Acids Research* 28(1): 235-242.
- Bernadó, P. and D. I. Svergun (2012). "Analysis of intrinsically disordered proteins by small-angle X-ray scattering." *Methods Mol Biol* 896: 107-122.
- Bibi, E., P. Gros and H. R. Kaback (1993). "Functional expression of mouse *mdr1* in *Escherichia coli*." *Proceedings of the National Academy of Sciences of the United States of America* 90(19): 9209-9213.
- Blanchet, C. E. and D. I. Svergun (2013). "Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution." *Annual Review of Physical Chemistry* 64(1): 37-54.
- Boehm, H. J., M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbbers, N. Meunier-Keller and F. Mueller (2000). "Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening." *J Med Chem* 43(14): 2664-2674.
- Boldon, L., F. Laliberte and L. Liu (2015). "Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application." *Nano Reviews* 6: 10.3402/nano.v3406.25661.
- Booth, D. S., A. Avila-Sakar and Y. Cheng (2011). "Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition." *Journal of visualized experiments : JoVE*(58): 3227.

- Brennich, M. E., A. R. Round and S. Hutin (2017). "Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline." *JoVE*(119): e54861.
- Brilot, A. F., J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, R. Henderson and N. Grigorieff (2012). "Beam-induced motion of vitrified specimen on holey carbon film." *J Struct Biol* 177(3): 630-637.
- Brodsky, J. L. and W. R. Skach (2011). "Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems." *Curr Opin Cell Biol* 23(4): 464-475.
- Cant, N. (2013). Expression and Purification of the Cystic Fibrosis Transmembrane Conductance Regulator from *Saccharomyces cerevisiae* for High-Resolution Structural Studies. Ph.D Traditional thesis, The Univeristy of Manchester.
- Chakraborti, S., K. Hatti and N. Srinivasan (2021). "'All That Glitters Is Not Gold': High-Resolution Crystal Structures of Ligand-Protein Complexes Need Not Always Represent Confident Binding Poses." *Int J Mol Sci* 22(13).
- Chang, K. T., J. Guo, A. di Ronza and M. Sardiello (2018). "Aminode: Identification of Evolutionary Constraints in the Human Proteome." *Scientific Reports* 8(1): 1357.
- Chen, P.-c. and J. S. Hub (2015). "Structural Properties of Protein–Detergent Complexes from SAXS and MD Simulations." *The Journal of Physical Chemistry Letters* 6(24): 5116-5121.
- Cheung, M., N. Kajimura, F. Makino, M. Ashihara, T. Miyata, T. Kato, K. Namba and A. J. Blocker (2013). "A method to achieve homogeneous dispersion of large transmembrane complexes within the holes of carbon films for electron cryomicroscopy." *J Struct Biol* 182(1): 51-56.
- Chin, S., M. Hung, A. Won, Y.-S. Wu, S. Ahmadi, D. Yang, S. Elmallah, K. Toutah, C. M. Hamilton and R. N. Young (2018). "Lipophilicity of the cystic fibrosis drug, ivacaftor (VX-770), and its destabilizing effect on the major CF-causing mutation: F508del." *Molecular pharmacology* 94(2): 917-925.

- Choi, C.-H. (2005). "ABC transporters as multidrug resistance mechanisms and the development of chemosensitizers for their reversal." *Cancer Cell International* 5: 30-30.
- Chufan, E. E., K. Kapoor, H. M. Sim, S. Singh, T. T. Talele, S. R. Durell and S. V. Ambudkar (2013). "Multiple transport-active binding sites are available for a single substrate on human P-glycoprotein (ABCB1)." *PLoS One* 8(12): e82463.
- Cole, J., J. Nissink and R. Taylor (2005). *Protein-Ligand Docking and Virtual Screening with GOLD*.
- Consortium, T. U. (2021). "UniProt: the universal protein knowledgebase in 2021." *Nucleic Acids Res* 49(D1): D480-d489.
- Cox, M. H., P. Kapoor, D. A. Briggs and I. D. Kerr (2018). "Residues contributing to drug transport by ABCG2 are localised to multiple drug-binding pockets." *Biochemical Journal* 475(9): 1553-1567.
- Cramer, J., S. Kopp, S. E. Bates, P. Chiba and G. F. Ecker (2007). "Multispecificity of drug transporters: probing inhibitor selectivity for the human drug efflux transporters ABCB1 and ABCG2." *ChemMedChem* 2(12): 1783-1788.
- Crowther, R. A. (2016). *The resolution revolution : recent advances in cryoEM*. Cambridge, MA, Academic Press is an imprint of Elsevier.
- Crowther, R. A., D. J. DeRosier and A. Klug (1970). "The reconstruction of a three-dimensional structure from projections and its application to electron microscopy." *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 317(1530): 319-340.
- Csanády, L. and B. Töröcsik (2019). "Cystic fibrosis drug ivacaftor stimulates CFTR channels at picomolar concentrations." *eLife* 8: e46450.
- Czuba, L. C., K. M. Hillgren and P. W. Swaan (2018). "Post-translational modifications of transporters." *Pharmacology & therapeutics* 192: 88-99.
- Daniel, F. and S. D. I. (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." *Journal of Applied Crystallography* 42(2): 342-346.

- Dassa, E. and P. Bouige (2001). "The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms." *Res Microbiol* 152(3-4): 211-229.
- Dawson, R. J. and K. P. Locher (2006). "Structure of a bacterial multidrug ABC transporter." *Nature* 443(7108): 180-185.
- Dawson, R. J. and K. P. Locher (2007). "Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP." *FEBS Lett* 581(5): 935-938.
- Dean M, Rando R, Ishikawa T and L. V. "ATP binding cassette transporters (ABC)." from <https://www.genenames.org/cgi-bin/genefamilies/set/417>.
- Dean, M., A. Rzhetsky and R. Allikmets (2001). "The human ATP-binding cassette (ABC) transporter superfamily." *Genome Res* 11(7): 1156-1166.
- Debye, P. (1915). "Zerstreuung von Röntgenstrahlen." *Annalen der Physik* 351(6): 809-823.
- Deisenhofer, J., O. Epp, K. Miki, R. Huber and H. Michel (1985). "Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution." *Nature* 318(6047): 618-624.
- D'Imprima, E., D. Floris, M. Joppe, R. Sánchez, M. Grininger and W. Kühlbrandt (2019). "Protein denaturation at the air-water interface and how to prevent it." *eLife* 8: e42747.
- Doyle, L. A., W. Yang, L. V. Abruzzo, T. Krogmann, Y. Gao, A. K. Rishi and D. D. Ross (1998). "A multidrug resistance transporter from human MCF-7 breast cancer cells." *Proceedings of the National Academy of Sciences of the United States of America* 95(26): 15665-15670.
- Drew, D., S. Newstead, Y. Sonoda, H. Kim, G. von Heijne and S. Iwata (2008). "GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*." *Nat Protoc* 3(5): 784-798.
- Drew, D. E., G. von Heijne, P. Nordlund and J. W. de Gier (2001). "Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*." *FEBS Lett* 507(2): 220-224.

- Dubochet, J. and A. W. McDowell (1981). "VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY." *Journal of Microscopy* 124(3): 3-4.
- Ejendal, K. F., N. K. Diop, L. C. Schweiger and C. A. Hrycyna (2006). "The nature of amino acid 482 of human ABCG2 affects substrate transport and ATP hydrolysis but not substrate binding." *Protein Sci* 15(7): 1597-1607.
- Esser, L., F. Zhou, K. M. Pluchino, J. Shiloach, J. Ma, W. K. Tang, C. Gutierrez, A. Zhang, S. Shukla, J. P. Madigan, T. Zhou, P. D. Kwong, S. V. Ambudkar, M. M. Gottesman and D. Xia (2017). "Structures of the Multidrug Transporter P-glycoprotein Reveal Asymmetric ATP Binding and the Mechanism of Polyspecificity." *J Biol Chem* 292(2): 446-461.
- Evans, G. L., B. Ni, C. A. Hrycyna, D. Chen, S. V. Ambudkar, I. Pastan, U. A. Germann and M. M. Gottesman (1995). "Heterologous expression systems for P-glycoprotein: E. coli, yeast, and baculovirus." *J Bioenerg Biomembr* 27(1): 43-52.
- FDA (2019). FDA approves new breakthrough therapy for cystic fibrosis. , U.S. Food and drug administration.
- Forster, S., L. Apostol and W. Bras (2010). "Scatter: software for the analysis of nano- and mesoscale small-angle scattering." *Journal of Applied Crystallography* 43(3): 639-646.
- Franke, D., M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries and D. I. Svergun (2017). "ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions." *Journal of Applied Crystallography* 50(4): 1212-1225.
- Frederik, P. M. and D. H. W. Hubert (2005). *Cryoelectron Microscopy of Liposomes. Methods in Enzymology*, Academic Press. 391: 431-448.
- Freeze, H. H. and C. Kranz (2010). "Endoglycosidase and glycoamidase release of N-linked glycans." *Current protocols in molecular biology* Chapter 17: 10.1002/0471142727.mb0471141713as0471142789-0471142717.0471142713A.
- Frey, S., K. Sott, M. Smedh, T. Millat, P. Dahl, O. Wolkenhauer and M. Goksör (2011). "A mathematical analysis of nuclear intensity dynamics for Mig1-GFP under consideration

- of bleaching effects and background noise in *Saccharomyces cerevisiae*." *Mol Biosyst* 7(1): 215-223.
- Fu, Y., J. Zhao and Z. Chen (2018). "Insights into the Molecular Mechanisms of Protein-Ligand Interactions by Molecular Docking and Molecular Dynamics Simulation: A Case of Oligopeptide Binding Protein." *Computational and mathematical methods in medicine* 2018: 3502514-3502514.
- Gasteiger, J. and M. Marsili (1980). "Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges." *Tetrahedron* 36(22): 3219-3228.
- Goldie, K. N., P. Abeyrathne, F. Kebbel, M. Chami, P. Ringler and H. Stahlberg (2014). *Cryo-electron Microscopy of Membrane Proteins. Electron Microscopy: Methods and Protocols*. J. Kuo. Totowa, NJ, Humana Press: 325-341.
- Gose, T., T. Shafi, Y. Fukuda, S. Das, Y. Wang, A. Allcock, A. Gavan McHarg, J. Lynch, T. Chen, I. Tamai, A. Shelat, R. C. Ford and J. D. Schuetz (2020). "ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket." *Faseb j* 34(4): 4890-4903.
- Gottesman M.M and I Pastan (1988). "The multidrug transporter, a double-edged sword." *Journal of Biological Chemistry* 263: 12163-12166.
- Grant, T., A. Rohou and N. Grigorieff (2018). "cisTEM, user-friendly software for single-particle image processing." *eLife* 7: e35383.
- Grigorieff, N. and S. C. Harrison (2011). "Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy." *Current opinion in structural biology* 21(2): 265-273.
- Haider, A. J., M. H. Cox, N. Jones, A. J. Goode, K. S. Bridge, K. Wong, D. Briggs and I. D. Kerr (2015). "Identification of residues in ABCG2 affecting protein trafficking and drug transport, using co-evolutionary analysis of ABCG sequences." *Biosci Rep* 35(4).
- Halgren, T. A. (1996). "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94." *Journal of Computational Chemistry* 17(5-6): 490-519.

- Hauer, F., C. Gerle, N. Fischer, A. Oshima, K. Shinzawa-Itoh, S. Shimada, K. Yokoyama, Y. Fujiyoshi and H. Stark (2015). "GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM." *Structure* 23(9): 1769-1775.
- Hazai, E., I. Hazai, I. Ragueneau-Majlessi, S. P. Chung, Z. Bikadi and Q. Mao (2013). "Predicting substrates of the human breast cancer resistance protein using a support vector machine method." *BMC bioinformatics* 14: 130-130.
- Henderson, E. (2011). *The beginner's guide to SAXS data processing and analysis*. MS, University of Georgia.
- Hendlich, M., F. Rippmann and G. Barnickel (1997). "LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins." *J Mol Graph Model* 15(6): 359-363, 389.
- Hiller, S., R. G. Garces, T. J. Malia, V. Y. Orekhov, M. Colombini and G. Wagner (2008). "Solution structure of the integral human membrane protein VDAC-1 in detergent micelles." *Science* 321(5893): 1206-1210.
- Hoffmaster, K. A., R. Z. Turncliff, E. L. LeCluyse, R. B. Kim, P. J. Meier and K. L. Brouwer (2004). "P-glycoprotein expression, localization, and function in sandwich-cultured primary rat and human hepatocytes: relevance to the hepatobiliary disposition of a model opioid peptide." *Pharm Res* 21(7): 1294-1302.
- Hura, G., C. Hodge, D. Rosenberg, D. Guzenko, J. Duarte, B. Monastyrskyy, S. Grudinin, A. Kryshtafovych, J. Tainer, K. Fidelis and S. Tsutakawa (2019). "Small angle X-ray scattering-assisted protein structure prediction in CASP13 and emergence of solution structure differences." *Proteins: Structure, Function, and Bioinformatics* 87.
- I. Svergun, D., M. Koch, P. A. Timmins and R. May (2013). *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*.
- Jackson, S. M., I. Manolaridis, J. Kowal, M. Zechner, N. M. I. Taylor, M. Bause, S. Bauer, R. Bartholomaeus, G. Bernhardt, B. Koenig, A. Buschauer, H. Stahlberg, K. H. Altmann and



- K. P. Locher (2018). "Structural basis of small-molecule inhibition of human multidrug transporter ABCG2." *Nat Struct Mol Biol* 25(4): 333-340.
- Jackson, S. M., I. Manolaridis, J. Kowal, M. Zechner, N. M. I. Taylor, M. Bause, S. Bauer, R. Bartholomaeus, G. Bernhardt, B. Koenig, A. Buschauer, H. Stahlberg, K.-H. Altmann and K. P. Locher (2018). "Structural basis of small-molecule inhibition of human multidrug transporter ABCG2." *Nature Structural & Molecular Biology* 25(4): 333-340.
- Jacques, D. A. and J. Trehwella (2010). "Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls." *Protein Sci* 19(4): 642-657.
- Jardetzky, O. (1966). "Simple allosteric model for membrane pumps." *Nature* 211(5052): 969-970.
- Jeffries, C. M., M. A. Graewert, C. E. Blanchet, D. B. Langley, A. E. Whitten and D. I. Svergun (2016). "Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments." *Nat Protoc* 11(11): 2122-2153.
- Jeong, H.-S., H.-N. Park, J.-G. Kim and J.-K. Hyun (2013). "Critical importance of the correction of contrast transfer function for transmission electron microscopy-mediated structural biology." *Journal of Analytical Science and Technology* 4(1): 14.
- Jin, F., C. Shen, Y. Wang, M. Wang, M. Sun and M. Hattori (2021). "Fluorescence-detection size-exclusion chromatography utilizing nanobody technology for expression screening of membrane proteins." *Communications Biology* 4(1): 366.
- Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). "Development and validation of a genetic algorithm for flexible docking" Edited by F. E. Cohen." *Journal of Molecular Biology* 267(3): 727-748.
- Juliano, R. L. and V. Ling (1976). "A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants." *Biochim Biophys Acta* 455(1): 152-162.
- Kampjut, D., J. Steiner and L. A. Sazanov (2021). "Cryo-EM grid optimization for membrane proteins." *iScience* 24(3): 102139-102139.

- Karssen, A. M., O. Meijer, D. Pons and E. R. De Kloet (2004). "Localization of mRNA expression of P-glycoprotein at the blood-brain barrier and in the hippocampus." *Ann N Y Acad Sci* 1032: 308-311.
- Kesidis, A., P. Depping, A. Lodé, A. Vaitopoulou, R. M. Bill, A. D. Goddard and A. J. Rothnie (2020). "Expression of eukaryotic membrane proteins in eukaryotic and prokaryotic hosts." *Methods* 180: 3-18.
- Khunweeraphong, N. and K. Kuchler (2020). "The first intracellular loop is essential for the catalytic cycle of the human ABCG2 multidrug resistance transporter." *FEBS letters* 594(23): 4059-4075.
- Khunweeraphong, N., D. Szöllősi, T. Stockner and K. Kuchler (2019). "The ABCG2 multidrug transporter is a pump gated by a valve and an extracellular lid." *Nature Communications* 10(1): 5433.
- Kim, S., J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton (2021). "PubChem in 2021: new data content and improved web interfaces." *Nucleic Acids Res* 49(D1): D1388-d1395.
- Kim, Y. and J. Chen (2018). "Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation." *Science* 359(6378): 915-919.
- Kirton, S. B., C. W. Murray, M. L. Verdonk and R. D. Taylor (2005). "Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins." *Proteins* 58(4): 836-844.
- Koch, M. H. J., P. Vachette and D. I. Svergun (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." *Quarterly Reviews of Biophysics* 36(2): 147-227.
- Kohlstaedt, M., I. von der Hocht, F. Hilbers, Y. Thielmann and H. Michel (2015). "Development of a Thermofluor assay for stability determination of membrane proteins using the Na(+)/H(+) antiporter NhaA and cytochrome c oxidase." *Acta Crystallogr D Biol Crystallogr* 71(Pt 5): 1112-1122.

- Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." *Journal of Applied Crystallography* 36(5): 1277-1282.
- Korasick, D. A. and J. J. Tanner (2018). "Determination of protein oligomeric structure from small-angle X-ray scattering." *Protein science : a publication of the Protein Society* 27(4): 814-824.
- Koutsioubas, A. (2017). "Low-Resolution Structure of Detergent-Solubilized Membrane Proteins from Small-Angle Scattering Data." *Biophysical Journal* 113(11): 2373-2382.
- Kowal, J., D. Ni, S. M. Jackson, I. Manolaridis, H. Stahlberg and K. P. Locher (2021). "Structural Basis of Drug Recognition by the Multidrug Transporter ABCG2." *J Mol Biol* 433(13): 166980.
- Krchniakova, M., J. Skoda, J. Neradil, P. Chlapek and R. Veselska (2020). "Repurposing Tyrosine Kinase Inhibitors to Overcome Multidrug Resistance in Cancer: A Focus on Transporters and Lysosomal Sequestration." *Int J Mol Sci* 21(9).
- Kühlbrandt, W. (2014). "The Resolution Revolution." *Science* 343(6178): 1443-1444.
- Lacapere, J. J., E. Pebay-Peyroula, J. M. Neumann and C. Etchebest (2007). "Determining membrane protein structures: still a challenge!" *Trends Biochem Sci* 32(6): 259-270.
- LaConte, L. E. W., S. Srivastava and K. Mukherjee (2017). "Probing Protein Kinase-ATP Interactions Using a Fluorescent ATP Analog." *Methods in molecular biology* (Clifton, N.J.) 1647: 171-183.
- Le, C. A., D. S. Harvey and S. G. Aller (2020). "Structural definition of polyspecific compensatory ligand recognition by P-glycoprotein." *IUCrJ* 7(Pt 4): 663-672.
- Le, M.-T., V.-N. Hoang, D.-N. Nguyen, T.-H.-L. Bui, T.-V. Phan, P. N.-H. Huynh, T.-D. Tran and K.-M. Thai (2021). "Structure-Based Discovery of ABCG2 Inhibitors: A Homology Protein-Based Pharmacophore Modeling and Molecular Docking Approach." *Molecules* (Basel, Switzerland) 26(11): 3115.

- Lee, J.-Y., L. N. Kinch, D. M. Borek, J. Wang, J. Wang, I. L. Urbatsch, X.-S. Xie, N. V. Grishin, J. C. Cohen, Z. Otwinowski, H. H. Hobbs and D. M. Rosenbaum (2016). "Crystal structure of the human sterol transporter ABCG5/ABCG8." *Nature* 533(7604): 561-564.
- Lerner-Marmarosh, N., K. Gimi, I. L. Urbatsch, P. Gros and A. E. Senior (1999). "Large scale purification of detergent-soluble P-glycoprotein from *Pichia pastoris* cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins." *J Biol Chem* 274(49): 34711-34718.
- Li, J., K. F. Jaimes and S. G. Aller (2014). "Refined structures of mouse P-glycoprotein." *Protein Sci* 23(1): 34-46.
- Li, L., C. C. Koh, D. Reker, J. B. Brown, H. Wang, N. K. Lee, H.-h. Liow, H. Dai, H.-M. Fan, L. Chen and D.-Q. Wei (2019). "Predicting protein-ligand interactions based on bow-pharmacological space and Bayesian additive regression trees." *Scientific Reports* 9(1): 7703.
- Li, M. J., M. Guttman and W. M. Atkins (2018). "Conformational dynamics of P-glycoprotein in lipid nanodiscs and detergent micelles reveal complex motions on a wide time scale." *The Journal of biological chemistry* 293(17): 6297-6307.
- Li, X., P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard and Y. Cheng (2013). "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM." *Nature Methods* 10(6): 584-590.
- Lichtenberg, D., H. Ahyayauch and F. M. Goñi (2013). "The mechanism of detergent solubilization of lipid bilayers." *Biophysical journal* 105(2): 289-299.
- Lingam, S. (2017). Expression And Purification Of The Multidrug Resistance Protein P-Glycoprotein And High-Throughput Assay Development For Drug Discovery. Ph.D., University of Manchester.
- Lingam, S., N. Thonghin and R. C. Ford (2017). "Investigation of the effects of the CFTR potentiator ivacaftor on human P-glycoprotein (ABCB1)." *Scientific reports* 7(1): 17481-17481.

- Linton, K. J., Holland, I Barry (2011). The ABC Transporters of Human Physiology and Disease.
- Liu, F., Z. Zhang, A. Levit, J. Levring, K. K. Touhara, B. K. Shoichet and J. Chen (2019). "Structural identification of a hotspot on CFTR for potentiation." *Science* 364(6446): 1184.
- Lodish H, Berk A, Zipursky S.L., Matsudaira P, Baltimore D and Darnell J (2000). *Molecular cell biology*. New York, W. H. Freeman.
- Loll, P. J. (2014). "Membrane proteins, detergents and crystals: what is the state of the art?" *Acta crystallographica. Section F, Structural biology communications* 70(Pt 12): 1576-1583.
- Loo, T. W. and D. M. Clarke (2002). "Location of the Rhodamine-binding Site in the Human Multidrug Resistance P-glycoprotein \*." *Journal of Biological Chemistry* 277(46): 44332-44338.
- Lusvarghi, S. and S. V. Ambudkar (2019). "ATP-dependent thermostabilization of human P-glycoprotein (ABCB1) is blocked by modulators." *The Biochemical journal* 476(24): 3737-3750.
- Lysko, K. A., R. Carlson, R. Taverna, J. Snow and J. F. Brandts (1981). "Protein involvement in structural transition of erythrocyte ghosts. Use of thermal gel analysis to detect protein aggregation." *Biochemistry* 20(19): 5570-5576.
- Lyumkis, D. (2019). "Challenges and opportunities in cryo-EM single-particle analysis." *The Journal of biological chemistry* 294(13): 5181-5197.
- Mahmood, T. and P.-C. Yang (2012). "Western blot: technique, theory, and trouble shooting." *North American journal of medical sciences* 4(9): 429-434.
- Majorovits, E., B. Barton, K. Schultheiß, F. Pérez-Willard, D. Gerthsen and R. R. Schröder (2007). "Optimizing phase contrast in transmission electron microscopy with an electrostatic (Boersch) phase plate." *Ultramicroscopy* 107(2): 213-226.
- Manalastas-Cantos, K., P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S. Molodenskiy, A. Panjkovich, H. D. T. Mertens, A. Gruzinov, C. Borges, C. M. Jeffries, D. I.

- Svergun and D. Franke (2021). "ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis." *Journal of Applied Crystallography* 54(1): 343-355.
- Maniatis, T., E. F. Fritsch and J. Sambrook (1982). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y, Cold Spring Harbor Laboratory.
- Manolaridis, I., S. M. Jackson, N. M. I. Taylor, J. Kowal, H. Stahlberg and K. P. Locher (2018). "Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states." *Nature* 563(7731): 426-430.
- Mao, Q., G. Conseil, A. Gupta, S. P. Cole and J. D. Unadkat (2004). "Functional expression of the human breast cancer resistance protein in *Pichia pastoris*." *Biochem Biophys Res Commun* 320(3): 730-737.
- Marcoux, J., 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 (2013). "Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump." *Proc Natl Acad Sci U S A* 110(24): 9704-9709.
- Marighetti, F., K. Steggemann, M. Hanl and M. Wiese (2013). "Synthesis and quantitative structure-activity relationships of selective BCRP inhibitors." *ChemMedChem* 8(1): 125-135.
- Marques, S. M., L. Šupolíková, L. Molčanová, K. Šmejkal, D. Bednar and I. Slaninová (2021). "Screening of Natural Compounds as P-Glycoprotein Inhibitors against Multidrug Resistance." *Biomedicines* 9(4).
- Martinec, O., M. Huliciak, F. Staud, F. Cecka, I. Vokral and L. Cervený (2019). "Anti-HIV and Anti-Hepatitis C Virus Drugs Inhibit P-Glycoprotein Efflux Activity in Caco-2 Cells and Precision-Cut Rat and Human Intestinal Slices." *Antimicrobial agents and chemotherapy* 63(11): e00910-00919.
- Mazard, T., A. Causse, J. Simony, W. Leconet, N. Vezzio-Vie, A. Torro, M. Jarlier, A. Evrard, M. Del Rio, E. Assenat, P. Martineau, M. Ychou, B. Robert and C. Gongora (2013). "Sorafenib

- Overcomes Irinotecan Resistance in Colorectal Cancer by Inhibiting the ABCG2 Drug-Efflux Pump." *Molecular Cancer Therapeutics* 12(10): 2121.
- McDevitt, C. A., C. A. Shintre, J. G. Grossmann, N. L. Pollock, S. M. Prince, R. Callaghan and R. C. Ford (2008). "Structural insights into P-glycoprotein (ABCB1) by small angle X-ray scattering and electron crystallography." *FEBS Lett* 582(19): 2950-2956.
- McPherson, A. (2004). "Introduction to protein crystallization." *Methods* 34(3): 254-265.
- Mellors, A. and D. R. Sutherland (1994). "Tools to cleave glycoproteins." *Trends in Biotechnology* 12(1): 15-18.
- Melo, F., R. Sánchez and A. Sali (2002). "Statistical potentials for fold assessment." *Protein Science* 11(2): 430-448.
- Miyake, K., L. Mickley, T. Litman, Z. Zhan, R. Robey, B. Cristensen, M. Brangi, L. Greenberger, M. Dean, T. Fojo and S. E. Bates (1999). "Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes." *Cancer Res* 59(1): 8-13.
- Molecular Operating Environment (2021). Chemical Computing Group ULC,. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7,.
- Molodenskiy, D. S., H. D. T. Mertens and D. I. Svergun (2020). "An automated data processing and analysis pipeline for transmembrane proteins in detergent solutions." *Scientific Reports* 10(1): 8081.
- Molsa, M., T. Heikkinen, J. Hakkola, K. Hakala, O. Wallerman, M. Wadelius, C. Wadelius and K. Laine (2005). "Functional role of P-glycoprotein in the human blood-placental barrier." *Clin Pharmacol Ther* 78(2): 123-131.
- Montecinos-Franjola, F., B. L. Bauer, J. A. Mears and R. Ramachandran (2020). "GFP fluorescence tagging alters dynamin-related protein 1 oligomerization dynamics and creates disassembly-refractory puncta to mediate mitochondrial fission." *Scientific Reports* 10(1): 14777.

- Müller, H., J. Jin, R. Danev, J. Spence, H. Padmore and R. M. Glaeser (2010). "Design of an electron microscope phase plate using a focused continuous-wave laser." *New Journal of Physics* 12(7): 073011.
- Murata, K. and M. Wolf (2018). "Cryo-electron microscopy for structural analysis of dynamic biological macromolecules." *Biochim Biophys Acta Gen Subj* 1862(2): 324-334.
- Naïm, M., S. Bhat, K. N. Rankin, S. Dennis, S. F. Chowdhury, I. Siddiqi, P. Drabik, T. Sulea, C. I. Bayly, A. Jakalian and E. O. Purisima (2007). "Solvated Interaction Energy (SIE) for Scoring Protein–Ligand Binding Affinities. 1. Exploring the Parameter Space." *Journal of Chemical Information and Modeling* 47(1): 122-133.
- Newstead, S., H. Kim, G. von Heijne, S. Iwata and D. Drew (2007). "High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*." *Proceedings of the National Academy of Sciences of the United States of America* 104(35): 13936-13941.
- Ni, Z., Z. Bikadi, D. L. Shuster, C. Zhao, M. F. Rosenberg and Q. Mao (2011). "Identification of proline residues in or near the transmembrane helices of the human breast cancer resistance protein (BCRP/ABCG2) that are important for transport activity and substrate specificity." *Biochemistry* 50(37): 8057-8066.
- Nicklisch, S. C., S. D. Rees, A. P. McGrath, T. Gökirmak, L. T. Bonito, L. M. Vermeer, C. Cregger, G. Loewen, S. Sandin, G. Chang and A. Hamdoun (2016). "Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure." *Sci Adv* 2(4): e1600001.
- Noble, A. J., V. P. Dandey, H. Wei, J. Brasch, J. Chase, P. Acharya, Y. Z. Tan, Z. Zhang, L. Y. Kim, G. Scapin, M. Rapp, E. T. Eng, W. J. Rice, A. Cheng, C. J. Negro, L. Shapiro, P. D. Kwong, D. Jeruzalmi, A. des Georges, C. S. Potter and B. Carragher (2018). "Routine single particle CryoEM sample and grid characterization by tomography." *eLife* 7: e34257.
- Noble, A. J., H. Wei, V. P. Dandey, Z. Zhang, Y. Z. Tan, C. S. Potter and B. Carragher (2018). "Reducing effects of particle adsorption to the air–water interface in cryo-EM." *Nature Methods* 15(10): 793-795.



- Nogales, E. and Sjors H. W. Scheres (2015). "Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity." *Molecular Cell* 58(4): 677-689.
- Nosol, K., K. Romane, R. N. Irobalieva, A. Alam, J. Kowal, N. Fujita and K. P. Locher (2020). "Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter ABCB1." *Proc Natl Acad Sci U S A* 117(42): 26245-26253.
- O'Connor, C. M. (2021). "Regulation of the GAL1 promoter."
- O'Dell, W. (2021). *Recombinant Protein Expression: Eukaryotic hosts*, Academic Press.
- Orlando, B. J. and M. Liao (2020). "ABCG2 transports anticancer drugs via a closed-to-open switch." *Nat Commun* 11(1): 2264.
- O'Ryan, L., T. Rimington, N. Cant and R. C. Ford (2012). "Expression and purification of the cystic fibrosis transmembrane conductance regulator protein in *Saccharomyces cerevisiae*." *J Vis Exp*(61).
- Ozvegy, C., T. Litman, G. Szakacs, Z. Nagy, S. Bates, A. Varadi and B. Sarkadi (2001). "Functional characterization of the human multidrug transporter, ABCG2, expressed in insect cells." *Biochem Biophys Res Commun* 285(1): 111-117.
- Passmore, L. A. and C. J. Russo (2016). "Specimen Preparation for High-Resolution Cryo-EM." *Methods in enzymology* 579: 51-86.
- Penczek, P. A., R. A. Grassucci and J. Frank (1994). "The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles." *Ultramicroscopy* 53(3): 251-270.
- Pérez, J. and A. Koutsioubas (2015). "Memprot: a program to model the detergent corona around a membrane protein based on SEC-SAXS data." *Acta Crystallogr D Biol Crystallogr* 71(Pt 1): 86-93.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem* 25(13): 1605-1612.

- Pick, A., H. Müller, R. Mayer, B. Haenisch, I. K. Pajeva, M. Weigt, H. Bönisch, C. E. Müller and M. Wiese (2011). "Structure-activity relationships of flavonoids as inhibitors of breast cancer resistance protein (BCRP)." *Bioorg Med Chem* 19(6): 2090-2102.
- Pollock, N., N. Cant, T. Rimington and R. C. Ford (2014). "Purification of the cystic fibrosis transmembrane conductance regulator protein expressed in *Saccharomyces cerevisiae*." *Journal of visualized experiments : JoVE*(87): 51447.
- Qu, Q., P. L. Russell and F. J. Sharom (2003). "Stoichiometry and affinity of nucleotide binding to P-glycoprotein during the catalytic cycle." *Biochemistry* 42(4): 1170-1177.
- Rabindran, S. K., D. D. Ross, L. A. Doyle, W. Yang and L. M. Greenberger (2000). "Fumitremorgin C reverses multidrug resistance in cells transfected with the breast cancer resistance protein." *Cancer research* 60(1): 47-50.
- Ren, X.-Q., T. Furukawa, M. Yamamoto, S. Aoki, M. Kobayashi, M. Nakagawa and S.-i. Akiyama (2006). "A Functional Role of Intracellular Loops of Human Multidrug Resistance Protein 1." *The Journal of Biochemistry* 140(3): 313-318.
- Rice, S. A. (1956). "Small angle scattering of X-rays. A. Guinier and G. Fournet. Translated by C. B. Wilson and with a bibliographical appendix by K. L. Yudowitch. Wiley, New York, 1955. 268 pp. \$7.50." *Journal of Polymer Science* 19(93): 594-594.
- Rivas-Pardo, J. A., A. Herrera-Morande, V. Castro-Fernandez, F. J. Fernandez, M. C. Vega and V. Guixé (2013). "Crystal structure, SAXS and kinetic mechanism of hyperthermophilic ADP-dependent glucokinase from *Thermococcus litoralis* reveal a conserved mechanism for catalysis." *PloS one* 8(6): e66687-e66687.
- Robey, R. W., K. M. Pluchino, M. D. Hall, A. T. Fojo, S. E. Bates and M. M. Gottesman (2018). "Revisiting the role of ABC transporters in multidrug-resistant cancer." *Nature Reviews Cancer* 18(7): 452-464.
- Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." *Journal of Structural Biology* 192(2): 216-221.

- Roninson, I. B., J. E. Chin, K. G. Choi, P. Gros, D. E. Housman, A. Fojo, D. W. Shen, M. M. Gottesman and I. Pastan (1986). "Isolation of human *mdr* DNA sequences amplified in multidrug-resistant KB carcinoma cells." *Proceedings of the National Academy of Sciences of the United States of America* 83(12): 4538-4542.
- Rosenstein, R. E., C. Sanjurjo and D. P. Cardinali (1989). "gamma Aminobutyric acid uptake, release, and effect on  $^{36}\text{Cl}^-$  influx in bovine pineal gland." *J Neural Transm* 77(2-3): 141-152.
- Ryan, T. M., J. Trewhella, J. M. Murphy, J. R. Keown, L. Casey, F. G. Pearce, D. C. Goldstone, K. Chen, Z. Luo, B. Kobe, C. A. McDevitt, S. A. Watkin, A. M. Hawley, S. T. Mudie, V. Samardzic Boban and N. Kirby (2018). "An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis." *Journal of Applied Crystallography* 51(1): 97-111.
- Saikia, S. and M. Bordoloi (2019). "Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective." *Current Drug Targets* 20(5): 501-521.
- Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." *J Mol Biol* 234(3): 779-815.
- Sarkadi, B., E. M. Price, R. C. Boucher, U. A. Germann and G. A. Scarborough (1992). "Expression of the human multidrug resistance cDNA in insect cells generates a high activity drug-stimulated membrane ATPase." *J Biol Chem* 267(7): 4854-4858.
- Scarborough, G. A. (1995). "Drug-stimulated ATPase activity of the human P-glycoprotein." *Journal of bioenergetics and biomembranes* 27(1): 37-41.
- Schinkel, A. H., S. Kemp, M. Dollé, G. Rudenko and E. Wagenaar (1993). "N-glycosylation and deletion mutants of the human MDR1 P-glycoprotein." *J Biol Chem* 268(10): 7474-7481.
- Schneidman-Duhovny, D., M. Hammel and A. Sali (2010). "FoXS: a web server for rapid computation and fitting of SAXS profiles." *Nucleic acids research* 38(Web Server issue): W540-W544.

- Schneidman-Duhovny, D., M. Hammel, J. A. Tainer and A. Sali (2013). "Accurate SAXS profile computation and its assessment by contrast variation experiments." *Biophysical journal* 105(4): 962-974.
- Seddon, A. M., P. Curnow and P. J. Booth (2004). "Membrane proteins, lipids and detergents: not just a soap opera." *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1666(1): 105-117.
- Senior, A. E., M. K. al-Shawi and I. L. Urbatsch (1995). "The catalytic cycle of P-glycoprotein." *FEBS Lett* 377(3): 285-289.
- Senisterra, G., I. Chau and M. Vedadi (2011). "Thermal Denaturation Assays in Chemical Biology." *ASSAY and Drug Development Technologies* 10(2): 128-136.
- Sgro, G. G. and T. R. D. Costa (2018). "Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis." *Frontiers in Molecular Biosciences* 5(74).
- Shen, M.-Y. and A. Sali (2006). "Statistical potential for assessment and prediction of protein structures." *Protein science : a publication of the Protein Society* 15(11): 2507-2524.
- Shi, Z., S. Parmar, X.-X. Peng, T. Shen, R. W. Robey, S. E. Bates, L.-W. Fu, Y. Shao, Y.-M. Chen, F. Zang and Z.-S. Chen (2009). "The epidermal growth factor tyrosine kinase inhibitor AG1478 and erlotinib reverse ABCG2-mediated drug resistance." *Oncol Rep* 21(2): 483-489.
- Shintre, C. A. (2009). Structural studies of P-glycoprotein using small angle scattering and electron microscopy. Phd Traditional thesis, The Univeristy of manchester
- Shiozawa, K., M. Oka, H. Soda, M. Yoshikawa, Y. Ikegami, J. Tsurutani, K. Nakatomi, Y. Nakamura, S. Doi, T. Kitazaki, Y. Mizuta, K. Murase, H. Yoshida, D. D. Ross and S. Kohno (2004). "Reversal of breast cancer resistance protein (BCRP/ABCG2)-mediated drug resistance by novobiocin, a coumermycin antibiotic." *Int J Cancer* 108(1): 146-151.
- Shukla, S., A. Kouanda, L. Silverton, T. T. Talele and S. V. Ambudkar (2014). "Pharmacophore modeling of nilotinib as an inhibitor of ATP-binding cassette drug transporters and BCR-

- ABL kinase using a three-dimensional quantitative structure-activity relationship approach." *Mol Pharm* 11(7): 2313-2322.
- Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson and D. G. Higgins (2011) "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." *Molecular systems biology* 7, 539 DOI: 10.1038/msb.2011.75.
- Sigworth, F. J. (2016). "Principles of cryo-EM single-particle image processing." *Microscopy (Oxf)* 65(1): 57-67.
- Smith, S. M. (2011). "Strategies for the purification of membrane proteins." *Methods Mol Biol* 681: 485-496.
- Sodani, K., A. Patel, R. J. Kathawala and Z.-S. Chen (2012). "Multidrug resistance associated proteins in multidrug resistance." *Chinese journal of cancer* 31(2): 58-72.
- Sonoda, Y., S. Newstead, N.-J. Hu, Y. Alguel, E. Nji, K. Beis, S. Yashiro, C. Lee, J. Leung, A. D. Cameron, B. Byrne, S. Iwata and D. Drew (2011). "Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures." *Structure (London, England : 1993)* 19(1): 17-25.
- Srinivasan, V., A. J. Pierik and R. Lill (2014). "Crystal structures of nucleotide-free and glutathione-bound mitochondrial ABC transporter Atm1." *Science* 343(6175): 1137-1140.
- Stansfeld, P. J., J. E. Goose, M. Caffrey, E. P. Carpenter, J. L. Parker, S. Newstead and M. S. P. Sansom (2015). "MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes." *Structure (London, England : 1993)* 23(7): 1350-1361.
- Svergun, D., M. Koch, P. Timmins and R. May (2013). *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*.
- Szewczyk, P., H. Tao, A. P. McGrath, M. Villaluz, S. D. Rees, S. C. Lee, R. Doshi, I. L. Urbatsch, Q. Zhang and G. Chang (2015). "Snapshots of ligand entry, malleable binding and induced

- helical movement in P-glycoprotein." *Acta Crystallogr D Biol Crystallogr* 71(Pt 3): 732-741.
- Szöllősi, D., P. Chiba, G. Szakacs and T. Stockner (2020). "Conversion of chemical to mechanical energy by the nucleotide binding domains of ABCB1." *Scientific reports* 10(1): 2589-2589.
- Taylor, N. M. I., I. Manolaridis, S. M. Jackson, J. Kowal, H. Stahlberg and K. P. Locher (2017). "Structure of the human multidrug transporter ABCG2." *Nature* 546(7659): 504-509.
- Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Applied microbiology and biotechnology* 72(2): 211.
- Terstappen, G. C. and A. Reggiani (2001). "In silico research in drug discovery." *Trends Pharmacol Sci* 22(1): 23-26.
- Thonghin, N., R. F. Collins, A. Barbieri, T. Shafi, A. Siebert and R. C. Ford (2018). "Novel features in the structure of P-glycoprotein (ABCB1) in the post-hydrolytic state as determined at 7.9 Å resolution." *BMC Structural Biology* 18(1): 17.
- Thonghin, N., V. Kargas, J. Clews and R. C. Ford (2018). "Cryo-electron microscopy of membrane proteins." *Methods* 147: 176-186.
- Tjernberg, A., N. Markova, W. J. Griffiths and D. Hallén (2006). "DMSO-Related Effects in Protein Characterization." *Journal of Biomolecular Screening* 11(2): 131-137.
- Tzvi Tzfira, R. b. (2005). "Recombinant Gene Expression: Reviews and Protocols. Second Edition. *Methods in Molecular Biology*, Volume 267. Edited by Paulina Balbás and Argelia Lorence." *The Quarterly Review of Biology* 80(4): 475-475.
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigartyo CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J and P. F. "ABCG2." *Tissue-based map*

of the human proteome. Retrieved 19th november, 2017, from <https://www.proteinatlas.org/ENSG00000118777-ABCG2/tissue>.

ULC, C. c. g. (2019). Molecular Operating Environment (MOE).

Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor (2003). "Improved protein–ligand docking using GOLD." *Proteins: Structure, Function, and Bioinformatics* 52(4): 609-623.

Verhalen, B., R. Dastvan, S. Thangapandian, Y. Peskova, H. A. Koteiche, R. K. Nakamoto, E. Tajkhorshid and H. S. McHaourab (2017). "Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein." *Nature* 543(7647): 738-741.

Verrier, P. J., D. Bird, B. Burla, E. Dassa, C. Forestier, M. Geisler, M. Klein, U. Kolukisaoglu, Y. Lee, E. Martinoia, A. Murphy, P. A. Rea, L. Samuels, B. Schulz, E. J. Spalding, K. Yazaki and F. L. Theodoulou (2008). "Plant ABC proteins--a unified nomenclature and updated inventory." *Trends Plant Sci* 13(4): 151-159.

Vieira Gomes, A. M., T. Souza Carmo, L. Silva Carvalho, F. Mendonça Bahia and N. S. Parachin (2018). "Comparison of Yeasts as Hosts for Recombinant Protein Production." *Microorganisms* 6(2): 38.

Vinothkumar, K. R. (2015). "Membrane protein structures without crystals, by single particle electron cryomicroscopy." *Curr Opin Struct Biol* 33: 103-114.

von Richter, O., O. Burk, M. F. Fromm, K. P. Thon, M. Eichelbaum and K. T. Kivisto (2004). "Cytochrome P450 3A4 and P-glycoprotein expression in human small intestinal enterocytes and hepatocytes: a comparative analysis in paired tissue specimens." *Clin Pharmacol Ther* 75(3): 172-183.

Wagner, S., M. L. Bader, D. Drew and J. W. de Gier (2006). "Rationalizing membrane protein overexpression." *Trends Biotechnol* 24(8): 364-371.

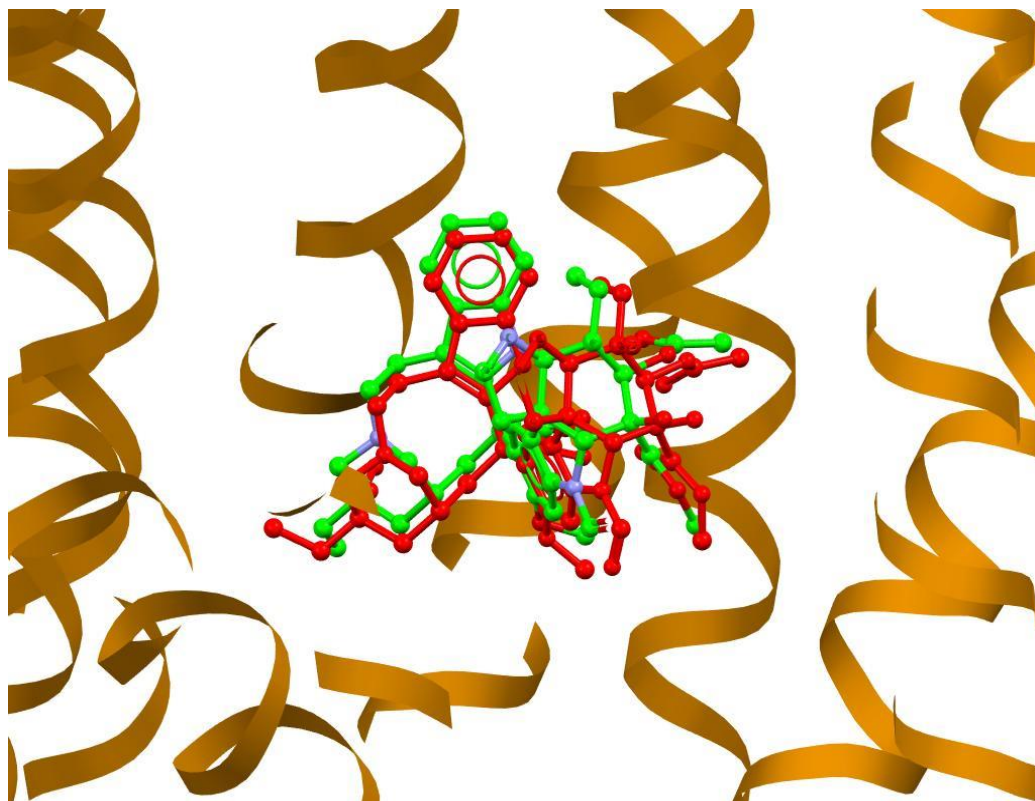
Wallin, E. and G. von Heijne (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." *Protein Sci* 7(4): 1029-1038.

- Wang, X., T. Furukawa, T. Nitanda, M. Okamoto, Y. Sugimoto, S.-I. Akiyama and M. Baba (2003). "Breast cancer resistance protein (BCRP/ABCG2) induces cellular resistance to HIV-1 nucleoside reverse transcriptase inhibitors." *Molecular pharmacology* 63(1): 65-72.
- Ward, A. B., 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 (2013). "Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain." *Proc Natl Acad Sci U S A* 110(33): 13386-13391.
- Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." *Bioinformatics* 25(9): 1189-1191.
- Weiss, H. M. and R. Grisshammer (2002). "Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*." *Eur J Biochem* 269(1): 82-92.
- Weiss, J., J. Rose, C. H. Storch, N. Ketabi-Kiyanvash, A. Sauer, W. E. Haefeli and T. Efferth (2007). "Modulation of human BCRP (ABCG2) activity by anti-HIV drugs." *Journal of Antimicrobial Chemotherapy* 59(2): 238-245.
- Weissenberger, G., R. J. M. Henderikx and P. J. Peters (2021). "Understanding the invisible hands of sample preparation for cryo-EM." *Nature Methods* 18(5): 463-471.
- White, S. H. (2009). "Biophysical dissection of membrane proteins." *Nature* 459(7245): 344-346.
- Woodward, O. M., D. N. Tukaye, J. Cui, P. Greenwell, L. M. Constantoulakis, B. S. Parker, A. Rao, M. Köttgen, P. C. Maloney and W. B. Guggino (2013). "Gout-causing Q141K mutation in ABCG2 leads to instability of the nucleotide-binding domain and can be corrected with small molecules." *Proceedings of the National Academy of Sciences of the United States of America* 110(13): 5223-5228.

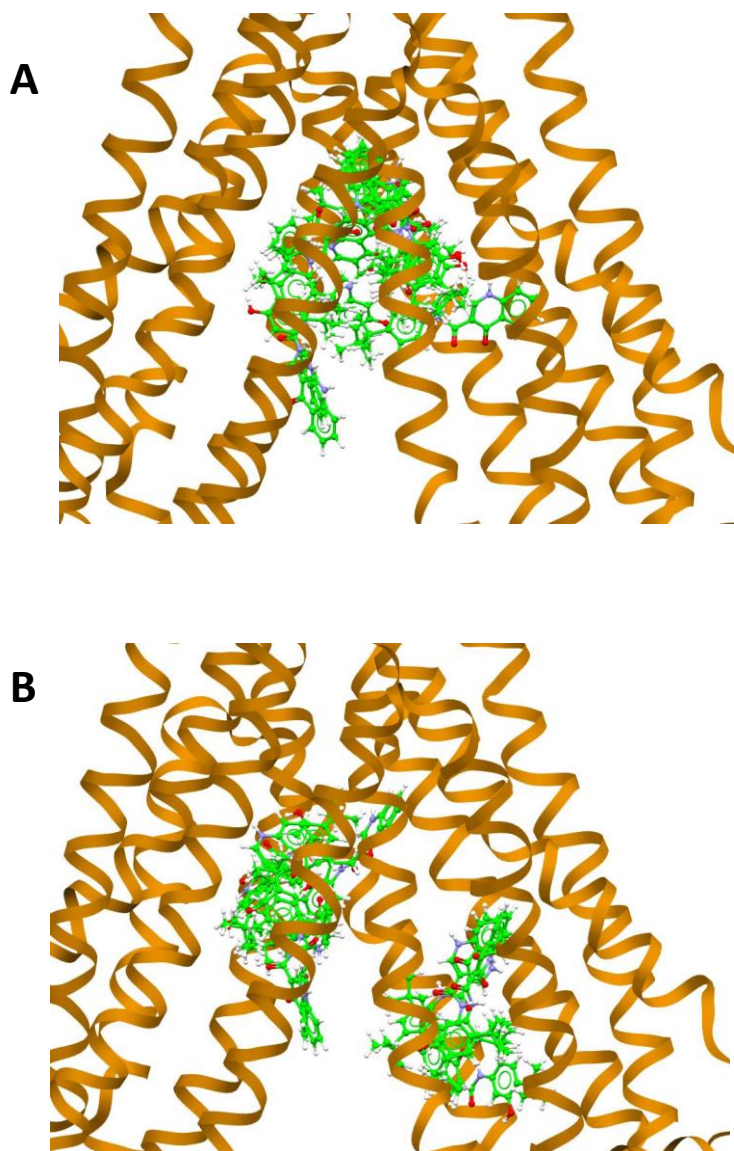


- Woolfson, M. M. (1997). *An Introduction to X-ray Crystallography*. Cambridge, Cambridge University Press.
- Xu, J., Y. Liu, Y. Yang, S. Bates and J. T. Zhang (2004). "Characterization of oligomeric human half-ABC transporter ATP-binding cassette G2." *J Biol Chem* 279(19): 19781-19789.
- Yang, S.-Y. (2010). "Pharmacophore modeling and applications in drug discovery: challenges and recent advances." *Drug Discovery Today* 15(11): 444-450.
- Yu, Q., D. Ni, J. Kowal, I. Manolaridis, S. M. Jackson, H. Stahlberg and K. P. Locher (2021). "Structures of ABCG2 under turnover conditions reveal a key step in the drug transport mechanism." *Nature Communications* 12(1): 4376.
- Zacharius, R. M., T. E. Zell, J. H. Morrison and J. J. Woodlock (1969). "Glycoprotein staining following electrophoresis on acrylamide gels." *Analytical Biochemistry* 30(1): 148-152.
- Zhang, K. (2016). "Gctf: Real-time CTF determination and correction." *J Struct Biol* 193(1): 1-12.
- Zhao, Z. and E. Tajkhorshid (2021). "GOLEM: Automated and Robust Cryo-EM-Guided Ligand Docking with Explicit Water Molecules." *Biophysical Journal* 120(3, Supplement 1): 290a.
- Zheng, S. Q., E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." *Nature Methods* 14(4): 331-332.
- Zhou, H.-X. and O. Bilsel (2014). "SAXS/SANS probe of intermolecular interactions in concentrated protein solutions." *Biophysical journal* 106(4): 771-773.

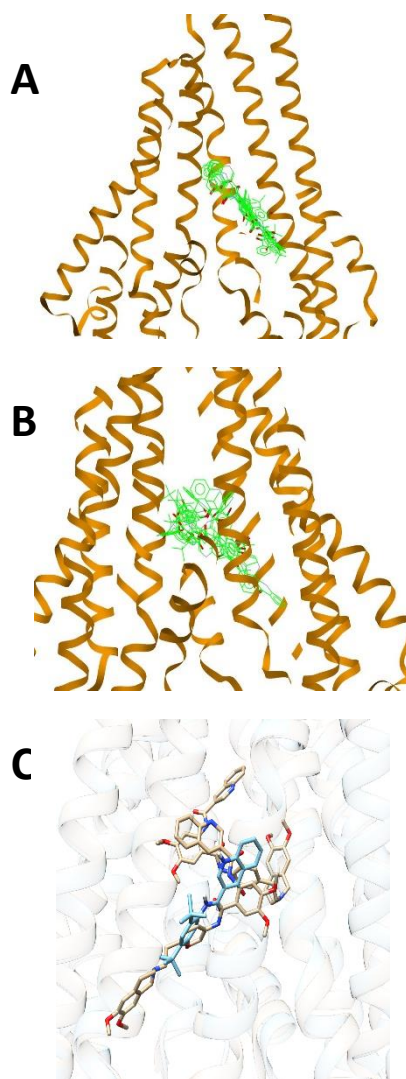
## Appendix



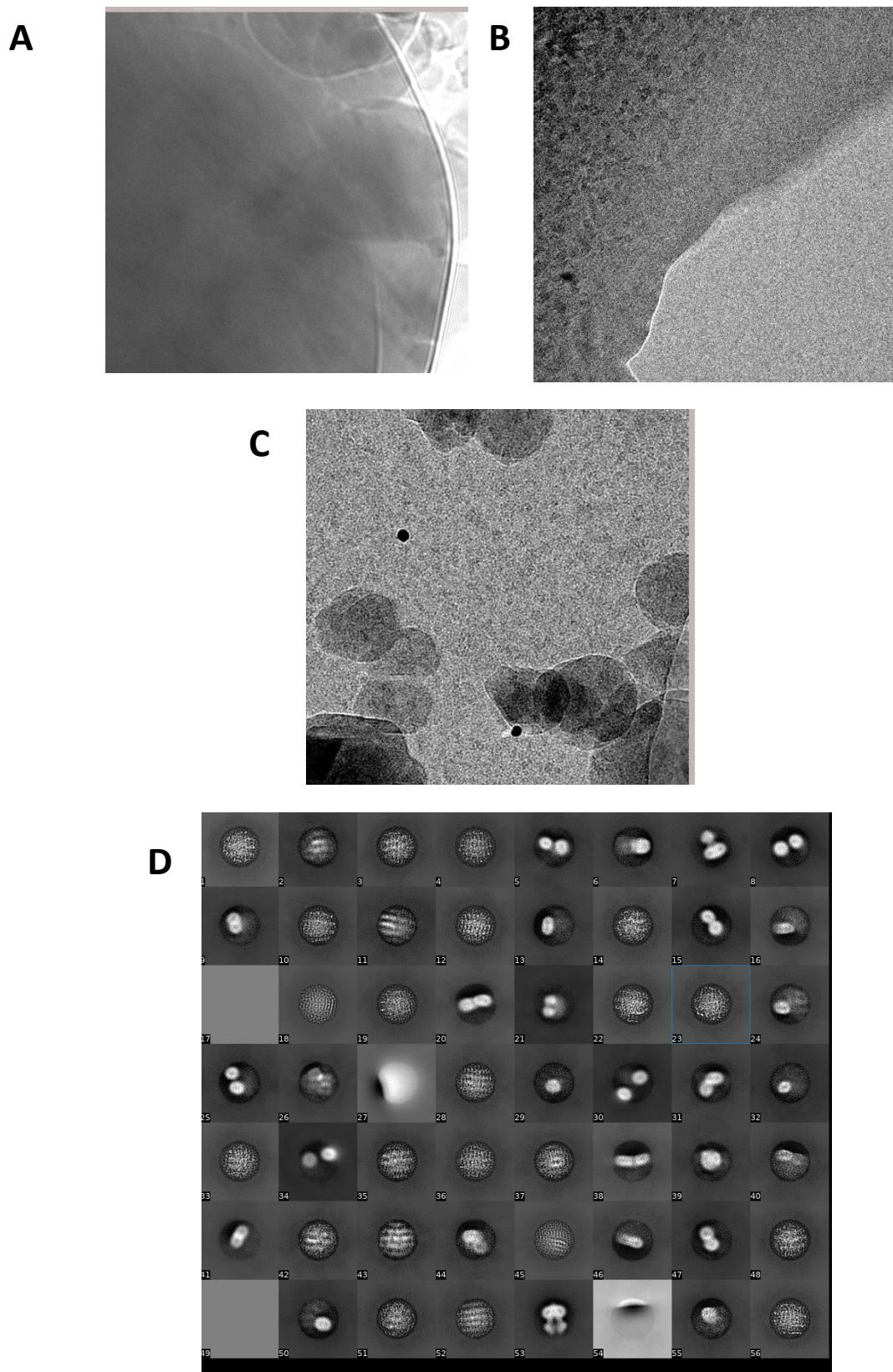
Appendix Figure 5.1. Redocking of vincristine with human P-glycoprotein. Docked conformation (green) and experimentally determined conformations



Appendix Figure 5.2 top ten conformations ivacaftor with mP-gp (4KSB). (A) top ten conformation form Chemplp. (B) top ten conformations from GOLD score

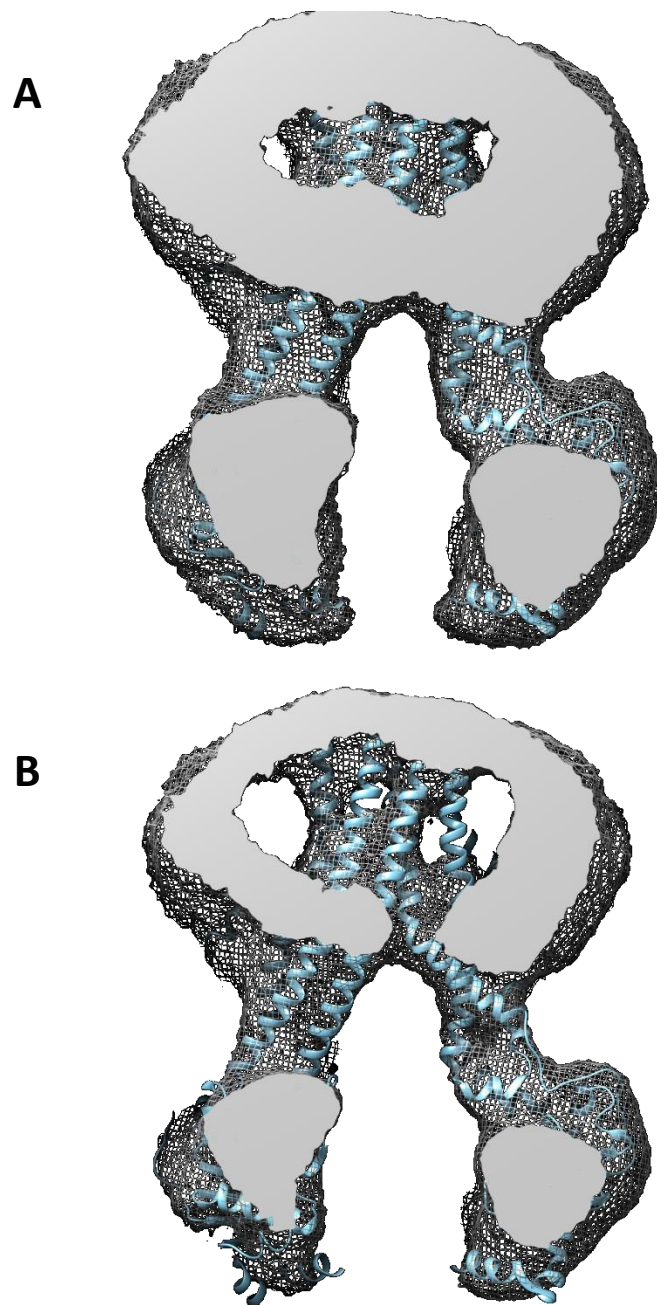


Appendix Figure 5.3. Top ten conformations from ivacaftor and human P-gp (7A69) docking. (A) chemplp (B) goldscore (C) location of ivacaftor (cyan) in comparison to tariquidar (brown) binding site identified from cryo-EM.

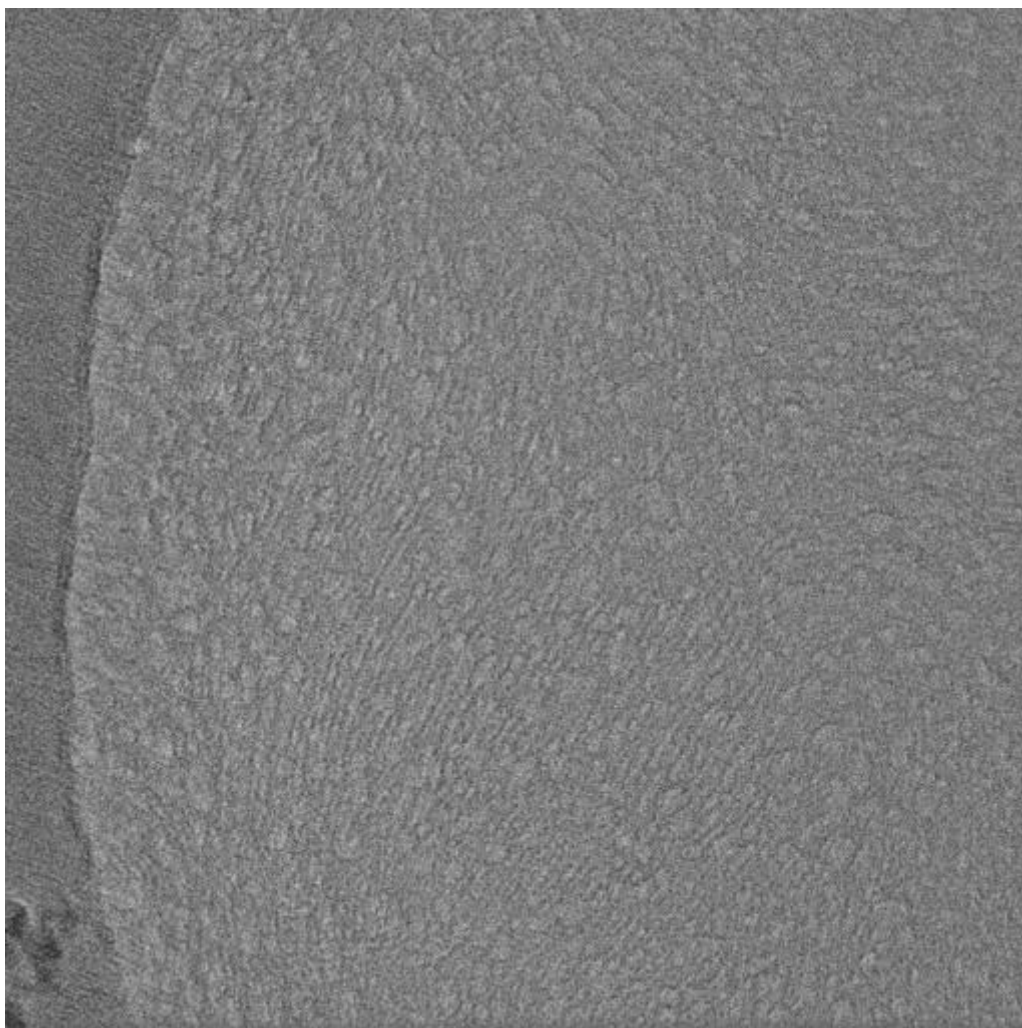


Appendix Figure 5.4. representative bad images (A) thick ice contamination. (B) Lensing effect thin ice in the middle and thick ice on the edges. (C) carbon support (D) initial 2D classification showing P-gp like particles, frog eye like micelles and carbon support





Appendix Figure 5.5 ABCB6 structure from high resolution dataset. 3D map generation using 12802 particles to compare the resolution from B1 data. (A) after refinement (B) at density level 1



Appendix Figure 5.6 image of DDM treated grids. It shows thumb print like pattern and no particles

Appendix Table 5.1 list of hits identified from virtual screening of FDA library against ABCB1

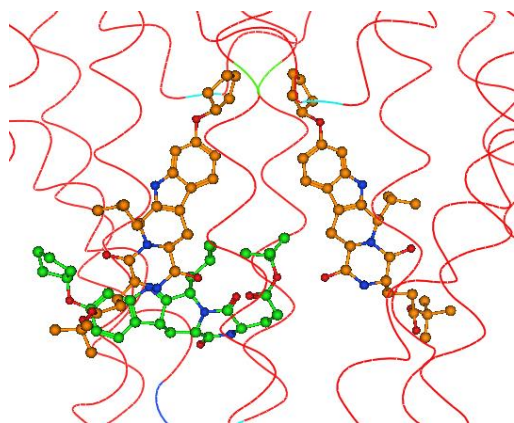
<b>Drugs</b>	<b>Target</b>	<b>Class</b>
miltefosine	akt	kinase
perifosine	akt	kinase
Ldk378	ALK	kinase
barasartib	aurora kinase	kinase
abt-199	bcl-2 family	kinase
abt 263	bcl-2 family	kinase
ponatinib	bcr-abl family	kinase
bosutinib	bcr -abl family	kinase
imatinib	bcr -abl family	kinase
nilotinib	bcr-abl family	kinase
saracatinib	bcr-abl family	kinase
masitinib	c-kit	kinase
cabozantinib	c-met	kinase
emd 1214063	c-met	kinase
dinaciclib	cyclin dependent kinase	kinase
ly2835219	cyclin dependent kinase	kinase
erlotinib	egfr	kinase
azd 9291	egfr	kinase
neratinib	egfr	kinase
lapatinib	egfr	kinase
co-1686	egfr	kinase
GLPG0634	jak	kinase
tg101348	JAK	kinase
ispinesib	ksp	kinase
cobimetinib	MEK1/2	kinase
trametinib	mek1/2	kinase
nintedanib	pdgfr	kinase
AZD7545	PDHK	kinase
sar245409	pi3k	kinase
enzastaurin	pkc	kinase
bi6727	plk	kinase
r788	spleen tyrosine kinase	kinase
vandetanib	vegfr	kinase
pazopanib	vegfr	kinase
vatalanib	vegfr	kinase
foretinib	vegfr	kinase
dutasteride	5 alpha reductase	non kinase
ziprasidone	5-HT receptor	non kinase
iloperidone	5-HT receptor	non kinase
benzethonium	AchR	non kinase
rocuronium	AchR	non kinase



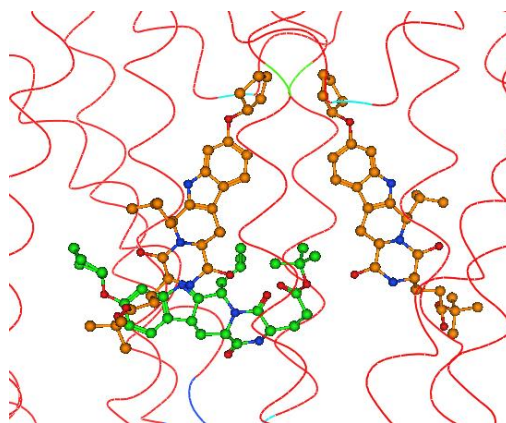
vecuronium	AchR	non kinase
pancuronium	AchR	non kinase
atracurium	AchR	non kinase
neftopidil	adrenergic receptor	non kinase
carvedilol	adrenergic receptor	non kinase
salmeterol	adrenergic receptor	non kinase
cisatracurim	adrenergic receptor	non kinase
DAPT	amyloid A	non kinase
olmesartan	angiotensin	non kinase
AHU-377(sacubitril)	angiotensin receptor	non kinase
valnemulin	anti infection	non kinase
tigecycline	antibiotic	non kinase
cefoperazone	antibiotic	non kinase
natamycin	antibiotic	non kinase
salinomycin	antibiotic	non kinase
sertaconazole	anti-infection	non kinase
fenticonazole	anti-infection	non kinase
bifonazole	anti-infections	non kinase
butoconazole	anti-infections	non kinase
cetylpyridinium	anti-infections	non kinase
vinblastin	autophagy	non kinase
posaconazole	c14 alpha demthylase	non kinase
cinacalcet	calcimemmatic agent	non kinase
manidipine	calcium channel	non kinase
flunarizine	calcium channel	non kinase
econazole	calcium channel/anti-infections	non kinase
torcetrapib	cetp	non kinase
evacetrapib	cetp	non kinase
vx 661	cftr	non kinase
ezetimibe	cholesterol abdorbance	non kinase
montelukast	cis lt1 receptor	non kinase
KPT-330	CRM1	non kinase
cobicistat	cyp3a/cyp45	non kinase
bleomycin	dna synthesis	non kinase
quetiapine	dopamine receptor	non kinase
droperidol	dopamine receptor	non kinase
brexpiprazole	dopamine receptor	non kinase
alvelestat	elastase	non kinase
bosentan	endothelium receptor	non kinase
eprosartan	angiotensin receptor	non kinase
fulvestrant	estrogen progestron receptor	non kinase
estradiol	estrogen/progestron receptor	non kinase
quizartinib	flt3	non kinase
latrepirdine	GluR	non kinase
alarelin	GnRH	non kinase
riocigaut	guanylate acetate	non kinase

ledipasvir	HCV	non kinase
telaprevir	HCV protease	non kinase
asunaprevir	HCV protease	non kinase
simeprevir	HCV protease	non kinase
daclatasvir	HCV protease	non kinase
mocetinostat	HDAC	non kinase
ly2940680	hedgehog	non kinase
fexofenadine	histamine receptor	non kinase
azelastine	histamine receptor	non kinase
loratadine	histamine receptor	non kinase
rupatadine	histamine receptor	non kinase
rilpivirine	HIV	non kinase
maraviroc	HIV	non kinase
nelfinavir	HIV integrase	non kinase
atazanavir	HIV integrase	non kinase
darunavir	HIV protease	non kinase
lopinavir	HIV protease	non kinase
saquinavir	HIV protease	non kinase
ritonavir	HIV protease	non kinase
mevastatin	HMG-CoA reductase	non kinase
fluvastatin	HMG-CoA reductase	non kinase
pitavastatin	HMG-CoA reductase	non kinase
atrovastatin	HMG-CoA reductase	non kinase
ym155	iap	non kinase
birinapant	iap	non kinase
cilengitide	integrin	non kinase
monomethyl auristatin	microtubilin	non kinase
docetaxel	microtubilin	non kinase
paclitaxel	microtubilin	non kinase
tolterodine	muscrinic receptor	non kinase
fenoprofen	NSAID	non kinase
alvimopan	opoid receptor	non kinase
famprofazone	others	non kinase
ranolazine	others	non kinase
betacarotene	others	non kinase
elacridar	others	non kinase
docosanol	others	non kinase
domiphen	others	non kinase
phenylbutazone	others	non kinase
aprepitant	p/nk1	non kinase
birb 796(doramapimod)	p38	non kinase
rg7388	p53	non kinase
BMN 673	PARP	non kinase
ly335979-zosuquidar	P-gp	non kinase
dronedarone	potassium channel	non kinase
dronedarone	potassium channel	non kinase

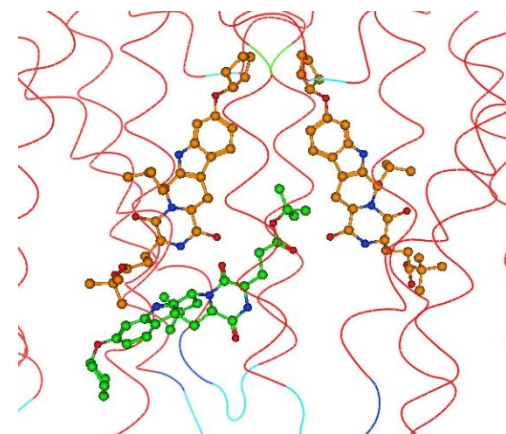
troglitazone	ppar	non kinase
rosiglitazone	ppar	non kinase
carfilzomib	proteosome	non kinase
mln2238	proteosome	non kinase
bortezomib	proteosomes	non kinase
cep-18770	proteosomes	non kinase
losartan	RAAS	non kinase
telmisattan	RAS	non kinase
salirasib	RasGAP	non kinase
aliskiren	renin	non kinase
canagliflozin	SGLT	non kinase
edoxaban	thrombin	non kinase
bibr 1048	thrombin	non kinase
dabigatran	thrombin	non kinase
amsacrine	topoisomerase	non kinase
irinotecan	topoisomerase	non kinase
novobiocin	topoisomerase	non kinase
teniposide	topoisomerase	non kinase
conivaptan	vessopresin receptor	non kinase
retinyl (vitamin A)		
palmitate		non kinase
angiotensin ii	angiotensin receptor	non kinase



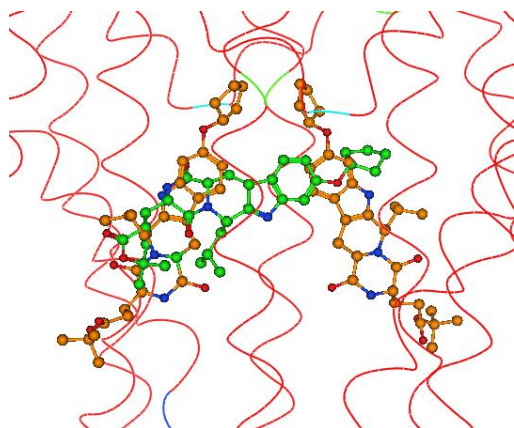
A-AlphaPMI\_alphaHB



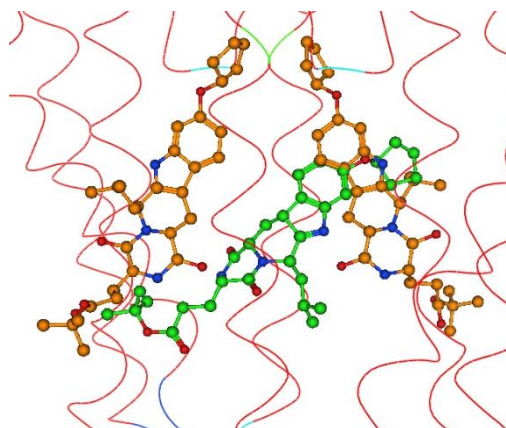
B-AlphaPMI\_londonDG



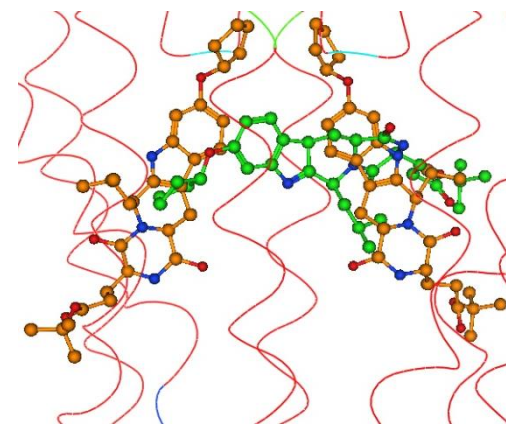
C-MatchTri-londonDG



D-AlphaTri\_londonDG

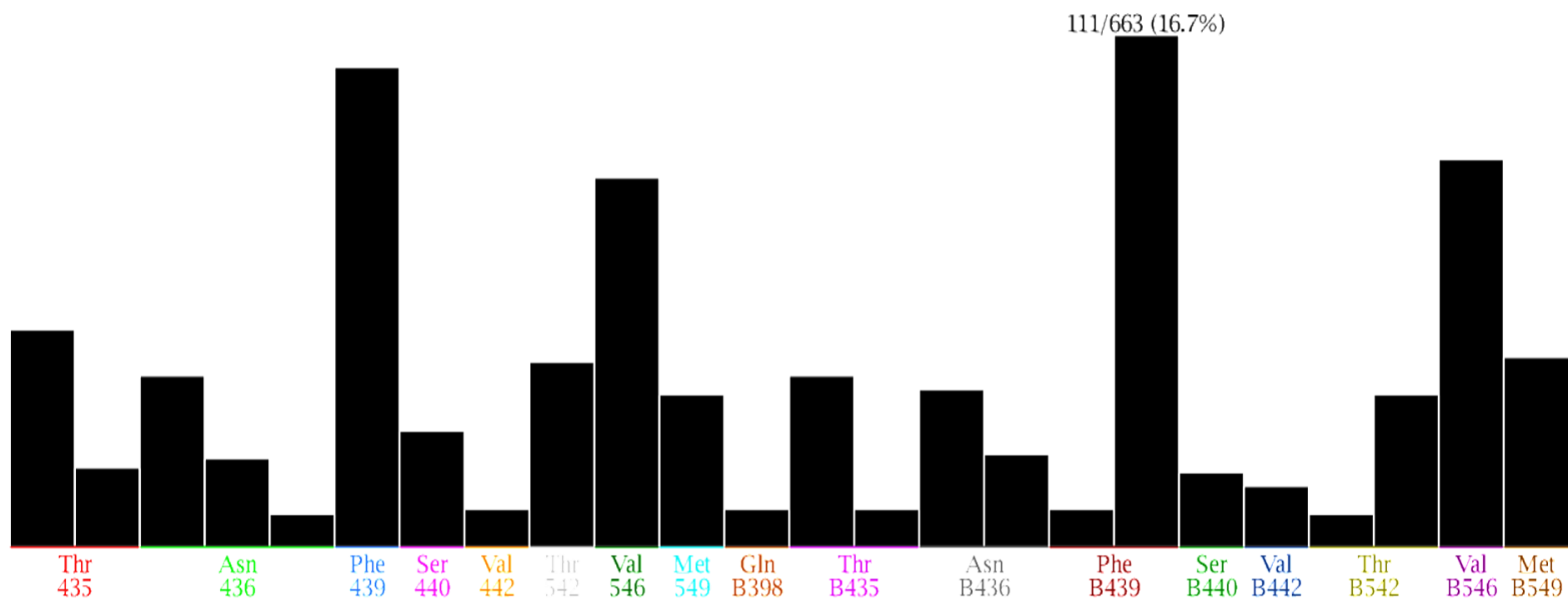


E-ProxyTri\_affinityDG

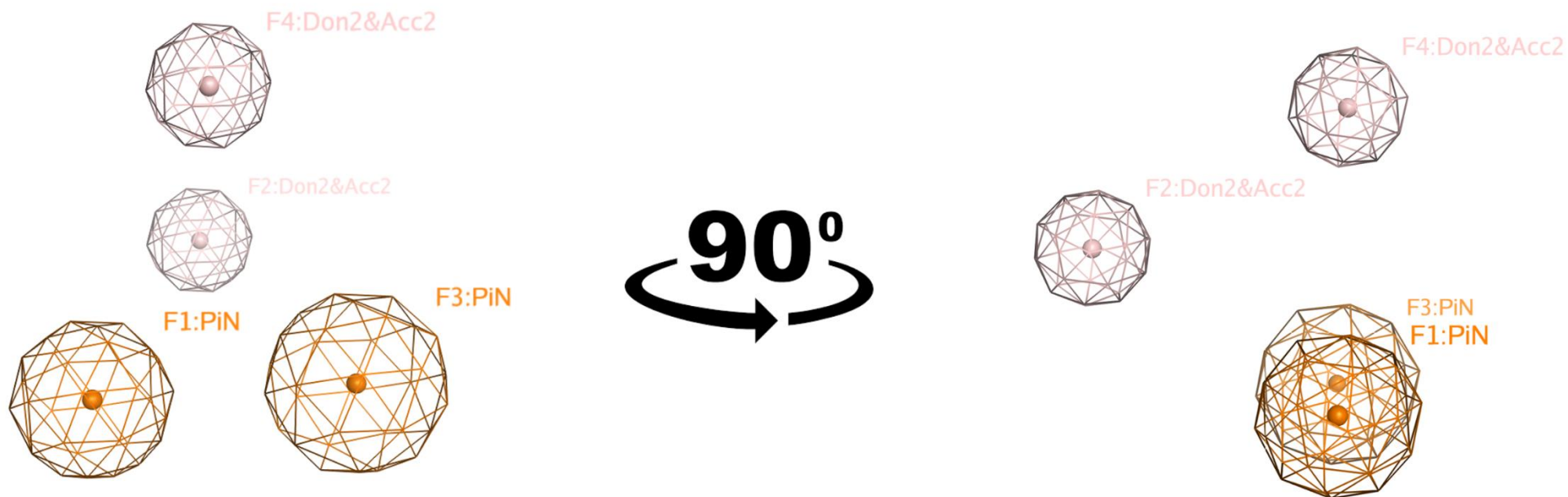


F-ProxyTri\_LondonDG

Appendix Figure 6.1. optimisation of docking protocol for ABCG2 using different placement methods and scoring functions implemented in the MOE



Appendix Figure 6.2. PLIF analysis of structurally diverse compounds with ABCG2.



Appendix Figure 6.3. 3D orientation of topotecan model 1.

Appendix Table 6.1. list of hits from kinase class of drug against ABCG2 from virtual screening studies

Item Name	Target
perifosine	akt
CH5422802	ALK
LDK378	ALK
Barasertib (AZD1152-HQPA)	Aurora Kinase
VX-680	Aurora Kinase
(MK-457, Tozasertib)	
trifluoperazine	autophagy
ABT199	Bcl-2
ABT-263	Bcl-2 family
bosutinib	Bcr-Abl
Nilotinib	Bcr-Abl
(AMN-107)	
Ponatinib (AP24534)	Bcr-Abl
Saracatinib (AZD0530)	Bcr-Abl
PCI-32765 (Ibrutinib)	BTk
Imatinib	c-Kit
(STI571)	
Masitinib	c-Kit
(AB1010)	
cabozantinib	c-MET
crizotinib	c-MET
EMD-1214063	c-MET
amuvatinib	c-RET
regorafenib	c-RET
Deferasirox	CDK
Dinaciclib (SCH727965)	CDK
LEE011	CDK
LY2835219	CDK
Palbociclib (PD0332991)	CDK
Roscovitine (Seliciclib, CYC202)	CDK
Afatinib (BIBW2992)	EGFR
AZD-9291	EGFR
CO-1686	EGFR
(AVL-301)	
Dacomitinib (PF299804, PF299)	EGFR
Erlotinib	EGFR

Gefitinib (ZD1839)	EGFR
icotinib	EGFR
Lapatinib	EGFR
Neratinib (HKI-272)	EGFR
Pelitinib (EKB-569)	EGFR
Dovitinib (TKI-258, CHIR-258)	FGFR
linsitinib	insulin-like GFR
GLPG0634	JAK
LY2784544	JAK
TG101348 (SAR302503)	JAK
ispinesib	ksp
Trametinib (GSK1120212)	MEK1/2
Vemurafenib (PLX4032, RG7204)	MEK1/2
Crenolanib (CP-868596)	PDGFR
Nintedanib (BIBF 1120)	PDGFR
Pazopanib (GW-786034)	PDGFR
BYL-719	PI3K
CAL-101 (Idelalisib)	PI3K
GDC-0941	PI3K
IPI-145 (INK1197)	PI3K
Quercetin	PI3K
SAR245409 (XL765)	PI3K
CX-4945 (Silmitasertib)	PKC
Enzastaurin (LY317615)	PKC
Sotrastaurin (AEB071)	PKC
BI6727 (Volasertib)	PLK
Dabrafenib (GSK2118436)	Raf
R788	STK
-=Axitinib (AG 013736)	VEGFR
Cediranib (AZD217)	VEGFR



Dovitinib	VEGFR
Foretinib (GSK1363089)	VEGFR
Pazopanib	VEGFR
Sunitinib	VEGFR
Tivozanib	VEGFR
(AV-951)	
Vandetanib (ZD6474)	VEGFR

Appendix Table 6.2. list of hits from non kinase class of drug against ABCG2 from virtual screening studies.

<b>drug</b>	<b>target</b>
lurasidone	5HT receptor
ziprasidone	5HT receptor
mosapride	5HT receptor
vilazodone	5HT receptor
ziprasidone	5HT receptor
benzethonium chloride	AchR
gallamine trithiodide	AchR
otilonium bromide	AchR
istradefyline	adenosine receptor
carvidelol	adrenergic receptor
cisatracurium	adrenergic receptor
doxazosin	adrenergic receptor
indacaterol	adrenergic receptor
ivabradine	adrenergic receptor
nebivolol	adrenergic receptor
neftopidil	adrenergic receptor
salmeterol xinofoate	adrenergic receptor
silodosin	adrenergic receptor
vilanterol	adrenergic receptor
indacaterol	adrenergic receptor
fumagilin	aminopeptidase
DAPT	amyloid
MDV3100(enzalutamide	androgen receptor
OMD-201	androgen receptor
AHU-377	angiotensin
eprosartan	angiotensin
angiotensin II	angiotensin II
ascomycin	antibiotic
bacitracin	antibiotic
cefoperazone	antibiotic
ceftriaxone	antibiotic
cetylpyridinium	antibiotic
colistin	antibiotic
difloxacin	antibiotic
enrofloxacin	antibiotic
tebipenempivoxil	antibiotic
tigecyclin	antibiotic
butenafine	antifungal

butoconazole	antifungal
fenticonazole	antifungal
ketoconazole	antifungal
pasoconazole	antifungal
mesoridazine	anti-psycotic
prochlorperazine	anti-psycotic
asunaprevir	antiviral
atazanavir	antiviral
daclastavir	antiviral
danoprevir	antiviral
elvitagavir	antiviral
ladipasvir	antiviral
lopinavir	antiviral
maraviroc	antiviral
MK2048	antiviral
MK-5172	antiviral
nelfinavir	antiviral
nelfinavir	antiviral
ritonavir	antiviral
S/GSK1349572	antiviral
saquinavir	antiviral
semiprevir	antiviral
telaprevir	antiviral
closantal	anti-worms
bendamustine	apoptosis inducer
cinacalcet	calcimemetic agent
azelnidipine	calcium chennel
cinapazide	calcium chennel
flunarizine	calcium chennel
manidipine	calcium chennel
rimanobant	cannabinoid receptor
CP-945598	cannabinoid receptors
evacetrapib	CETP
torcetrapib	CETP
vx-661	cftr
VX809	cftr
ezetamibe	cholesterol absorbance
acemetacin	COX
indomethacin	COX
licofelone	COX
parecoxib	COX
KPT-330	CRM1
plerixafor	CXCR
celecoxib	cyclooxygenase
cobicistat	cyp3a/cyp450

dronedarone	cyp3a/cyp450
pralatrexate	DHFR
bleomycin	DNA synthesis
brexpiprazole	dopamine receptor
droperidol	dopamine receptor
haloperidole	dopamine receptor
GLimipride	DPP4
linagliptin	DPP4
sitagliptin	DPP4
alvelestat	elastase
bosentan	endothelium receptor
VRT752271	ERK
bazedoxifene	estrogen progestron receptor
fulvestrant	estrogen progestron receptor
estradiol cypionate	estrogen/progestron receptor
rivaroxaban	factor Xa
Quizartinib	FLT3
BMS-708163	gamma sceatase
latreperdine	GluR
leuprolide	GnRH
mubritinib	GSK-3
tideglusib	GSK-3
Riocigaut	guanilate cyclase
CUDC-101	HDAC
etinostat	HDAC
mocetinostat	HDAC
parcinostat	HDAC
GDC-0449(vismodegib)	hedgehog
LY2940670	hedgehog
azelastine	histamine receptor
ebastine	histamine receptor
fexofenadine	histamine receptor
mizolastin	histamine receptor
terfenadine	histamine receptor
fluvastatin	HMG- coA reductase
atrovastatin	HMG-CoA reductase
pitavastatin	HMG-CoA reductase
AT13387	HSP
birinapant	IAP
YM155	IAP
Diacerine	il receptor
ispinesib	ksp
casanthranol	laxative
orlistat	lipase inhibitor
docosanol	long chain fatty acid alcohol with multiple uses cosmetic, preliminary antiviral;

batimstat	Matrix metalloproteinase
monomethyl auristatin	microtubule/tubulin
paclitaxel	microtubule/tubulin
GSK2126458	mTOR
thioridazine	multiple effects, anti-psychotic, anti-histamine
etofenamate	NSAID
famprofazone	NSAID
fenoprofen	NSAID
vertoporphin	other
adapalene	others
elacridar	others
BIRB796	p38
ly2228820	p38
MK-4827	PARP
olaprib	PARP
rucaparib	PARP
avanafil	PDE inhibitor
dipyridamole	PDE inhibitor
roflumilast	PDE inhibitor
sildenafil	PDE inhibitor
varafenafil	PDE inhibitor
LY335979 (zosuquidar)	P-gp, structure with tariquidar analogue is present)
clofazimine	phospholipase
IPI-145	PI3K
SAR245409	PI3K
amiodarone	potassium channel
repaglinide	potassium channel
bortezomib	proteasome
carfilzomib	proteasome
CEP-18770	proteasome
MLN2238	proteasome
MLN9708	proteasome
rosuvastatin	RAAS
telmisartan	RAAS
tamibarotene	RAR
fenretinide	RAR/RXR
salirasib	Ras-GAP
aliskiren	renin
glibenclamide	saccharometabolite
canagliflozin	SGLT
dapagliflozin	SGLT
empagliflozin	SGLT
LDE225 di phosphate	smoothed
flecainide	sodium channel
aprepitant	SP/NK receptor
apixaban	thrombin

BIBR-1048	thrombin
dabigatran	thrombin
eltrombopag	thrombopoietin receptor
amsacrine	topoisomerase
camptothecin	topoisomerase
doxorubicin	topoisomerase
epirubicin	topoisomerase
idarubicin	topoisomerase
irinotecan	topoisomerase
moxifloxacin	topoisomerase
lonafarnib	transferase
Tipifarnib	transferase
conivaptan	vessopressin receptor



## Article

# Structure of ABCB1/P-Glycoprotein in the Presence of the CFTR Potentiator Ivacaftor

Alessandro Barbieri <sup>1,2</sup> , Nopnithi Thonghin <sup>1,3</sup> , Talha Shafi <sup>1</sup>, Stephen M. Prince <sup>1</sup>, Richard F. Collins <sup>1</sup> and Robert C. Ford <sup>1,\*</sup>

- <sup>1</sup> School of Biological Sciences, Faculty of Biology Medicine and Health, The University of Manchester, Oxford Road, Manchester M13 9PT, UK; alessandro.barbieri@postgrad.manchester.ac.uk (A.B.); nopnithi@g.swu.ac.th (N.T.); talha.shafi@manchester.ac.uk (T.S.); steve.prince@manchester.ac.uk (S.M.P.); Richard.Collins@manchester.ac.uk (R.F.C.)
- <sup>2</sup> Bioinformatics Institute (BII), Agency for Science, Technology, and Research (A\*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore
- <sup>3</sup> Department of Biology, Faculty of Science, Srinakharinwirot University, 114 Sukhumvit 23, Wattana District, Bangkok 10110, Thailand
- \* Correspondence: robert.ford@manchester.ac.uk

**Abstract:** ABCB1/P-glycoprotein is an ATP binding cassette transporter that is involved in the clearance of xenobiotics, and it affects the disposition of many drugs in the body. Conformational flexibility of the protein within the membrane is an intrinsic part of its mechanism of action, but this has made structural studies challenging. Here, we have studied different conformations of P-glycoprotein simultaneously in the presence of ivacaftor, a known competitive inhibitor. In order to conduct this, we used high contrast cryo-electron microscopy imaging with a Volta phase plate. We associate the presence of ivacaftor with the appearance of an additional density in one of the conformational states detected. The additional density is in the central aqueous cavity and is associated with a wider separation of the two halves of the transporter in the inward-facing state. Conformational changes to the nucleotide-binding domains are also observed and may help to explain the stimulation of ATPase activity that occurs when transported substrate is bound in many ATP binding cassette transporters.

**Keywords:** P-glycoprotein; ABCB1; ABCC7; ABC transporter; ivacaftor; drug binding; Volta phase plate



**Citation:** Barbieri, A.; Thonghin, N.; Shafi, T.; Prince, S.M.; Collins, R.F.; Ford, R.C. Structure of ABCB1/P-Glycoprotein in the Presence of the CFTR Potentiator Ivacaftor. *Membranes* **2021**, *11*, 923. <https://doi.org/10.3390/membranes11120923>

Academic Editor: Olga Vinogradova

Received: 29 October 2021

Accepted: 23 November 2021

Published: 25 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

P-glycoprotein (P-gp) is an ATP-dependent multi-drug transporter of the ATP-binding cassette (ABC) family that is involved in the active efflux of many drugs and xenobiotics from cells [1]. Overexpression or activation of P-gp can lead to multi-drug resistance in cancer cells [2–5]. Drugs and inhibitors are thought to bind to the inward-facing conformation of P-gp [6–9], which is more favored in the post hydrolytic or nucleotide-free conditions [10–12]. The drug is then translocated across the plasma membrane by transition of the protein to the outward-facing state, which is more favored when ATP is bound [12]. The dynamic conformational nature of P-gp (and other ABC family members [13]) has challenged structural studies. Nevertheless, P-gp structures have been obtained by X-ray crystallography [14–18] and cryo-electron microscopy (cryo-EM) [7–9,19–21]. Stabilizing antibodies and ATP hydrolysis-inactivating mutations have been exploited to generate ~4 Å resolution structural data by cryo-EM [7,9,19]. Similarly, trapping the protein in a post-hydrolytic state with vanadate and ATP yielded an ~8 Å resolution structure [21]. However, under these experimental conditions, allosteric changes could be altered or inhibited.

Recent studies have demonstrated that the novel cystic fibrosis (CF) therapeutic ivacaftor is a competitive inhibitor or transported substrate of P-gp [22,23]. This drug is a potentiator of the channel activity of an P-gp homolog, the cystic fibrosis transmembrane

conductance regulator CFTR/ABCC7 [24], and ivacaftor is now widely used in the clinic, both on its own as well as in combination with CFTR-corrector compounds such that roughly 90% of CF patients may now be treatable [25,26]. A prior structure for ivacaftor bound to CFTR showed the binding mode of the drug, which was on the external surface of the membrane-spanning region and within the hydrophobic region of the detergent micelle [27]. The drug itself is highly hydrophobic and was used at a relatively high concentration, but mutagenesis studies and experiments with another potentiator drug with a slightly different structure were undertaken to provide further evidence that the position of the drug in CFTR represented a common binding mode. A further interesting observation in the study was that there was minimal conformational change upon drug binding, and because of the rigid nature of the drug, conformational selection (of the drug) may not occur. Structure–activity relationships (SAR) of ivacaftor homologs with potentiating and inhibiting properties identified the quinoline nitrogen as crucial for activity via hydrogen bonding.  $\pi$ - $\pi$  stacking with nearby residues by the same quinoline group was also found to be important. The planar nature of the drug was also shown to be a significant factor in the SAR study. The hydroxyl group of the hydroxyphenyl moiety is polar with a computed  $pK_a$  of 11 (ChEMBL2010601) and points towards the guanidino group of R933 in the structural model [24,27], implying polar interactions that are too distant (6 Å) for hydrogen bonding.

In this study, we examined the P-gp structure in the presence of ivacaftor. In order to minimize non-specific binding of the hydrophobic drug, we employed a concentration likely to be encountered in vivo [28] and at a sub-stoichiometric ratio with the protein. This approach had the advantage of allowing both drug-bound and drug-free forms of the protein to coexist as a mixture in the protein/drug solution immediately prior to flash freezing and cryo-electron microscopy. We reasoned that if drug binding caused a conformational change in P-gp, we should be able to distinguish different conformations using 3D classification of the particles. If no major change in the overall 3D structure occurred (as was the case for CFTR [27,29]), we expected to obtain a single 3D structure but with fractional occupancy of the drug at its binding site, leading to weak density for ivacaftor in the experimental 3D density map. A third possibility existed: that drug binding would cause conformational changes, but that discrimination of these different conformations would not be possible by the image processing routines. In this case, we would expect a smeared, low-resolution 3D map for the protein with no information about drug binding. In order to minimize the chances of the latter scenario, we employed the Volta phase plate to collect very high signal:noise data with minimal under focus. There are difficulties in the use of the Volta phase plate that restrict its current wide application [29], and there are currently only a few examples of studies resulting in resolution better than 3 Å [29]. Nevertheless, previous studies with 3D maps of resolution similar to those reported here have allowed drugs and small ligands such as sterols and nucleotides to be positioned and their binding interactions modeled within the protein of interest [30–42]. In this report, we aim to compare apo and drug-bound states within a single cryo-EM specimen. If successful, the approach we describe should minimize specimen preparation and biological variation as well as allowing relatively physiological levels of drugs to be employed.

## 2. Materials and Methods

*P. pastoris* harboring the *opti-mdr3* gene was kindly provided by Prof. Ina L. Urbatsch, Texas Tech University [43]. Cell culture was performed using the shake-flask method described by Beaudet and co-workers with minor modifications [44]. Overexpression of the gene product (*abcb1a*) was initiated by the addition of methanol, and the expression was prolonged following an optimized procedure [21]. Cell pellets were collected and stored at  $-80\text{ }^{\circ}\text{C}$ .

Cell rupture was conducted using glass bead beating, and microsomal membranes were prepared as described in (4). Microsomal membranes were diluted to 2.5 mg/mL



prior to solubilization in detergent-containing buffer (50 mM Tris pH 8.0, 10% (*v/v*) glycerol, 50 mM NaCl, 1 mM 2-mercaptoethanol and 2% (*w/v*) DDM). Solubilized materials were processed through multi-step protein purification, described previously [21,22]. Purified protein was concentrated to 5–10 mg/mL using a Vivaspin concentrator with 100-kDa cut-off, flash-frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ .

Drug binding and P-gp thermostability were assayed in 3 ways using an UNCLE (Unchained laboratories) instrument: concentrated P-gp protein (1  $\mu\text{g}$  per sample) was diluted into 10  $\mu\text{L}$  of buffer (100 mM Tris-HCl pH 8, 150 mM NaCl, 10% glycerol, 0.1% DDM and 0.02% CHS) containing different concentrations of ivacaftor. The experiments were also repeated with the further addition of 100 ng of CPM dye (7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin), which acts as a reporter of solvent-exposed cysteine residues [22,45–47]. The mixtures were loaded into capillaries at  $4^{\circ}\text{C}$ , and the latter were inserted into pre-chilled 16 well UNCLE capillary cassettes. Fluorescence emission spectra were recorded between 200 nm and 700 nm with an excitation wavelength of 266 nm. Static light scattering was recorded simultaneously using the 266 nm laser source. Data were gathered between  $16^{\circ}\text{C}$  and  $90^{\circ}\text{C}$  with a stepwise increase in the temperature ( $2^{\circ}\text{C}$  increments) followed by the recording of spectra for all the samples. The total heating run required approximately 70 min. Raw data were exported and then analyzed for tryptophan fluorescence, static light scattering, and CPM fluorescence using the GraphPad Prism software package.

P-gp was initially buffer-exchanged into a detergent/glycerol-free buffer. The protein was diluted to 1.1 mg/mL and incubated with 2  $\mu\text{M}$  Ivacaftor (stock: 100  $\mu\text{M}$  in DMSO) for 30 min on ice. Quantifoil 200 or 400 Au grids with a 1.3-micron spacing pattern were pre-treated prior to protein deposition, as described previously [21]. Briefly, grids were surface-cleansed by multiple chloroform washes to eliminate hydrophobic residues, then glow-discharged. An FEI Vitrobot MkIV was employed to facilitate sample vitrification where 3  $\mu\text{L}$  of protein was gently loaded onto the middle of the grid, immediately blotted for 6 s, and flash-frozen in liquid ethane. Grids were assessed for specimen quality (i.e., ice thickness, particle distribution, etc.) using a Polara G30 before shipping to the eBIC UK National facility for high-resolution data acquisition on an FEI Titan Krios G2 microscope. Data were recorded at 300 kV using a 20 mV energy filter and a Gatan K2 electron detector. Images were acquired via the Volta phase plate [48] with a phase shift range between 0.15 and 0.8 radian and fixed defocus at  $-0.5\ \mu\text{m}$ . A total of  $\sim 64\ \text{e}^{-}\ \text{\AA}^{-2}$  were used over 40 frames spanning 10 s of exposure, and early and late frames were discarded. The magnification was calibrated at  $1.043\ \text{\AA}/\text{pixel}$ . A total of 2436 movie stacks were collected, and beam-induced image shift was eliminated using MotionCorr2 [49].

The data processing of the 2436 images was entirely performed with *cisTEM* [50]. A total of 1030 images were eliminated for possessing poor CTF fitting profiles and exceeding defocus values. The final set of 1406 images were divided into 2 similarly-populated subsets (696i and 710i) to independently assess the reproducibility of the 2D classification procedure. Ab-initio 3D reconstruction was performed employing  $\sim 50,000$  particles derived from several 2D classification runs of the 696i subsets. For each of the 5 classes that were generated ab-initio, a PDB model was built using Molecular Dynamics Flexible Fitting (MDFF) with NAMD 2.12, as detailed previously [21]. These models were used as a measure for comparing NBD separation in the five classes. Ab-initio 3D maps were then low pass filtered to  $10\ \text{\AA}$  and employed as starting templates for 3D refinement using the full final set of 104,000 selected particles. After full 3D refinement of the maps, real-space refinement of atomic models against the maps was performed within the PHENIX package [51], with the MDFF-derived models used as the initial inputs. Half-maps and Fourier shell correlation curves were generated employing the *cisTEM* and *calculate\_fsc* routine. Local resolution was also estimated using the program ResMap [52]. Rigid-docking of ivacaftor was performed with the fit-in-map routine of UCSF Chimera [53] with a  $5\ \text{\AA}$ -resolution simulated map of ivacaftor atoms at  $0.252\ (7 \times \text{SD})$  and with optimization for overlap with the experimental (additional) density in the P-gp central cavity. Fitting was subsequently

minimized in Schrodinger's Maestro (Schrödinger Release 2021-2: Maestro, Schrödinger, LLC, New York, NY, USA, 2021), concomitantly allowing the rotamer adjustments of neighboring residues. The final result was verified to maintain the occupancy of the cryo-EM density. Nucleotide-binding pocket volumes of NBD1 and NBD2 were estimated by summation of all voxels with lower than mean density values that were within 5 Å of a positioned nucleotide. This was carried out with the *matchmaker*, *color zone/split map*, and *calculate volume and area* functions of Chimera [52] and using the NBDs in the 6q81 P-gp atomic model for positioning [21].

### 3. Results

#### 3.1. Ivacaftor Binding Affinity of Murine P-gp

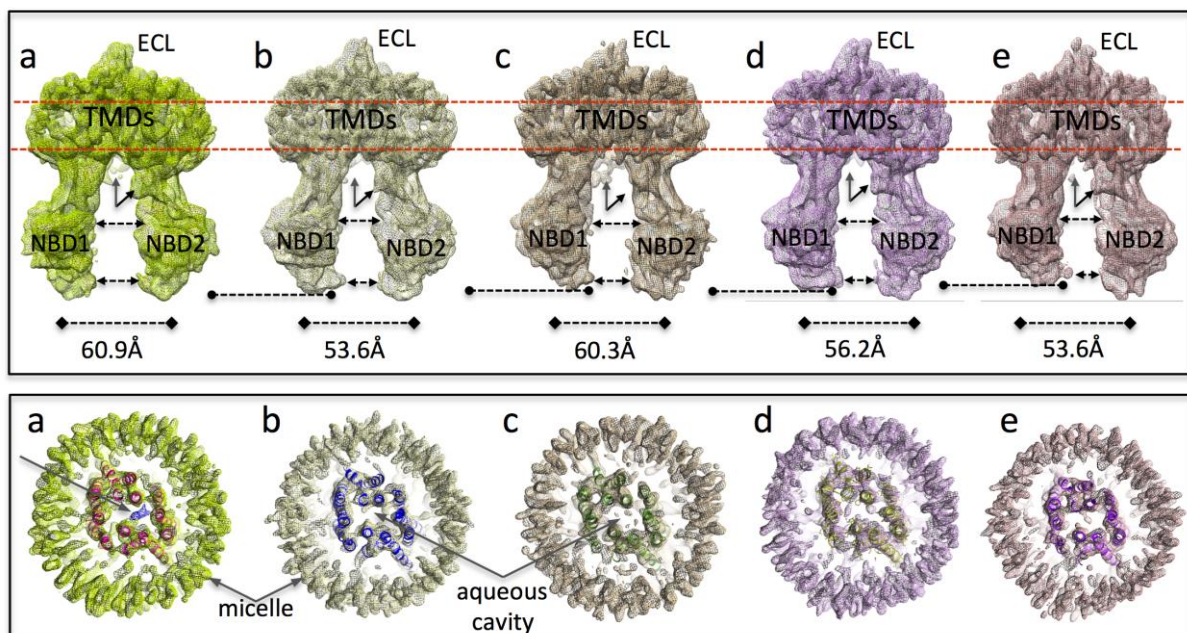
Ivacaftor has been identified as a competitive inhibitor of Hoechst 33,342 and digoxin transport by human P-gp [22,23]. Ivacaftor stimulated human P-gp ATPase activity eight-fold, suggesting it could also be considered as a transported allocrite, and its dissociation constant ( $K_D$ ) was estimated at 0.3  $\mu$ M using a fluorescent dye transport assay with the purified protein [22]. Similarly, a half-maximal inhibitory concentration ( $IC_{50}$ ) of 0.2  $\mu$ M was estimated using an in-vivo assay [23]. We tested ivacaftor binding to murine P-gp using various methods: Static light scattering and tryptophan fluorescence changes yielded  $K_D$  values of 0.2  $\mu$ M (static light scattering) and  $\cong 1$   $\mu$ M (tryptophan fluorescence) (Supplementary Figure S1a,b). The binding of a sulfhydryl-reactive fluorescent dye (7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methyl-coumarin, CPM) to solvent-exposed P-gp cysteine residues detected an unfolding transition between 30–50 °C that was sensitive to ivacaftor with a  $K_D$  estimated at 0.2  $\mu$ M (Supplementary Figure S1c). Hence, we concluded that the affinity for ivacaftor of murine P-gp was similar to human P-gp. Moreover, the detection of light scattering changes upon ivacaftor binding implied conformational changes in P-gp (e.g., due to an induced fit). Cryo-EM studies were subsequently performed with a concentration of ivacaftor (2  $\mu$ M) that was well above the  $K_D$ , although at this level it was still at a sub-stoichiometric ratio versus protein (estimated at 1.1 mg/mL, or roughly 8  $\mu$ M using a molecular mass of 150kDa for P-gp).

#### 3.2. Cryo-EM of P-gp in the Presence of Ivacaftor

Images (2436) were recorded with the Volta phase plate, and after initial rejection of poor quality images by eye, the contrast transfer function (CTF) of each was estimated. A total of 1537 images gave CTF fits that were judged to be reliable whilst 868 were rejected at this stage. A second criterion was applied to eliminate all images affected by excessive defocus values, and a total of 131 further images were eliminated, with a final set of 1406 images (Supplementary Figure S2). As expected, the P-gp particles were readily identifiable in the images in the inward-facing conformation (Supplementary Figure S3), and automated particle picking was performed. Reference-free classification of the projections of the resulting picked particles allowed a further pruning of the dataset by deleting any non-P-gp contaminants as well as small P-gp aggregates (Supplementary Figure S4). Several iterations of 2D classification were completed to remove bad particles, and this was judged to be complete when consistent 2D classes were obtained. The reproducibility of the 2D classification was then assessed by splitting the dataset into two roughly equally populated sub-sets, which individually showed the same outcome (Supplementary Figure S2).

*Ab-initio* 3D classes of the final P-gp dataset were generated from a subset of only ~50,000 particles, and these are shown in Supplementary Figure S5. At this early stage, the 3D classes were relatively low resolution (~7 to 8 Å, as estimated by Resmap [54]). Nevertheless, the two procedures clearly identified only inward-facing conformations, even when five classes were allowed. Sets of particles displaying a wider separation of the NBDs (classes *a* and *c*) were also picked out by the image processing package. For each of the five low-resolution *ab-initio* maps, a PDB model was derived using Molecular Dynamics Flexible Fitting (see Methods), and results were compared. Utilizing the center of mass of the NBDs as a reference, classes *b*, *d*, and *e* showed a narrower separation of

the NBDs (from 56 Å to 53 Å), while both classes *a* and *c* were found to have a wider distance (60 Å). A moderate low-pass filter was applied to the *ab-initio* maps to avoid over-fitting effects. CTF parameters were also further refined at this stage with per-micrograph CTF fits using the dedicated *cisTEM* functionality for this. Refinement of the full dataset (104,000 particles) starting with each of the five *cisTEM* *ab-initio* 3D classes was carried out with a combination of auto- and manual-refinement, and an atomic model was generated for each resulting refined 3D map using the PHENIX *real space refinement* routine [51]. The 3D maps, compared at a density threshold yielding a map volume of 153,000 Å<sup>3</sup> are displayed in Figure 1.



**Figure 1.** Differentiation of the five 3D maps (a–e): Maps are displayed at a density level enclosing a volume of 153,000 Å<sup>3</sup>. Upper panel: (1) NBD1 rotates and hinges upwards towards the TMDs; with this increasing from maps (a) through to (e), (dashed line with circles indicates the lower edge of NBD1). (2) NBD1–NBD2 separation, (dashed lines with arrowheads) is greatest for map (a,c), least for map (e) and intermediate for maps (b,d). (3) Density for the N-terminus (rearwards, vertical arrow) and the end of the NBD1–NBD2 linker (slanted arrow, front) is present in map (a), whilst other maps lack the N-terminus density (maps (b,d)) or lack the linker density (c). (4) The density for the extracellular loop (ECL) between TM helices 1 and 2 is weaker for maps (c,e). Lower panel: slices through the transmembrane regions of the maps (as indicated by the red dashed lines in the upper panel). Real-space refined atomic models that were generated are also displayed using ribbon representation in contrasting colors. The diameter of the micelles in maps (a,c) are larger; map (e) has the smallest micelle with some distortion from circular. Map (a) contains a central density (blue mesh, arrow) not accounted for by the real-space refined model. Other maps have only small features in the aqueous central cavity. Map (e) shows the most compact transmembrane organization.

The resolution of each map was assessed by splitting each data set in half and separately refining each half dataset. The correspondence between half-maps was analyzed in reciprocal space using Fourier shell correlation (FSC) as well as using the Resmap routine, which calculates local resolution in the maps [54,55]. The resolution assessments as well as other data pertaining to the five refined maps and the five atomic models, are summarised in Table 1. Each 3D map could be generated from fewer particles than in prior studies of P-gp by cryo-EM [7,19,21] due to the high contrast of individual particle projections produced with the Volta phase plate. However, the resolution achieved was estimated at 4–6 Å for all the maps (Table 1, Supplementary Figures S6–S8). Map resolution was not correlated with particle numbers, consistent with the idea that, for structural studies of unstabilized P-gp, intrinsic protein flexibility is likely to be a major factor in determining the resolution achievable. Interestingly, map *a* displayed the lowest resolution of all the

maps in the various tests employed, although the overall map-to-model correlation was better (Table 1, Supplementary Figure S8). The Resmap analysis of map *a*, implied that variation between the half-maps was mainly confined to the periphery of the NBDs and the micelle (Supplementary Figure S6). Conversely, map *e*, with the closest approach of the NBDs, was estimated to have the highest resolution by the Resmap algorithm [54] and by Fourier shell correlation [55] (Supplementary Figure S7). In general, the map-to-model correlation was lower for the NBDs compared to the TMDs in all maps (Supplementary Figure S8).

**Table 1.** Summary of all the data for reciprocal space resolution assessment of the five maps as well as the real-space correlation coefficient (CC) between the atomic model and the relevant map.

Parameter.	Map <i>a</i>	Map <i>b</i>	Map <i>c</i>	Map <i>d</i>	Map <i>e</i>
Global Resolution (Å, FSC = 0.143) from unmasked half maps (CisTEM)	5.4	4.3	4.3	4.2	4.2
Global Resolution (Å, FSC = 0.143) from map to model fit (Phenix)	4.6	4.4	4.5	4.4	4.4
Resolution (Å) from Resmap with masked half maps (Mean, Mode)	5.0, 4.0	4.5, 4.0	4.6, 4.0	4.5, 4.0	4.4, 4.0
CC map to model fit (mean from 1182 residues, Phenix)	0.73	0.67	0.63	0.69	0.65

### 3.3. Interpretation of the Differences between the 3D Maps

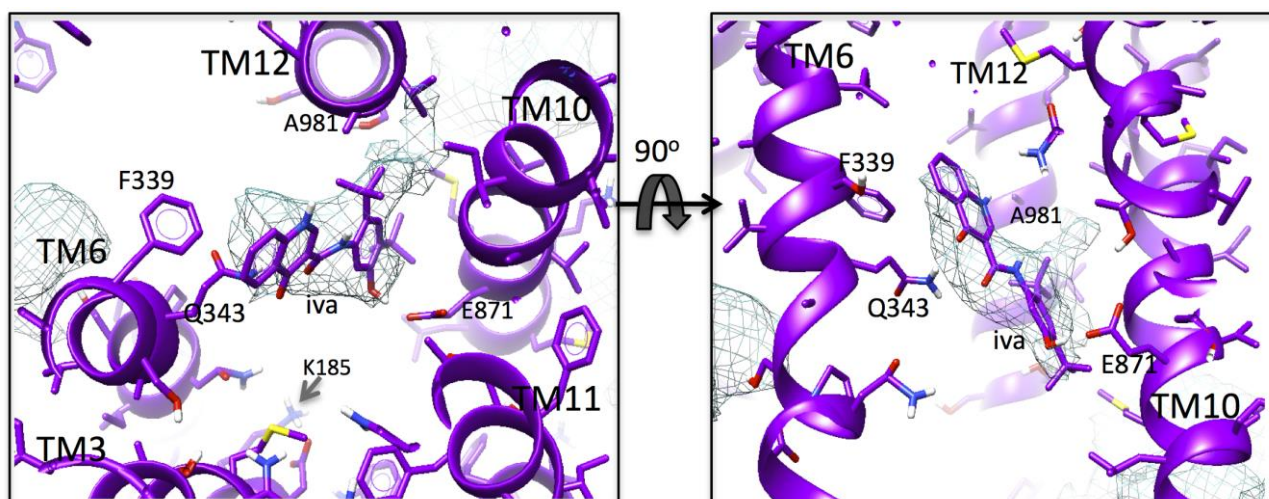
The various 3D maps showed the typical hinging of the two halves of the molecule, giving differences in NBD separation, but it was apparent that NBD1 flexing relative to the other P-gp domains was also being differentiated by the 3D classification algorithm: NBD1 appeared to progressively rotate upwards towards the TMDs when one examines maps *a* through *e*; (Figure 1, upper panel). Separation of NBD1 from NBD2 was greatest for maps *a* and *c*; was narrower in maps *b* and *d*; whilst map *e* had the closest approach of the NBDs (Figure 1, upper panel). As previously reported [21], densities assigned to the N-terminus and to the end of the NBD1-NBD2 linker on the opposite side of the molecule can be observed in the maps, especially map *a* (Figure 1 upper panel). Maps *b* and *d* showed additional density for the linker but lacked density for the N-terminal region. In contrast, map *c* showed the latter but appeared to lack the linker density. The density for the extracellular loop (ECL) between TM helices 1 and 2 was weaker for maps *c* and *e*, perhaps representing increased disorder in this loop, which may be core glycosylated in the expression system used. These symmetry-breaking features of P-gp likely aid the image processing via discrimination of projections of alternative 180°-rotated orientations of particles (Supplementary Figure S9).

The transmembrane regions (Figure 1, lower panel) also showed some subtle differentiation between the five 3D maps. The diameter of the micelles in maps *a* and *c* are noticeably larger; whilst map *e* has the smallest micelle, displaying some distortion from a circular annulus. These differences may be a reflection of the degree of separation of the two halves of the protein, which is greatest for maps *a* and *c* and least for map *e*. The latter map shows the most compaction of the transmembrane  $\alpha$ -helices. The transmembrane regions of all five maps (Figure 1, lower panel) displayed the large central aqueous cavity that connects to the cytoplasm as well as to the lipid bilayer (via the two lateral gaps formed between TM helices 4 and 6 and between 10 and 12). These lateral gaps are a common feature of type IV ABC transporters such as Sav1866, MsbA, and P-gp [56]. In P-gp the central aqueous cavity has been found to house the binding sites for transported drugs and inhibitors [7,14,16]. We, therefore, closely examined this cavity for evidence of ivacaftor binding in the five maps. The criteria we used were that: (i) any features due to ivacaftor should be visible at the same density level used for the protein (ii) that at this density level, the feature(s) should be roughly rod-shaped, and (iii) have a continuous volume equivalent in size to that expected for the drug. Finally, (iv) once fitted, the ivacaftor molecule should not have serious clashes with nearby residues in a fitted P-gp atomic model (i.e., none that could not be remedied with rotamer adjustment). Only map *a* contained a single feature



that satisfied all these criteria (blue mesh, arrow, Figure 1, lower panel). All the other maps did show small features in the aqueous central cavity (Figure 1 lower panel), especially at lower density thresholds, but none of these met the first three criteria.

The additional density in the aqueous cavity of map *a* has a rocket-shape with two fin-like protrusions at one end. At the opposite end to the protrusions, the density is close to TM6, and in the fitted atomic model, it was close to F339 and Q343 (Figure 2). In accordance with the last criterion, a single ivacaftor molecule could be fitted into this density, and the overall rocket shape favored a fit with the 2,4 *tert*-butyl groups of the ivacaftor molecule corresponding to the two ‘fins’ of the rocket-shaped density (CC = 0.960 vs. CC = 0.938 for the 180° rotated version). The density is surrounded by residues contributed by TM helices 6, 10, and 12, and hence is positioned more in one lobe of the aqueous cavity (Figure 1). Residues F339 and Q343 within TM6 in the refined atomic model are close enough to form  $\pi$ - $\pi$  (F339) and H-bonding (Q343) interactions with the quinoline group of the modeled ivacaftor molecule (Figure 2), and the phenolic hydroxyl group of the drug could also form H-bonding interactions with E871 within TM10 in this model. Small residues G868 and G985 in TM helices 10 and 12, respectively, allowed the fitting of the bulky *tert*-butyl groups of ivacaftor without steric clashes with the protein. Although the cavity is aqueous, I864, M872, L875 in TM10, and A981 in TM12 form a hydrophobic patch on its surface that appears to be compatible with the positioning of the hydrophobic drug (Figure 2). K185 on the cytoplasmic side of TM3 in the aqueous cavity is within ~10 Å of the additional density. This is somewhat reminiscent of R933 in human CFTR [27].

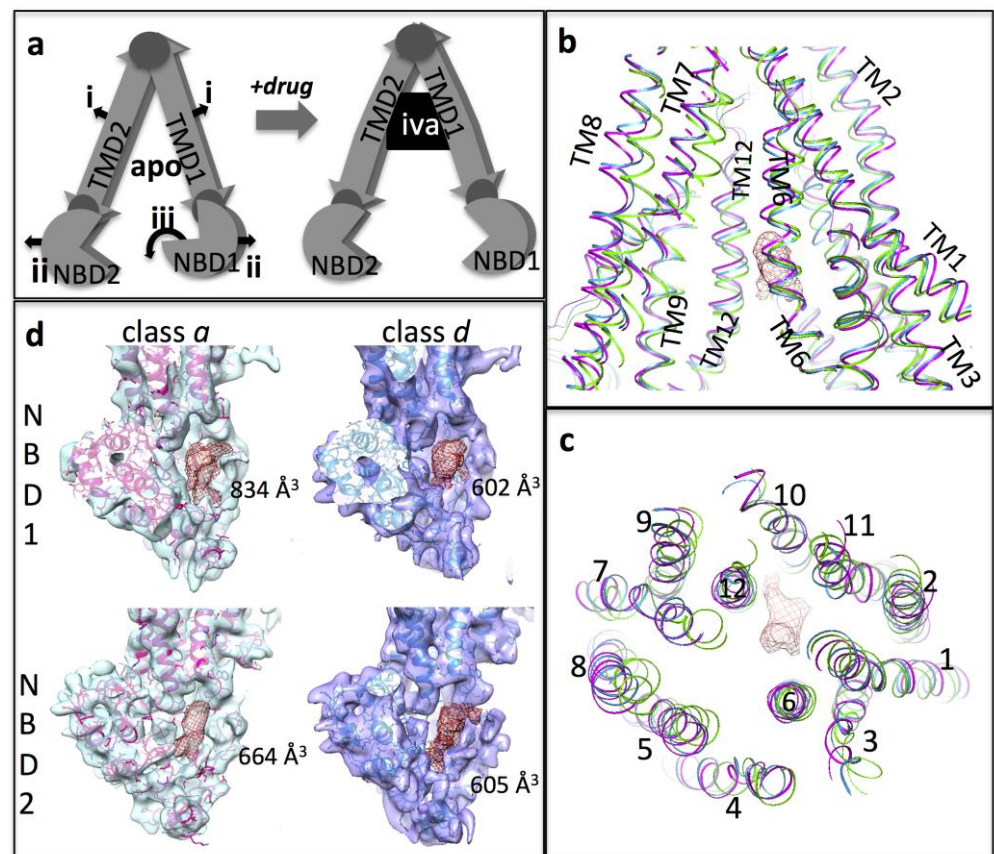


**Figure 2.** Additional density: Magnified, orthogonal views of additional density not accounted for by the fitted model in map *a* (blue mesh). The drug (*iva*) is modeled into the additional central density. The density is surrounded by residues in TM helices 6, 10, and 12 in the real-space refined atomic model. The 2,4 *tert*-butyl groups at one end of the ivacaftor molecule favored its orientation as shown within the rocket-shaped density. In this model, F339 and Q343 within TM6 could form  $\pi$ - $\pi$  and H-bonding interactions with the ivacaftor quinoline group. Small (glycine) residues at 868 and 985 in TM helices 10 and 12 may be important for accommodating the bulky *tert*-butyl groups at the other end of the drug, and A981 may also satisfy hydrophobicity requirements for these groups. Similarly, H-bonding with E871 by the phenolic hydroxyl group of ivacaftor may occur in this fitting. K185 in TM3 extends up into the central cavity to within 10 Å of this group.

### 3.4. Conformational Changes Associated with the Additional Density

Assuming the additional density in map *a* is due to drug, then its presence is associated with a wider separation of the two halves of P-gp, similar to observations for the co-crystal structures of murine P-glycoprotein with the marine pollutant and P-gp inhibitor BDE-100 [15,16,57]. Similarly, widening of the structure of the TmrAB heterodimeric peptide exporter was associated with substrate-binding [13]. We, therefore, examined whether other local changes in individual domains could be identified as being specific to map *a* versus the other four maps. In this line of conjecture, we assumed that maps *b* to *e*

were representative of multiple conformations of the *apo*-state. When map *a*'s atomic model was compared to the one most closely matching it: map *d* model (rmsd = 1.17 Å for 1018 atom pairs within 2 Å, 1.46 Å overall 1182 atom pairs), a local bulging outwards of the transmembrane helices were apparent (see Figure 3), similar to that observed for TmrAB [13]. When we compared map *a*'s model to that of map *e* (which was the most different—rmsd 1.40 Å over 519 atom pairs within 2 Å, 2.64 Å overall 1182 atom pairs), the outwards bulging of the TM helices upon ivacaftor accommodation was more noticeable. In both cases, TM helices 1, 2, 7, and 8 move most whilst helices 6 and 12 move little (Figure 3b,c), and the bulge of the TMDs resulted in a wider separation of the NBDs. As previously mentioned, NBD1 also showed a downward rotation away from the TMDs (Figure 3).



**Figure 3.** Allostery: (a) Cartoon representation of conformational changes discussed with regard to the presence of ivacaftor with P-gp. Left shows the assumed apo-state and indicates the movements of the TMDs and NBDs associated with the drug-bound state (right): (i) the TMD helices bow outwards and the central cavity enlarges. (ii) The NBDs become more separated. (iii) NBD1 rotates downwards. The small darker circles indicate hinge points between domains. (b,c) outward bulge of the TMDs in the atomic model from map a (pink ribbon) vs. models from map e (green) and map d (blue). The additional density associated with ivacaftor in map a is shown (red mesh). Panel b is a section viewed along the membrane plane; panel c is a section viewed from the extracellular surface. (d) NBD1 and NBD2 in experimental maps a and d viewed from between the NBDs (blue and purple semi-transparent surfaces). The real-space refined atomic models for P-gp fitted to the maps are in red and blue, respectively. The nucleotide-binding cavity is indicated with the red mesh and its volume is given in cubic Ångstrom.

The net result of this movement of the NBDs in the atomic models was a downward (cytoplasmic) shift of the Walker A and B regions relative to intracellular loops 1 and 4 that connect NBD1 from the TMD, and in map *a* this is manifested as an opening up of the nucleotide-binding site in NBD1 (Figure 3d). Similar changes but with a smaller

magnitude, were also observed for NBD2 in the maps and atomic models. Caution must be exercised when extrapolating from atomic models that have been fitted to worse-than-atomic resolution maps, but if these conformational shifts in the NBDs are valid, then they may be reflective of allosteric changes that lead to increased ATPase activity of P-gp upon binding of a transport substrate.

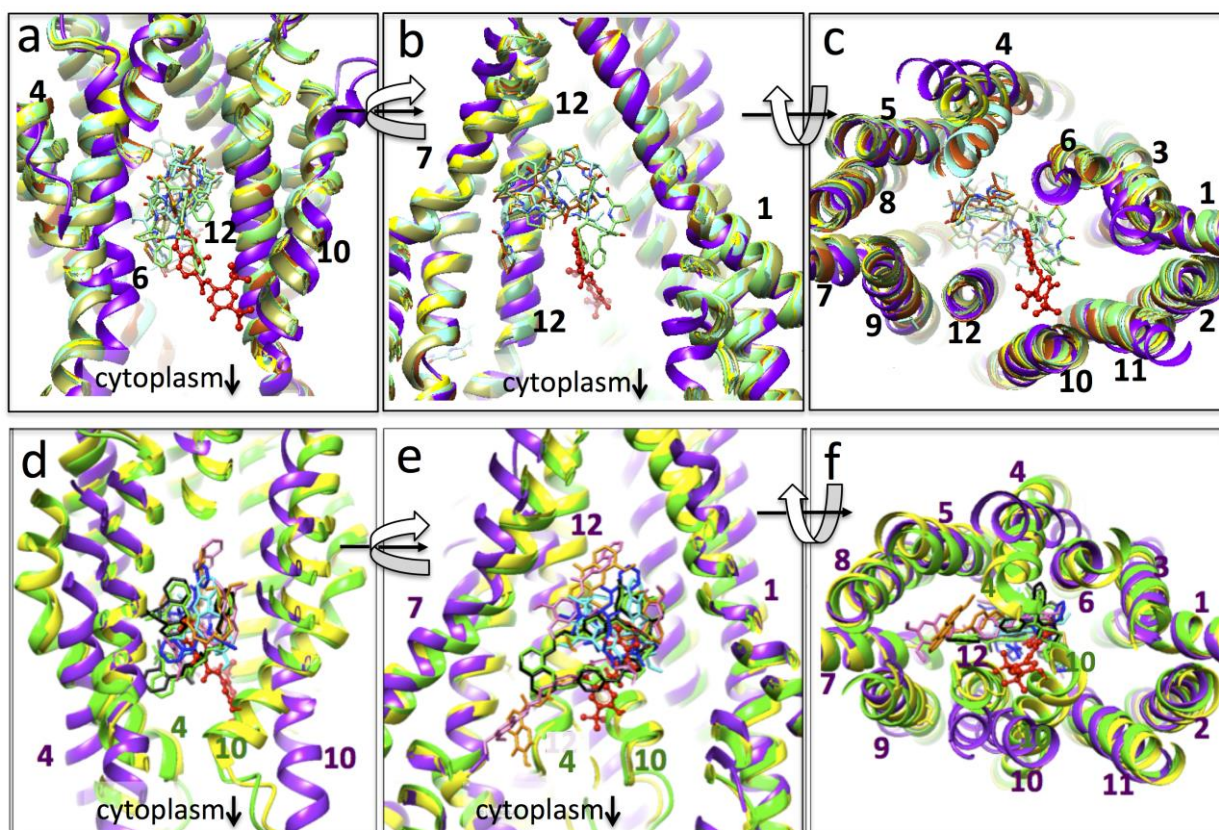
#### 4. Discussion

Structural data for 3D conformations of P-gp were obtained with the Volta phase plate; and with this methodology the resolution of the 3D information obtained was around 6–4 Å. Whether these resolution limitations can be overcome with a larger dataset remains to be established. The use of sub-stoichiometric levels of ligand may be a suitable approach in future cryo-EM studies, provided that ligand–protein complexes can be distinguished from ligand-free protein on the basis of different conformational status and/or by a sufficiently large additional density due to the ligand. Clearly, atomic details of ligand binding mode will depend on close to atomic resolution data, but the overall location of the ligand and some modeling of likely binding modes will be possible at intermediate resolution, as described here and in other studies for various other proteins and ligands [30–42].

Drug and inhibitor binding locations in P-gp have previously been studied for both mouse and human versions of the protein [8,9,15,16,58] as well as for a human/mouse chimeric construct [7]. For mouse P-gp, published structural data from X-ray crystallography is available for cyclic peptide inhibitors and for a marine pollutant BDE-100 that is also thought to act as an inhibitor of the protein [15,16]. For human P-gp, cryo-EM data are available for well-characterized substrates such as taxol [8,9] and vincristine [8] and also for some inhibitors such as elacridar, zosuquidar, and tariquidar [8]. Ivacaftor has been shown in-vitro to stimulate P-gp ATPase activity and to reduce transport by P-gp of Hoechst 33342 [22]. In humans, ivacaftor administration caused increased blood digoxin levels [23], implying inhibition of P-gp. These studies suggest that ivacaftor can compete with Hoechst 33342 transport and digoxin transport by P-gp, hence ivacaftor could be transported by P-gp. Alternatively, the data could be consistent with ivacaftor acting as a non-competitive or uncompetitive inhibitor of the protein's transport (but not ATPase) function. The Hoechst 33342 binding site (H-site) has been modeled to be asymmetrically-located in P-gp [59,60], but this position is in the opposite lobe of the internal cavity compared to the modeled site for digoxin (D-site) [60], and the density associated with ivacaftor described here. Hence from the data and models, one could propose that ivacaftor may compete for both digoxin binding and transport. Whereas for Hoechst 33342, transport, but not binding, would be affected by ivacaftor presence. These suggestions remain to be tested experimentally.

A comparison of murine P-gp models both with and without bound compounds is shown in Figure 4, panels a–c. Their conformations are similar to the murine P-gp model for map *a* that is described here, with root mean squared deviation (rmsd) between the main chain C $\alpha$  atoms of the models of around 4 Å over 1179 comparisons (i.e., not including disordered linker and N- and C-termini atoms). When these various atomic models are superimposed (Figure 4, upper panels), the positions of the inhibitors and drug (ivacaftor) are all found within the central aqueous cavity. However, the BDE-100 and cyclic peptide inhibitors appear to be located further towards the extracellular side of the membrane.





**Figure 4.** Location of drugs and inhibitors within the P-gp central binding cavity: (a–c): Murine P-gp models. Purple ribbon is the atomic model for map *a* in this paper (ivacaftor- red); yellow, PDBID 4q9h- apo-state; cyan, 4q9j- with QZ-Val; orange, 4q9i-QZ-Ala; taupe, 4q9k-QZ-Leu; green, 4q9l-QZ-Phe; turquoise, 4xwk-BDE-100. The modeled position of ivacaftor sits further towards the cytoplasm than the various inhibitors shown. Trans-membrane helices are numbered according to their position in the P-gp sequence. (a,b) are orthogonal views along the membrane plane; (c) is the view from the cytoplasm. Curved arrows indicate the 90° rotations involved. (d–f): Human and human/mouse chimera P-gp: no drug (6fn4, yellow ribbon); with zosuquidar (6fn1, green ribbon, green drugs, and black drugs-7a6f); tariquidar (7a6e, pink drugs); vincristine (7a69, cyan drug); taxol (6qex, blue drug); elacridar (7a6c, orange drugs). Wider separation of the murine NBDs necessitated alignment of the N- and C-terminal arms of the model (this paper, purple ribbon, red drug) with the human P-gp structures. Where major differences exist between the paths of transmembrane helices in human and murine models, they are numbered in both purple and green font.

A comparison of the mouse/ivacaftor P-gp model with human and human/mouse chimera P-gp models in the presence and absence of other small molecules is shown in Figure 4, panels d–f. Separate alignment of the two ‘wings’ of the inward-facing conformation of the mouse/ivacaftor structure was required for this comparison because the human and chimera P-gp models all show a closer approach of the NBDs. In these comparisons, both substrates and inhibitors co-localize to the same approximate position within the central cavity. As discussed by Nosol et al. [8], whilst the P-gp inhibitors display two separate molecules in the cavity, the transported substrates (and ivacaftor) display only one molecule in the cavity. Density associated with ivacaftor in map *a* is consistent with a single molecule, but again, in this comparison, it would be located closer to the cytoplasmic side of the membrane than the other compounds, with the exception of one of the elacridar molecules (orange stick representation, Figure 4).

Taken together, the data for ivacaftor is equivocal: the single density observed would imply that ivacaftor shares P-gp substrate characteristics, whilst its overall location in one lobe and over towards the cytoplasmic leaflet of the bilayer is suggestive of binding characteristics of an inhibitor according to the work of Nosol et al. [8]. It seems possible that the low concentration/stoichiometry of the drug employed in this study may result in a



single bound ivacaftor molecule at a higher affinity site, even if it were an inhibitor with two potential sites. However, it should be noted that not all inhibitors appear to bind in the same position: The cyclic peptide inhibitors and BDE-100 studied with murine P-gp occupied a location different from the elacridar, tariquidar, and zosuquidar inhibitors [14,15]. In addition, of note is that in the human P-gp models, transmembrane helices 4 and 10, were kinked inwards and appear to partially close off the internal cavity on the cytoplasmic side. This behavior has not been observed thus far in the murine P-gp models (Figure 4). The closest approach of the two halves of the human and chimera P-gp atomic models is also noteworthy, although this may result from different solubilization methods employed (lipids in the nanodiscs for human P-gp; detergent in micelles for murine and chimera P-gp). Whether these differences between human and murine P-gp structures represent interspecies variation, or instead alternative conformational states in the transport cycle remains to be established. At first sight, it would seem somewhat counterintuitive that drug binding to P-gp would induce a further separation of NBDs, rather than bringing them towards the NBD dimerized and outward-facing state. However, other studies have also detected similar conformational shifts (e.g., [13,15,16], including P-gp structures co-crystallized with BDE-100. We would propose that the opening of the NBD1 nucleotide-binding site may be a precursor to further conformational shifts induced by nucleotide that then promote the formation of the outward-facing state.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/membranes11120923/s1>, Figure S1: Murine P-gp affinity for ivacaftor measured by thermostability changes; Figure S2: CTF correction and initial selection of images step-by-step workflow; Figure S3: Cryo-EM images initially visualized with the *cisTEM* software suite; Figure S4: Initial classification of the automatically picked particles; Figure S5: 3D classification of particles; Figure S6: Assessment of map *a* resolution; Figure S7: Different assessments of the resolution of all five maps; Figure S8: Comparison of model-to-map fits; Figure S9: Asymmetry in P-gp. Examples of major asymmetrical features that distinguish between the N- and C-terminal halves of the protein.

**Author Contributions:** Conceptualization, R.C.F.; methodology, A.B., N.T., T.S., R.F.C.; software, A.B., R.F.C.; validation, A.B., T.S., S.M.P.; investigation, A.B., N.T., T.S., R.F.C.; data curation, A.B., R.C.F.; writing—original draft preparation, R.C.F., A.B.; writing—review and editing, R.C.F., A.B., S.M.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funding for access to the UK eBIC facility was via Block Allocation Grouping 16619. Funding for NT was supported by the Development and Promotion of Science and Technology Talent Project (DPST), the Institute for the Promotion of Teaching Science and Technology (IPST), Thailand. Funding for AB is a University of Manchester/ASTAR Singapore Ph.D. studentship. Funding for TS is by the Punjab education endowment fund (PEEF) with a Chief Minister Merit Scholarship (SSMS).

**Data Availability Statement:** Experimental density maps (*a*), (*e*), and the atomic models can be downloaded via the electron microscopy database (EMDB) and via the Protein Data Bank (RCSB) under the codes EMD-13059, 7OTG (map *a*), and EMD-13060, 7OTI (map *e*).

**Acknowledgments:** We would like to thank Daniel Clare and Alistair Siebert (eBIC) for their support and advice for data collection and image processing. We thank Hao Fan (A\*STAR Bioinformatics Institute, Singapore) for advice and guidance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Callaghan, R.; Ford, R.C.; Kerr, I.D. The translocation mechanism of P-glycoprotein. *FEBS Lett.* **2006**, *580*, 1056–1063. [[CrossRef](#)] [[PubMed](#)]
2. Higgins, C.F. ABC transporters: From microorganisms to man. *Annu. Rev. Cell Biol.* **1992**, *8*, 67–113. [[CrossRef](#)] [[PubMed](#)]
3. Higgins, C.F.; Callaghan, R.; Linton, K.J.; Rosenberg, M.F.; Ford, R.C. Structure of the multidrug resistance P-glycoprotein. *Semin. Cancer Biol.* **1997**, *8*, 135–142. [[CrossRef](#)] [[PubMed](#)]
4. Rosenberg, M.F.; Callaghan, R.; Ford, R.C.; Higgins, C.F. Structure of the multidrug resistance P-glycoprotein to 2.5 nm resolution determined by electron microscopy and image analysis. *J. Biol. Chem.* **1997**, *272*, 10685–10694. [[CrossRef](#)]

5. Rosenberg, M.F.; Velarde, G.; Ford, R.C.; Martin, C.; Berridge, G.; Kerr, I.D.; Callaghan, R.; Schmidlin, A.; Wooding, C.; Linton, K.J.; et al. Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *Embo J.* **2001**, *20*, 5615–5625. [[CrossRef](#)]
6. Locher, K.P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol.* **2016**, *23*, 487–493. [[CrossRef](#)]
7. Alam, A.; Kung, R.; Kowal, J.; McLeod, R.A.; Tremp, N.; Broude, E.V.; Roninson, I.B.; Stahlberg, H.; Locher, K.P. Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1973–E1982. [[CrossRef](#)]
8. Nosol, K.; Romane, K.; Irobalieva, R.N.; Alam, A.; Kowal, J.; Fujita, N.; Locher, K.P. Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter ABCB1. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 26245–26253. [[CrossRef](#)]
9. Alam, A.; Kowal, J.; Broude, E.; Roninson, I.; Locher, K.P. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. *Science* **2019**, *363*, 753–756. [[CrossRef](#)]
10. Ford, R.C.; Marshall-Sabey, D.; Schuetz, J. Linker Domains: Why ABC Transporters ‘Live in Fragments no Longer’. *Trends Biochem. Sci.* **2020**, *45*, 137–148. [[CrossRef](#)]
11. Ford, R.C.; Beis, K. Learning the ABCs one at a time: Structure and mechanism of ABC transporters. *Biochem. Soc. Trans.* **2019**, *47*, 23–36. [[CrossRef](#)]
12. Ford, R.C.; Hellmich, U.A. What monomeric nucleotide binding domains can teach us about dimeric ABC proteins. *FEBS Lett.* **2020**, *594*, 3857–3875. [[CrossRef](#)]
13. Hofmann, S.; Janulienė, D.; Mehdipour, A.R.; Thomas, C.; Stefan, E.; Bruchert, S.; Kuhn, B.T.; Geertsma, E.R.; Hummer, G.; Tampe, R.; et al. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **2019**, *571*, 580–583. [[CrossRef](#)] [[PubMed](#)]
14. Aller, S.G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P.M.; Trinh, Y.T.; Zhang, Q.; Urbatsch, I.L.; et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **2009**, *323*, 1718–1722. [[CrossRef](#)]
15. Le, C.A.; Harvey, D.S.; Aller, S.G. Structural definition of polyspecific compensatory ligand recognition by P-glycoprotein. *IUCr* **2020**, *7*, 663–672. [[CrossRef](#)]
16. Nicklisch, S.C.; Rees, S.D.; McGrath, A.P.; Gokirmak, T.; Bonito, L.T.; Vermeer, L.M.; Cregger, C.; Loewen, G.; Sandin, S.; Chang, G.; et al. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure. *Sci. Adv.* **2016**, *2*, e1600001. [[CrossRef](#)]
17. Esser, L.; Zhou, F.; Pluchino, K.M.; Shiloach, J.; Ma, J.; Tang, W.K.; Gutierrez, C.; Zhang, A.; Shukla, S.; Madigan, J.P.; et al. Structures of the Multidrug Transporter P-glycoprotein Reveal Asymmetric ATP Binding and the Mechanism of Polyspecificity. *J. Biol. Chem.* **2017**, *292*, 446–461. [[CrossRef](#)] [[PubMed](#)]
18. Li, J.; Jaimes, K.F.; Aller, S.G. Refined structures of mouse P-glycoprotein. *Protein Sci.* **2014**, *23*, 34–46. [[CrossRef](#)]
19. Kim, Y.; Chen, J. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. *Science* **2018**, *359*, 915–919. [[CrossRef](#)]
20. Frank, G.A.; Shukla, S.; Rao, P.; Borgnia, M.J.; Bartesaghi, A.; Merk, A.; Mobin, A.; Esser, L.; Earl, L.A.; Gottesman, M.M.; et al. Cryo-EM Analysis of the Conformational Landscape of Human P-glycoprotein (ABCB1) During its Catalytic Cycle. *Mol. Pharmacol.* **2016**, *90*, 35–41. [[CrossRef](#)] [[PubMed](#)]
21. Thonghin, N.; Collins, R.F.; Barbieri, A.; Shafi, T.; Siebert, A.; Ford, R.C. Novel features in the structure of P-glycoprotein (ABCB1) in the post-hydrolytic state as determined at 7.9 Å resolution. *BMC Struct. Biol.* **2018**, *18*, 17. [[CrossRef](#)]
22. Lingam, S.; Thonghin, N.; Ford, R.C. Investigation of the effects of the CFTR potentiator ivacaftor on human P-glycoprotein (ABCB1). *Sci. Rep.* **2017**, *7*, 17481. [[CrossRef](#)] [[PubMed](#)]
23. Robertson, S.M.; Luo, X.; Dubey, N.; Li, C.; Chavan, A.B.; Gilmartin, G.S.; Higgins, M.; Mahnke, L. Clinical drug-drug interaction assessment of ivacaftor as a potential inhibitor of cytochrome P450 and P-glycoprotein. *J. Clin. Pharmacol.* **2015**, *55*, 56–62. [[CrossRef](#)] [[PubMed](#)]
24. Van Goor, F.; Hadida, S.; Grootenhuys, P.D.; Burton, B.; Cao, D.; Neuberger, T.; Turnbull, A.; Singh, A.; Joubran, J.; Hazlewood, A.; et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18825–18830. [[CrossRef](#)]
25. Davies, J.C.; Moskowitz, S.M.; Brown, C.; Horsley, A.; Mall, M.A.; McKone, E.F.; Plant, B.J.; Prais, D.; Ramsey, B.W.; Taylor-Cousar, J.L.; et al. VX-659-Tezacaftor-Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N. Engl. J. Med.* **2018**, *379*, 1599–1611. [[CrossRef](#)] [[PubMed](#)]
26. Keating, D.; Marigowda, G.; Burr, L.; Daines, C.; Mall, M.A.; McKone, E.F.; Ramsey, B.W.; Rowe, S.M.; Sass, L.A.; Tullis, E.; et al. VX-445-Tezacaftor-Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N. Engl. J. Med.* **2018**, *379*, 1612–1620. [[CrossRef](#)]
27. Liu, F.; Zhang, Z.; Levit, A.; Levring, J.; Touhara, K.K.; Shoichet, B.K.; Chen, J. Structural identification of a hotspot on CFTR for potentiation. *Science* **2019**, *364*, 1184–1188. [[CrossRef](#)]
28. Liddy, A.M.; McLaughlin, G.; Schmitz, S.; D’Arcy, D.M.; Barry, M.G. The pharmacokinetic interaction between ivacaftor and ritonavir in healthy volunteers. *Br. J. Clin. Pharmacol.* **2017**, *83*, 2235–2241. [[CrossRef](#)]
29. Wang, H.W.; Fan, X. Challenges and opportunities in cryo-EM with phase plate. *Curr. Opin. Struct. Biol.* **2019**, *58*, 175–182. [[CrossRef](#)]

30. Hughes, T.E.T.; Lodowski, D.T.; Huynh, K.W.; Yazici, A.; Del Rosario, J.; Kapoor, A.; Basak, S.; Samanta, A.; Han, X.; Chakrapani, S.; et al. Structural basis of TRPV5 channel inhibition by econazole revealed by cryo-EM. *Nat. Struct. Mol. Biol.* **2018**, *25*, 53–60. [\[CrossRef\]](#)
31. Twomey, E.C.; Yelshanskaya, M.V.; Vassilevski, A.A.; Sobolevsky, A.I. Mechanisms of Channel Block in Calcium-Permeable AMPA Receptors. *Neuron* **2018**, *99*, 956–968.e954. [\[CrossRef\]](#)
32. Yelshanskaya, M.V.; Li, M.; Sobolevsky, A.I. Structure of an agonist-bound ionotropic glutamate receptor. *Science* **2014**, *345*, 1070–1074. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Shang, G.; Zhang, C.; Chen, Z.J.; Bai, X.C.; Zhang, X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP-AMP. *Nature* **2019**, *567*, 389–393. [\[CrossRef\]](#)
34. Martin, G.M.; Yoshioka, C.; Rex, E.A.; Fay, J.F.; Xie, Q.; Whorton, M.R.; Chen, J.Z.; Shyng, S.L. Cryo-EM structure of the ATP-sensitive potassium channel illuminates mechanisms of assembly and gating. *Elife* **2017**, *6*, e24149. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Kim, Y.; Jeong, E.; Jeong, J.H.; Kim, Y.; Cho, Y. Structural Basis for Activation of the Heterodimeric GABAB Receptor. *J. Mol. Biol.* **2020**, *432*, 5966–5984. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Polovinkin, L.; Hassaine, G.; Perot, J.; Neumann, E.; Jensen, A.A.; Lefebvre, S.N.; Corringer, P.J.; Neyton, J.; Chipot, C.; Dehez, F.; et al. Conformational transitions of the serotonin 5-HT<sub>3</sub> receptor. *Nature* **2018**, *563*, 275–279. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Zhang, M.; Gui, M.; Wang, Z.F.; Gorgulla, C.; Yu, J.J.; Wu, H.; Sun, Z.J.; Klenk, C.; Merklinger, L.; Morstein, L.; et al. Cryo-EM structure of an activated GPCR-G protein complex in lipid nanodiscs. *Nat. Struct. Mol. Biol.* **2021**, *28*, 258–267. [\[CrossRef\]](#)
38. Nys, M.; Wijckmans, E.; Farinha, A.; Yoluk, O.; Andersson, M.; Brams, M.; Spurny, R.; Peigneur, S.; Tytgat, J.; Lindahl, E.; et al. Allosteric binding site in a Cys-loop receptor ligand-binding domain unveiled in the crystal structure of ELIC in complex with chlorpromazine. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6696–E6703. [\[CrossRef\]](#)
39. Lu, W.; Du, J.; Goehring, A.; Gouaux, E. Cryo-EM structures of the triheteromeric NMDA receptor and its allosteric modulation. *Science* **2017**, *355*, eaal3729. [\[CrossRef\]](#)
40. Shaik, M.M.; Peng, H.; Lu, J.; Rits-Volloch, S.; Xu, C.; Liao, M.; Chen, B. Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* **2019**, *565*, 318–323. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Chen, H.; Liu, Y.; Li, X. Structure of human Dispatched-1 provides insights into Hedgehog ligand biogenesis. *Life Sci. Alliance* **2020**, *3*, e202000776. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Krintel, C.; Dorosz, J.; Larsen, A.H.; Thorsen, T.S.; Venskutonyte, R.; Mirza, O.; Gajhede, M.; Boesen, T.; Kastrop, J.S. Binding of a negative allosteric modulator and competitive antagonist can occur simultaneously at the ionotropic glutamate receptor GluA2. *FEBS J.* **2021**, *288*, 995–1007. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Bai, J.; Swartz, D.J.; Protasevich, I.I.; Brouillette, C.G.; Harrell, P.M.; Hildebrandt, E.; Gasser, B.; Mattanovich, D.; Ward, A.; Chang, G.; et al. A gene optimization strategy that enhances production of fully functional P-glycoprotein in *Pichia pastoris*. *PLoS ONE* **2011**, *6*, e22577. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Beaudet, L.; Urbatsch, I.L.; Gros, P. High-level expression of mouse Mdr3 P-glycoprotein in yeast *Pichia pastoris* and characterization of ATPase activity. *Methods Enzymol.* **1998**, *292*, 397–413. [\[CrossRef\]](#)
45. Thonghin, N.; Kargas, V.; Clews, J.; Ford, R.C. Cryo-electron microscopy of membrane proteins. *Methods* **2018**, *147*, 176–186. [\[CrossRef\]](#)
46. Kohlstaedt, M.; von der Hocht, I.; Hilbers, F.; Thielmann, Y.; Michel, H. Development of a Thermofluor assay for stability determination of membrane proteins using the Na<sup>+</sup>/H<sup>+</sup> antiporter NhaA and cytochrome c oxidase. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71*, 1112–1122. [\[CrossRef\]](#)
47. Meng, X.; Clews, J.; Kargas, V.; Wang, X.; Ford, R.C. The cystic fibrosis transmembrane conductance regulator (CFTR) and its stability. *Cell Mol. Life Sci.* **2017**, *74*, 23–38. [\[CrossRef\]](#)
48. Danev, R.; Buijsse, B.; Khoshouei, M.; Plitzko, J.M.; Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15635–15640. [\[CrossRef\]](#)
49. Li, X.; Mooney, P.; Zheng, S.; Booth, C.R.; Braunfeld, M.B.; Gubbens, S.; Agard, D.A.; Cheng, Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **2013**, *10*, 584–590. [\[CrossRef\]](#)
50. Grant, T.; Rohou, A.; Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife* **2018**, *7*, e35383. [\[CrossRef\]](#)
51. Afonine, P.V.; Poon, B.K.; Read, R.J.; Sobolev, O.V.; Terwilliger, T.C.; Urzhumtsev, A.; Adams, P.D. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **2018**, *74*, 531–544. [\[CrossRef\]](#)
52. Kucukelbir, A.; Sigworth, F.J.; Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **2014**, *11*, 63–65. [\[CrossRef\]](#)
53. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Swint-Kruse, L.; Brown, C.S. Resmap: Automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics* **2005**, *21*, 3327–3328. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Chen, S.; McMullan, G.; Faruqi, A.R.; Murshudov, G.N.; Short, J.M.; Scheres, S.H.; Henderson, R. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **2013**, *135*, 24–35. [\[CrossRef\]](#)

- 
56. Thomas, C.; Aller, S.G.; Beis, K.; Carpenter, E.P.; Chang, G.; Chen, L.; Dassa, E.; Dean, M.; Duong Van Hoa, F.; Ekiert, D.; et al. Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett.* **2020**, *594*, 3767–3775. [[CrossRef](#)] [[PubMed](#)]
  57. Pan, L.; Aller, S.G. Allosteric Role of Substrate Occupancy Toward the Alignment of P-glycoprotein Nucleotide Binding Domains. *Sci. Rep.* **2018**, *8*, 14643. [[CrossRef](#)]
  58. Shapiro, A.B.; Fox, K.; Lam, P.; Ling, V. Stimulation of P-glycoprotein-mediated drug transport by prazosin and progesterone. Evidence for a third drug-binding site. *Eur. J. Biochem.* **1999**, *259*, 841–850. [[CrossRef](#)]
  59. Pajeva, I.K.; Hanl, M.; Wiese, M. Protein contacts and ligand binding in the inward-facing model of human P-glycoprotein. *ChemMedChem* **2013**, *8*, 748–762. [[CrossRef](#)]
  60. Bocci, G.; Moreau, A.; Vayer, P.; Denizot, C.; Fardel, O.; Parmentier, Y. New insights in the in vitro characterisation and molecular modelling of the P-glycoprotein inhibitory promiscuity. *Eur. J. Pharm. Sci.* **2018**, *121*, 85–94. [[CrossRef](#)]

# ABCG2 requires a single aromatic amino acid to “clamp” substrates and inhibitors into the binding pocket

Tomoka Gose<sup>1</sup> | Talha Shafi<sup>2</sup> | Yu Fukuda<sup>1</sup> | Sourav Das<sup>3</sup> | Yao Wang<sup>1</sup> |  
 Alice Allcock<sup>1</sup> | Ailsa Gavan McHarg<sup>1</sup> | John Lynch<sup>1</sup> | Taosheng Chen<sup>3</sup> |  
 Ikumi Tamai<sup>4</sup> | Anang Shelat<sup>3</sup> | Robert C. Ford<sup>2</sup> | John D. Schuetz<sup>1</sup>

<sup>1</sup>Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN, USA

<sup>2</sup>School of Biological Sciences, The University of Manchester, Manchester, UK

<sup>3</sup>Department of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, Memphis, TN, USA

<sup>4</sup>Department of Pharmaceutical Sciences, Kanazawa University, Kanazawa, Japan

## Correspondence

John D. Schuetz, Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA.  
 Email: john.schuetz@stjude.org

## Funding information

NIH, ALSAC, Grant/Award Number: R01CA194057, R35-GM118041, P30CA21745, CA21865, CA96832

## Abstract

ATP-binding cassette sub-family G member 2 (ABCG2) is a homodimeric ATP-binding cassette (ABC) transporter that not only has a key role in helping cancer cells to evade the cytotoxic effects of chemotherapy, but also in protecting organisms from multiple xeno- and endobiotics. Structural studies indicate that substrate and inhibitor (ligands) binding to ABCG2 can be differentiated quantitatively by the number of amino acid contacts, with inhibitors displaying more contacts. Although binding is the obligate initial step in the transport cycle, there is no empirical evidence for one amino acid being primarily responsible for ligand binding. By mutagenesis and biochemical studies, we demonstrated that the phylogenetically conserved amino acid residue, F439, was critical for both transport and the binding of multiple substrates and inhibitors. Structural modeling implied that the  $\pi$ - $\pi$  interactions from each F439 monomer mediated the binding of a surprisingly diverse array of structurally unrelated substrates and inhibitors and that this symmetrical  $\pi$ - $\pi$  interaction “clamps” the ligand into the binding pocket. Key molecular features of diverse ABCG2 ligands using the  $\pi$ - $\pi$  clamp along with structural studies created a pharmacophore model. These novel findings have important therapeutic implications because key properties of ligands interacting with ABCG2 have been discovered. Furthermore, mechanistic insights have been revealed by demonstrating that for ABCG2 a single amino acid is essential for engaging and initiating transport of multiple drugs and xenobiotics.

## KEYWORDS

BCRP, cancer, CETSA, drug, thermal shift assay, kinase inhibitor, transporter

## 1 | INTRODUCTION

ATP-binding cassette (ABC) transporters have key roles in the energy-dependent movement of molecules across biological

membranes, including important chemotherapeutics. ATP-binding cassette sub-family G member 2 (ABCG2), also known as breast cancer resistance protein (BCRP), a ubiquitously expressed transporter, has an important physiological

**Abbreviations:** ABCG2, ATP-binding cassette sub-family G member 2; CETSA, cellular thermal shift assay;  $EC_{50}$ , half maximal effective concentration; EM, electron microscopy; FTC, fumitremorgin C; GPCR, G-protein coupled receptor;  $IC_{50}$ , half maximal inhibitory concentration; KI, kinase inhibitor; MEF, mouse embryonic fibroblast; MOE, molecular operating environment; PhA, pheophorbide a;  $T_m$ , melting temperature; TMDs, transmembrane domains; TMH, transmembrane helices.



role in creating a barrier to penetration of many cytotoxins into cells. For instance, ABCG2 protects hematopoietic progenitors from cytotoxins, both endogenous and chemotherapeutics, and has key protective roles in the blood-brain and blood-testis barriers.<sup>1-3</sup> Further, genetic variation in ABCG2 contributes to the susceptibility to toxicity from endogenous substances, for example, uric acid.<sup>4-6</sup> ABCG2 dysfunction is linked to hyperuricemia, which can result in gout, kidney disease, and hypertension, all of which are secondary to the impaired export of uric acid.<sup>4,5</sup> ABCG2, like its functional homologs ABCB1 (P-glycoprotein) and ABCC1 (MRP1), has a major protective role against many chemotherapeutics, effecting the pharmacokinetics of many commonly used drugs, and its expression correlates with a poor prognosis and treatment outcome of certain cancers (eg, acute myeloid leukemia<sup>7</sup>). ABCG2 has broad substrate specificity, with a seeming preference for ligands (defined here as either substrates and/or inhibitors) that are polycyclic, hydrophobic compounds.<sup>8</sup> Among the many medicines and xenobiotics, the therapeutically important kinase inhibitors (KIs) interact with ABCG2 as both substrate and inhibitor (see Table S1) producing important drug interactions.<sup>9</sup> Despite recent cryo-electron microscopy studies on the substrate and inhibitor binding pocket of ABCG2,<sup>10,11</sup> there is no knowledge or empirical evidence, in either cells or native cell membranes, for ligands preferring any one amino acid to mediate binding to ABCG2.

One of the crucial unanswered questions concerning multidrug transporters, such as ABCG2, is: Is there a specific amino acid that is key to substrate and inhibitor binding to ABCG2? The cryo-electron microscopy structures of ABCG2 are unclear as to whether there is a single amino acid that promotes binding to the binding pocket.<sup>10,11</sup> At present, it appears that inhibitors can be differentiated from substrates by both their greater size and sheer number and location of potential contacts with ABCG2.<sup>11</sup> For ABCG2, an analog of the inhibitor, Ko143 interacted with 16 residues distributed among transmembrane helices (TMH) 1, 2, and 5,<sup>10</sup> whereas a substrate, estrone-3-sulfate, interacted with fewer residues. The general interpretation was that the total number of contacts enabled tight binding of inhibitors (eg, tariquidar) compared to substrates (eg, estrone-sulfate). Thus, inhibitors, with more contacts were hypothesized to create a “wedge” that blocked substrate transport. Accordingly, one might

infer that disruption of inhibitor binding to ABCG2 would require the loss of more molecular contacts than that of a substrate, and that substrates and inhibitors might not depend on the the same contacts to bind to ABCG2.

Unexpectedly, for ABCG2, our studies discovered that in a native cellular environment, the substitution of a highly conserved amino-acid caused the loss of substrate transport and inhibitor binding. Our studies reveal a key feature of the amino acid F439, where, through  $\pi$ - $\pi$  stacking, it appears to form a molecular “clamp” in the ABCG2 binding pocket. This conserved feature might be important to other ABCG family members (eg, ABCG5/8). Importantly, this substitution of alanine for phenylalanine at position 439 results in the loss of ABCG2 interaction with multiple ligands, suggesting that this amino acid interaction is a key initiating step for ABCG2 transport and inhibitor interaction in cells.

## 2 | MATERIALS AND METHODS

### 2.1 | Chemicals

Fumitremorgin C (FTC) and Ko143 were obtained from Enzo Life Sciences. Hoechst 33342 was obtained from Life Technologies. Pheophorbide a (PhA) was obtained from Frontier Scientific Inc PhA was dissolved in DMSO and aliquots protected from light at  $-20^{\circ}\text{C}$ . Other compounds were stored at  $-20^{\circ}\text{C}$  at 10 mM in DMSO.

### 2.2 | Plasmid construction and site-directed mutagenesis

All site-directed mutations were generated in the MSCV-ABCG2-IRES-GFP expression plasmid using QuikChange II XL site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA). The primers used in this study are indicated in the Table 1.

### 2.3 | Cell culture and transfection

Mouse embryonic fibroblast (MEF) cell line was derived from *Abcg2*<sup>-/-</sup> knockout (KO) mice. The *Abcg2*-KO mice

**TABLE 1** Oligonucleotide primers used to generate ABCG2 mutations<sup>a</sup>

Mutants	Forward primer (5'→3')	Reverse primer (5'→3')
N436A	ctcttctctgacgaccGCCagtggttcagcagtg	acactgctgaacactggGCggtcgtcaggaagaagag
F439A	tgacgaccaaccagtgGTCagcagtggttcagccg	cggctgaacactgctgGCacactggttgctgca
F439W	tgacgaccaaccagtgGTCagcagtggttcagccg	cggctgaacactgctGCCaacactggttgctgca
F439Y	tgacgaccaaccagtgTAcagcagtggttcagccg	cggctgaacactgctgTaaactggttgctgca

<sup>a</sup>Capital letters indicate substitution.

were obtained from the Sorrentino lab.<sup>2</sup> The *Abcg2*-KO MEFs were engineered to express human ABCG2 (Uniprot: Q9UNQ0). Cells were cultured at 37°C with 5% of CO<sub>2</sub>. Cells were grown in Dulbecco's Modified Essential Medium (D-MEM) (Lonza, Basel, Switzerland) supplemented with 10% of fetal bovine serum (Gibco, Grand Island, NY), penicillin (100 U/mL), streptomycin (100 µg/mL) (Gibco, Grand Island, NY), and L-glutamine (2 mM) (Gibco, Grand Island, NY). The *Abcg2*-KO MEF cell line was transduced with the retroviruses harboring cDNAs encoding either the mutant ABCG2 N436A, F439A, F439W, or F439Y.

## 2.4 | Transport assays

The intracellular accumulation of fluorescent PhA, a specific ABCG2 substrate,<sup>12,13</sup> or Hoechst 33342 was used to determine the function of ABCG2. 10 µM of Fumitremogin C (FTC), a specific inhibitor of ABCG2 activity,<sup>14</sup> was used as a positive control. Cells were seeded into a 96-well cell culture plates at 10 000 cells/well and PhA or Hoechst 33342 (0.5 µM of final concentration) and either FTC or the test ligand added to the culture medium followed by an incubation for 1 hour (PhA) or 2 hours (Hoechst 33342) in the dark. Thereafter, the medium was removed, cells were washed twice with ice-cold PBS and the PhA fluorescence was then quantitatively determined with a Synergy H4 microplate reader (BioTek, Winooski, VT, USA) at an excitation wavelength of 405 nm (PhA) or 355 nm (Hoechst 33342) and an emission wavelength of 667 nm (PhA) or 460 nm (Hoechst 33342). The data were normalized by the proportion accumulated relative (%) to FTC inhibition (see Figure S1 for FTC dose-response curve). The accumulation of Hoechst 33342 was normalized to protein concentration as determined using the BCA Protein Assay Kit (Pierce, Rockford, IL, USA).

## 2.5 | Cellular thermal shift assay

The cellular thermal shift assay (CETSA)<sup>15-17</sup> was adapted for ABCG2 as follows: To establish the melting temperature ( $T_m$ ) for ABCG2, MEFs lacking *Abcg2* (KO), but harboring human ABCG2 were gently detached with a weak trypsin solution (0.05% of trypsin). After this, the cells were washed with PBS three times followed by resuspension in PBS supplemented with protease inhibitors (complete Tablets-Mini EASYpack, Roche, Basel, Switzerland). A portion of the cell suspension was reserved for protein determination by the BCA protein assay. The resuspended cells were then heated in a PCR machine for 3 minutes at various temperatures (59-75°C), followed by a 3 min incubation at room temperature. Samples were then treated with ice-cold PBS supplemented with NP-40 to a final concentration of 0.8%, snap-frozen in liquid nitrogen,

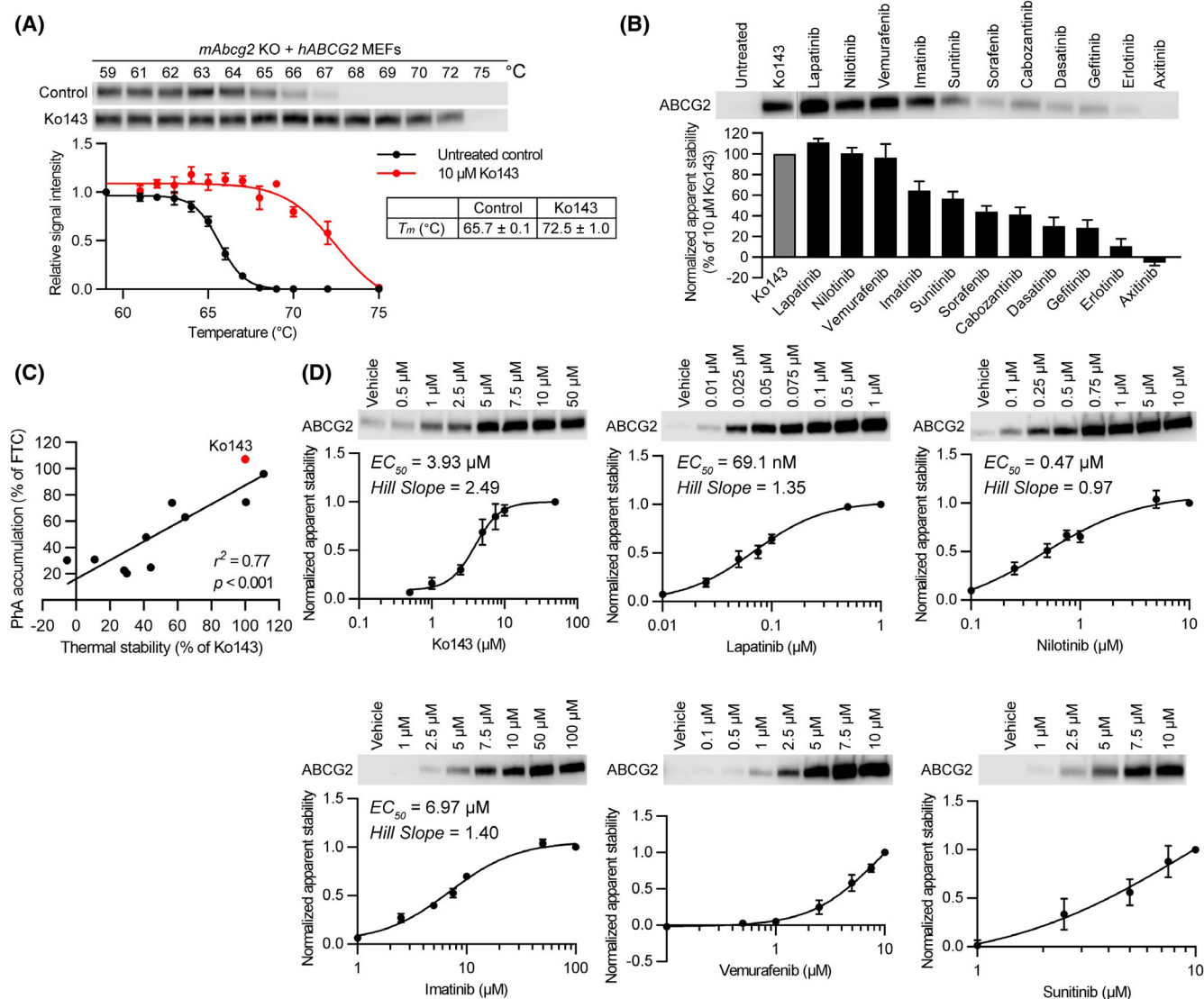
and then placed into a thermoshaker at 25°C to initiate thawing before being transferred to ice until the entire contents were thawed. This freeze-thaw cycle was repeated twice before ultracentrifugation to precipitate the denatured protein (20 minutes at 4°C and 100 000 g) and the soluble fractions were subjected to immunoblot analysis using an ABCG2 antibody. Based on the ABCG2 thermal denaturation curve, we empirically selected 67°C as the temperature to interrogate the ability of ligands to stabilize wildtype (WT) ABCG2. To determine the relative thermal stabilization of ABCG2 by ligands, cells were incubated with ligand for 60 minutes at 37°C before trypsin treatment. The immunoblot data were normalized with respect to the specific ABCG2 inhibitor, Ko143 and the signal intensity reported as % of Ko143 (Figure 1B). In isothermal dose-response experiments, the signal intensity was normalized to the intensity of the highest concentration.

## 2.6 | Isolation of plasma membranes

MEF cell lines were pelleted and washed with 1x PBS before suspension in a hypotonic buffer (0.5 mM of Tris (7.4), 0.1 mM of EDTA and protease inhibitors) and freezing. After gentle thawing, cells were pelleted by centrifugation, suspended in 2.5 mL of hypotonic buffer and spun for 40 minutes in an SW32Ti swinging bucket rotor at 100 000 g. Pellets were removed and resuspended into Homogenization Buffer (250 mM of sucrose, 50 mM of Tris, 250 µM of CaCl<sub>2</sub>, and protease inhibitors) and homogenized with 50 strokes of a tight dounce pestle. Following centrifugation at 500 g for 5 minutes, the pellets were then resuspended in Homogenization Buffer and homogenized again before an additional 500 g centrifugation step. The supernatants of each spin were pooled together and layered over a 5 mL 35% of sucrose pad formed in 50 mM of Tris (pH 7.4) and centrifuged in the SW32Ti rotor for 1.5 hours at 100 000 g. The interface was carefully removed and diluted with Dilution Buffer (50 mM of Tris, (pH 7.4), 25 mM of sucrose) and the crude membranes were pelleted at 100 000 g. Pellets were resuspended in 25 mM of Tris (pH 7.4) and incubated overnight at 4°C on ice. These pellets were then passed 10x through a 27 gauge needle before protein calculation. The membranes were typically usually used immediately or stored at -80°C for one thaw cycle.

## 2.7 | Membrane thermal shift

Human ABCG2-expressing membrane vesicles, from HEK293 cells, were purchased from Solvo Biotechnology (Budapest, Hungary) or prepared from MEFs. Thermal shift assay was adapted according to Ashok et al.<sup>18</sup> Briefly, membrane vesicles (5 µg protein/20 µL final reaction volume) were heated in a PCR machine for 3 minutes at various



**FIGURE 1** Relationship between kinase inhibitor-binding to ABCG2 and pheophorbide (PhA) transport inhibition. A, CETSA-melting curve of ABCG2 and ABCG2 CETSA curve shift upon ligand binding. Western blot for ABCG2 is shown in the inset. mAbcg2-KO MEFs expressing human ABCG2 were incubated with 10  $\mu$ M of Ko143 ( $n = 3$ , red) or 0.1% of DMSO (untreated control) ( $n = 7$ , black) for 60 min at 37°C. Samples were heated for 3 min at various temperatures (59–75°C). The signal intensity was normalized to the intensity of the 59°C sample. B, CETSA was performed to measure the binding ability of kinase inhibitors to ABCG2 in mAbcg2-KO MEFs expressing human ABCG2. mAbcg2-KO MEFs expressing human ABCG2 were incubated with 10  $\mu$ M of kinase inhibitor or 10  $\mu$ M of Ko143 (positive control) for 60 min at 37°C. Samples were heated for 3 min at 67°C. The immunoblot data were normalized with respect to Ko143 and the signal intensity reported as % of Ko143. Each bar represents the mean  $\pm$  SE ( $n = 5$ –7). C, Scatter plot showing significant positive correlation between kinase inhibitor binding to ABCG2 and transport inhibition. The calculated correlation coefficient ( $r^2$ ) between thermal stability in 10  $\mu$ M of kinase inhibitor or Ko143 treatment (Figure 1B) and PhA accumulation in 1  $\mu$ M of kinase inhibitor or Ko143 treatment (PhA transport inhibition, Figure S1) was 0.77 ( $P < .001$ ). D, mAbcg2-KO hABCG2 expressing MEFs were incubated with various concentrations of Ko143 or kinase inhibitor (Lapatinib, Nilotinib, Imatinib, Vemurafenib, or Sunitinib) for 60 min at 37°C. Dose-dependent stabilization of ABCG2 was assessed after heating samples at 67°C for 3 min. Each bar represents the mean  $\pm$  SE ( $n = 3$ –5)

temperatures (37–70°C) to establish a thermal denaturation curve. Samples were then treated with ice-cold PBS supplemented with NP-40 to a final concentration of 0.8%, snap-frozen in liquid nitrogen, and then placed into a thermoshaker at 25°C to initiate thawing before being transferred to ice after the entire contents were thawed. After this, the freeze-thaw cycle was repeated and ultracentrifugation performed

to precipitate the denatured protein (20 minutes at 4°C and 100 000 g). The supernatant was subjected to immunoblot analysis using an ABCG2 antibody (Enzo Life Sciences Inc, BXP53, 1:1000). Based on an extrapolation from the thermal denaturation curve that was predicted to produce 99% loss of protein in the supernatant, 65°C was selected for wildtype, 67°C for N436A, 61°C for F439A, 65°C for F439Y, and



67°C for F439W. For ABCG2-expressing HEK293 membrane vesicles, 62°C was selected based on the ABCG2 thermal denaturation curve. To assess the ability of ligands to thermally stabilize ABCG2, membranes were incubated with ligand for 60 minutes or with the non-hydrolyzable ATP analog, AMP-PNP for 30 minutes at 37°C. The signal intensity of the heated sample was normalized to the signal intensity of the unheated samples (WT and mutants) and the signal intensity reported as % of WT-ABCG2.

## 2.8 | Immunoblot analysis

Cell lysate was subjected to SDS-PAGE, and proteins were transferred to a nitrocellulose membrane. Immunoblot analysis was performed using primary antibody against ABCG2 (Enzo Life Sciences Inc, BXP53, 1:1000) and was subjected to further incubation with HRP-conjugated secondary antibodies raised against rat IgG (Jackson ImmunoResearch Laboratories Inc, West Grove, PA) as previously described.<sup>19</sup> Chemiluminescence signal was detected using LICOR Odyssey system. The bands were analyzed using LI-COR Image Studio Lite Ver 5.2 software for quantification of western blot signals.

## 2.9 | Pharmacophore model

A global pharmacophore model was created in molecular operating environment (MOE) by aligning all 20 ABCG2 ligands using the Flexible Alignment application with default parameters.<sup>20</sup> The best alignment, as determined by the “Grand Alignment Score,” was used to create a consensus pharmacophore model, the “Global Pharmacophore Model.”

## 2.10 | In silico ligand modeling

Docking experiments were performed on the inward facing electron microscopy (EM) structure of ABCG2 (PDB: 6ETI) using MOE 2019v01. The protein structure was prepared by protonating at physiological pH (7.4) followed by partial charge calculations using default parameters in MOE. Site finder implemented in the software was employed to identify residues in the TMDs (transmembrane domains) which were then manually selected to restrict the binding cavity to the solvent accessible residues inside the binding cavity. Ligands were drawn in MOE, which were then stochastically minimized for generating 3D conformations. Partial charges on ligands were calculated using the Gastegier PEOE method implemented in the software. The docking experiment employed the alpha triangle placement method along

with London-dG as scoring function. A total of 100 poses (without duplication) were generated for each docked ligand. Each pose was then minimized and rescored using the GBVI/WSA-dG scoring function.

## 2.11 | Cytotoxicity assays

CellTiter-Glo luminescent cell viability assay was used to determine the cytotoxicity of mitoxantrone, topotecan, Bis-POM-PMEA, and cladribine using the MEFs lacking *Abcg2* (KO) but harboring the human ABCG2-WT, ABCG2-F439A, or vector control. Cells were seeded into a 96-well cell culture plates at 1000 cells/well and cytotoxic drug was added to the culture medium followed by an incubation for 3 days.

## 2.12 | Data analysis and curve fitting

All data analysis and curve fitting were carried out using GraphPad Prism (v. 8.1. 2) (Prism; La Jolla, CA). Melting temperature curve ( $T_m$ ) or dose-response curve ( $EC_{50}$ ) were fit with the nonlinear regression with three parameters (1) or four parameters (variable slope) (2) and are shown in each figure according to the best-fit model as determined by  $r^2$  value.

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{((\text{LogEC}_{50} - X))}) \quad (1)$$

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{((\text{LogEC}_{50} - X) * \text{HillSlope}))}) \quad (2)$$

Top and Bottom are plateaus in the units of the Y axis.

Data shown are representative of mean  $\pm$  SE of at least three independent experiments. Apparent  $EC_{50}$  and  $IC_{50}$  values were calculated from dose-response data using nonlinear regression analysis.

## 3 | RESULTS

### 3.1 | CETSA reveals KIs bind with differential affinity to ABCG2

The cellular thermal shift assay (CETSA) was developed to assist in the identification of cancer chemotherapeutic protein targets.<sup>17,21</sup> The guiding principle is that a target protein will unfold and precipitate, upon heating, whereas a protein engaged with a ligand will generally require a higher temperature to unfold and precipitate.<sup>15-17,21</sup> To validate this approach, we constructed a thermal denaturation curve for

human ABCG2 (expressed in *abcg2* knockout mouse embryo fibroblasts (MEFs) to eliminate dimerization with the endogenous ABCG2) that revealed a  $T_m$  of 65.7°C. Upon treatment of cells with the well-known ABCG2 inhibitor, Ko143, the  $T_m$  was elevated almost 7 to 72.5°C (Figure 1A), a shift in the  $T_m$  of similar magnitude by Ko143 binding to purified ABCG2 was reported.<sup>10</sup>

To determine if there was a relationship between ligand induced thermal stability and transport, eleven KIs, many reported as substrates and inhibitors (Table S1), were evaluated for their ability to inhibit the transport of the ABCG2-specific substrate, PhA.<sup>12</sup> The KIs showed dose-dependent inhibition, and a wide range of potencies (Figure S1). To determine if these KIs thermally stabilized ABCG2, cells were incubated with 10  $\mu$ M of each KI followed by an incubation at 67°C. Some KIs induced a strong thermal shift while others produced either a weak or minimal thermal shift, suggesting either little or no ability to bind ABCG2 and provide thermal stability (Figure 1B). We then determined if there was a relationship between the CETSA measured ABCG2 thermal stabilization (expressed as a proportion (%) of Ko143 induced stabilization) and transport inhibition by comparing the % inhibition by the KI relative to that produced by FTC (see Figure S1). Among many of the KIs, the relationship between KI thermal stabilization of ABCG2 and transport inhibition was strong with a calculated correlation coefficient ( $r^2$ ) between ABCG2 thermal stability and transport inhibition of 0.77 ( $P < .001$ ) (Figure 1C). Vemurafenib showed strong binding to ABCG2, but weak transport inhibitory effect, however, even with inclusion of this outlier the correlation was still strong ( $r^2 = .58$ ,  $P = .004$ ).

We next hypothesized that the potency of a KI to produce an ABCG2 thermal shift might be related to their potency of ABCG2 transport inhibition. As a control, Ko143 was used because it has a known binding site on ABCG2.<sup>10,11</sup> Ko143 displayed dose-dependent stabilization of ABCG2 with an  $EC_{50}$  of 3.93  $\mu$ M and a Hill coefficient of 2.49 (Figure 1D), consistent with the ABCG2 structure displaying two potential Ko143 analogs in the binding pocket.<sup>10</sup> Next, the KIs were tested. Lapatinib displayed the lowest  $EC_{50}$  for binding of about 69 nM, whereas sunitinib was the least potent. Among the KIs, most showed a Hill slope slightly greater than one. Importantly, a strong correlation was observed between the KI thermal stabilization of ABCG2 (Figure 1B) and  $IC_{50}$  for transport inhibition (Figure S1) ( $r^2 = .88$ ,  $P = .002$ ) (Figure S2).

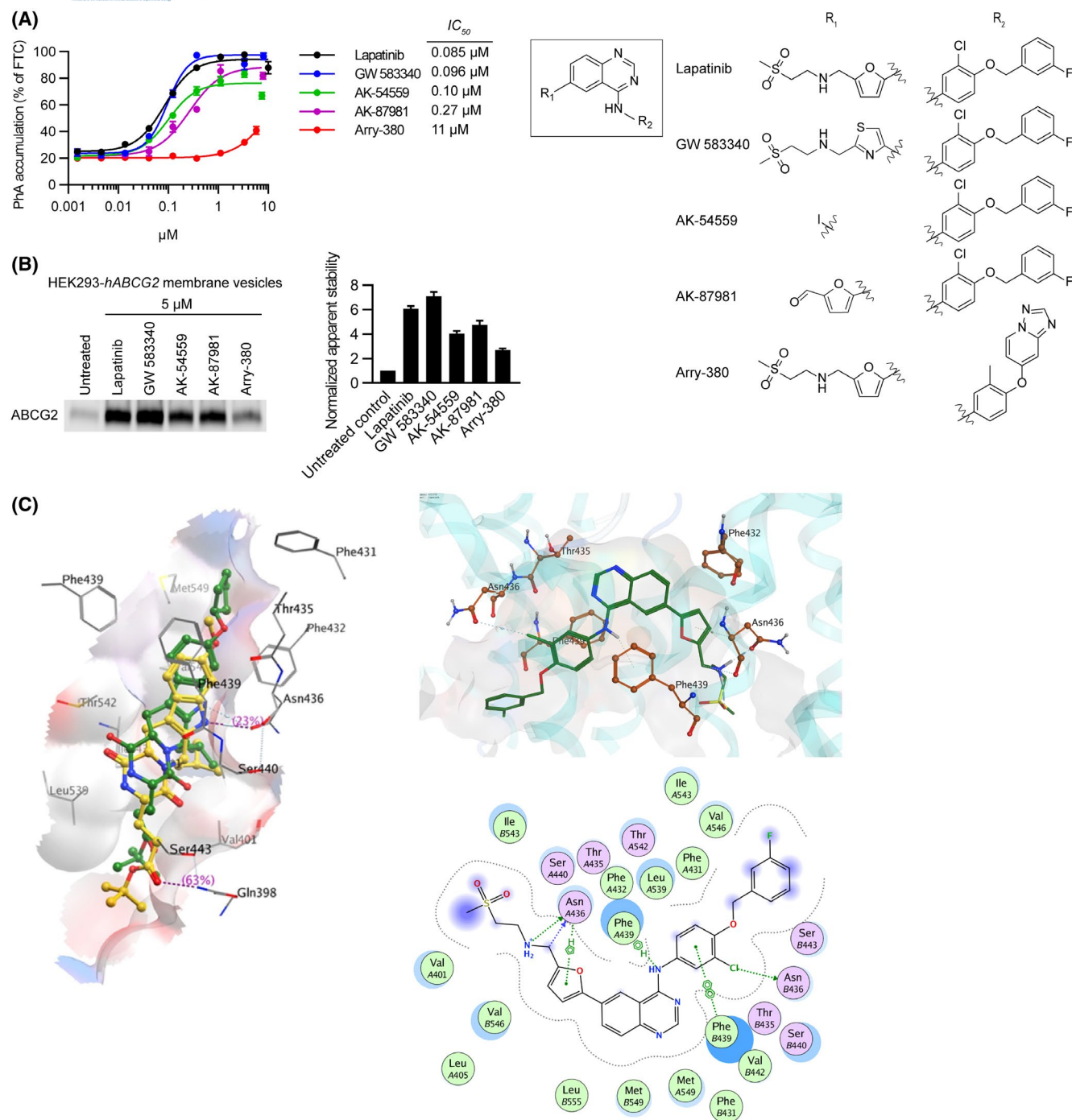
### 3.2 | Lapatinib analogs and structural model of binding

Because Lapatinib was the most potent in inducing a thermal shift among the KIs tested, we investigated if it had

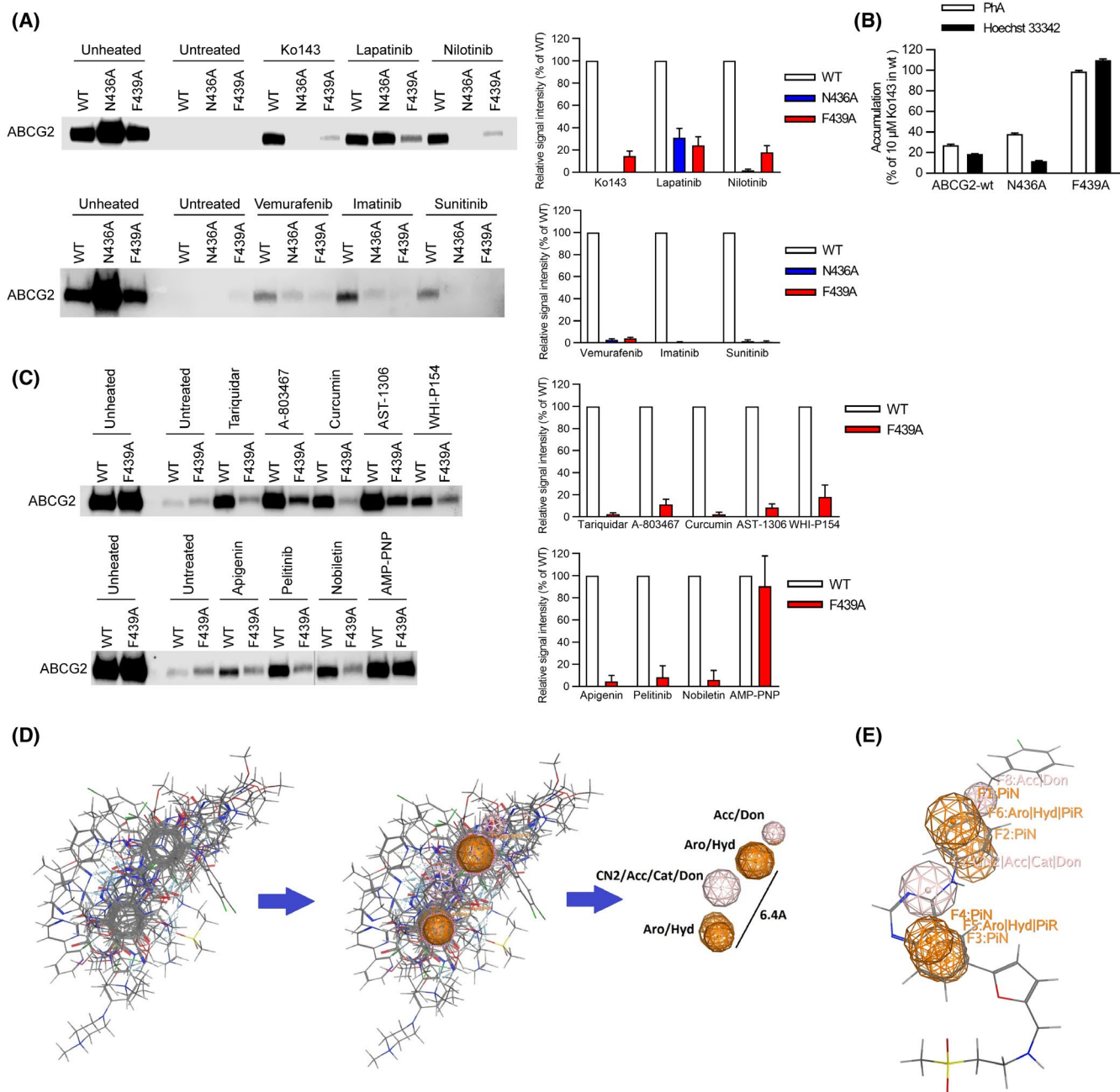
molecular features that were important determinants for interaction with ABCG2 (Figure 2). Four Lapatinib analogs were available to evaluate for their ability to inhibit ABCG2 mediated transport. Three of these had  $IC_{50}$  values similar to Lapatinib ( $IC_{50} = 0.085 \mu$ M), while the fourth, Arry-380 was 130-fold lower at 11  $\mu$ M (Figure 2A). These studies were then extended to assess the ability of these ligands to thermally stabilize ABCG2 by the membrane thermal shift assay. Lapatinib and GW583340 were most effective in stabilizing ABCG2 against thermal denaturation followed by AK-87981 and AK-54559. Paralleling the transport inhibition data, the least potent stabilizer was Arry-380 (Figure 2B). Structure-based docking experiments using ABCG2 (PDB: 6ETI) predicted the most likely binding poses of the KIs used in this study. We noted that a single amino acid, F439, was predicted to be involved in binding via  $\pi$ - $\pi$  stacking to all of the drugs. F439 is close to the two-fold symmetry axis of this dimeric transporter, and a F439 residue from each monomer participates in ligand binding. The top docking pose of Ko143 was consistent with the cryo-EM structure that would allow the binding of two molecules. In contrast, most of the other KIs, such as Lapatinib, had poses, where only one drug molecule was threaded in between the two F439 residues. Since none of the KIs have symmetry, the two F439 residues appear to form a “clamp” via  $\pi$ - $\pi$  interaction with the aromatic groups within the various drugs. These observations helped to elucidate the nature of the ABCG2 pharmacophore (generated separately below). Although F439 was identified as a universal interacting partner with all the KIs examined, the N436 was also predicted to hydrogen-bond with many of the drugs.

### 3.3 | Highly conserved F439 is required to confer ABCG2 thermal stability for multiple ligands

Based on the Lapatinib modeling studies (Figure 2C), which suggested that residues N436 and F439 frequently contribute to ligand interactions with ABCG2, we hypothesized that they might play a key role for KIs interacting with ABCG2. To investigate this, we developed ABCG2 retroviral expression vectors harboring site-directed mutants that converted either F439 to alanine or N436 to alanine. These were then transduced into *Abcg2*-knockout MEFs. Subsequently, membranes were prepared. These membranes display the same properties as cells in that ATP and its non-hydrolyzable analog (AMP-PNP) and stabilize ABCG2 against thermal denaturation (Figure 3B, Figure S4A-D). The  $T_m$  for F439A ( $T_m = 57^\circ\text{C}$ ) was lower than WT-ABCG2 (WT  $T_m = 59^\circ\text{C}$ ), whereas the  $T_m$  for N436A was higher at 62°C (see Figure S3B). To test the impact of



**FIGURE 2** Lapatinib analogs reveal important residues for interacting with ABCG2. A, 0.5  $\mu$ M of PhA was used as substrate and Fumitremorgin C (FTC) (positive control), Lapatinib, or Lapatinib analogs were used as inhibitors to measure inhibitory effect ( $IC_{50}$ ) of ABCG2. mAbcg2-KO MEFs expressing human ABCG2 were incubated with PhA and inhibitor for 60 min. Each bar represents the mean  $\pm$  SE ( $n = 4$ ). Molecular structures of Lapatinib and its analogs (right). Core structure is shown at the middle, and the individual  $R_1$  and  $R_2$  constituents of Lapatinib and its analogs GW 583340, AK-54559, AK-87981, and Arry-380 are shown right the core structure. B, HEK293-hABCG2 membrane vesicles were incubated with 5  $\mu$ M of Lapatinib or Lapatinib analogs for 60 min at 37°C. Samples were heated for 3 min at 62°C. Representative western blot is shown on the left side and the densitometry quantification data from three independent experiments is shown on the right side. C, Diagrammatic representation of Ko143 analog mz29 and lapatinib in the ABCG2 binding cavity. Left panel: the binding of Ko143 (yellow) and mz29 (green) with PDB 6ETI. Right panel: 3D (top side) and 2D (bottom side) representation of Lapatinib, a tyrosine kinase inhibitor, within the ABCG2 binding cavity



**FIGURE 3** A key amino acid residue in the binding pocket is required for binding to ABCG2. A, Role of N436 and F439 in kinase inhibitor binding assessed by thermal shift. Representative western blots are shown on the left side and the densitometry quantification data from three to four independent experiments are shown on the right side. Membrane vesicles prepared from mAbcg2-KO MEFs expressing human ABCG2 (WT, N436A, and F439A) were incubated with 10  $\mu$ M of kinase inhibitor for 60 min at 37°C. Samples were heated for 3 min at 65°C for WT-ABCG2 (white), 67°C for N436A (blue), and 61°C for F439A mutant (red). B, N436A retains transport of Hoechst 33342 (black) and PhA (white), whereas F439A has lost transport ability. Each bar represents the mean + SE (n = 8). C, F439 contribution to binding of various ligands to ABCG2. Representative western blots are shown on the left side and the densitometry quantification data from three independent experiments are shown on the right side. Membrane vesicles prepared from mAbcg2-KO MEFs expressing human ABCG2 (WT and F439A) were incubated with 10  $\mu$ M of ABCG2 ligand for 60 min or 1 mM of AMP-PNP for 30 min at 37°C. Samples were heated for 3 min at 65°C for WT-ABCG2 (white) and 61°C for F439A mutant (red). D, Best flexible alignment of all ABCG2 ligands after a 48 hr alignment run in MOE, was used to compute the consensus pharmacophore model that had two aromatic features at 6.4 Å to each other. E, Docked Lapatinib molecule mapped onto the consensus pharmacophore model, showing the connection between the docking model and the consensus pharmacophore model. Don = H-bond donor; Acc = H-bond acceptor; PiN = Ring normal; Aro = Aromatic; PiR = pi-Ring; Hyd = Hydrophobic; Cat = Cation (use MOE reference)



these mutations on KI stabilization of ABCG2, the denaturation temperature we selected was 65°C for wildtype, 67°C for N436A, and 61°C for the F439A mutant. These denaturation temperatures, predicted to precipitate greater than 99% of the ABCG2 compared to unheated samples, were selected after interpolation from their respective denaturation curves. Each membrane preparation was assessed for their level of ABCG2 expression in the absence of thermal denaturation (to account for any variation not only in ABCG2 expression of the mutants, but also between each membrane preparation), and then when heated to a denaturation temperature, where 99% of ABCG2 precipitated (labeled “untreated”). Both ABCG2-F439A and N436A strongly reduced Ko143 stabilization of ABCG2. The N436A affected the thermal stabilization of ABCG2 by many KIs, but Lapatinib partially stabilized N436A, and none of the KIs were capable of stabilizing the F439A-ABCG2 mutant to the same extent as WT-ABCG2 (Figure 3A). Importantly, unlike a previous report<sup>11</sup> the N436A-ABCG2 mutant retained PhA and Hoechst 33342 transport activity similar to WT-ABCG2. In contrast, the F439A-ABCG2 mutant was incapable of transporting PhA and Hoechst 33342 (Figure 3B). These results suggest that F439 is key for the binding of these ligands to ABCG2.

The strong broad effect of F439A on KI induced thermal stabilization of ABCG2 led us to more generally evaluate additional ABCG2 interacting ligands of diverse molecular features (AST-1306, Tariquidar, Nobiletin, Apgenin, Pelitinib, WHI-P154, Curcumin, A-803467) (Figure 3C). As a positive control for ABCG2 binding, the non-hydrolyzable ATP analog, AMP-PMP, stabilized both WT-ABCG2 and F439A, similarly, indicating nucleotide binding was unimpaired by the F439A substitution (Figure 3C). Thus, while WT-ABCG2 was stabilized by every additional ligand, none were capable of inducing strong thermal stabilization of the F439A-ABCG2 (Figure 3C).

We investigated the ligands for common features that might be affected by the F439 mutation. This was accomplished by flexibly aligning the compounds in MOE and extracting features from the alignment to generate a consensus pharmacophore (Figure 3D). The consensus “global” pharmacophore model had two aromatic/hydrophobic centers separated by 6.4Å which indicated that the aromatic centers will, on average be separated by approximately a two carbon spacer. The model also contained two hydrogen bond donor/hydrogen bond acceptor centers. A docked Lapatinib molecule is

overlaid on this consensus pharmacophore to illustrate the key features of the model (Figure 3E). This model likely accounts for Arry-380 displaying both weaker ABCG2 binding and inhibition compared to Lapatinib (see Figure 2A,B), by suggesting that the aromatic benzene ring of Lapatinib is important for binding and that the heterocyclic ring of Arry-380 might clash, thereby restricting its binding. We were able to map 12 out of the 20 ABCG2 ligands (Table 2) directly onto its pharmacophoric features. The model also mapped 13 of the 19 docked poses (best pose from each of the 19 docked ABCG2 ligands), with the correspondence indicating that both the docking and the pharmacophore model captured the essential features of these ABCG2 ligands (list is in Table S2).

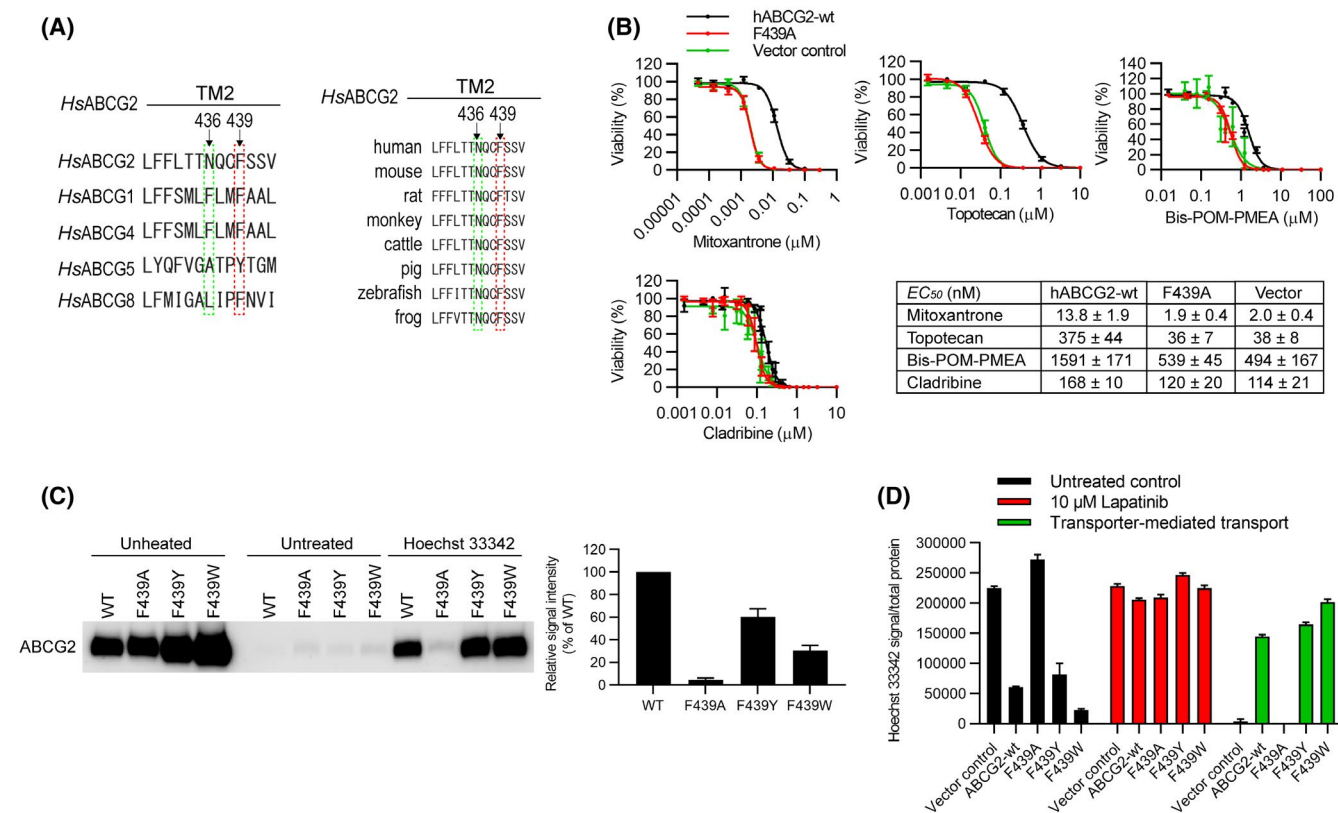
### 3.4 | F439 mutation affects substrate export with conservative amino acid substitutions restoring function

Among the ABCG family members, the position analogous to F439 in ABCG2 harbors an aromatic amino acid (Figure 4A, left) and as the sequence alignment shows, the amino acid F439 is conserved in ABCG2 orthologs from frogs to humans (Figure 4A, right). To further investigate the role of the amino acid F439 in substrate transport, we determined whether the F439A-ABCG2 was capable of conferring resistance (a proxy for transport) to a variety of cytotoxic drugs known to be exported by ABCG2.<sup>22–24</sup> Unlike WT-ABCG2, cells expressing F439A-ABCG2 were as sensitive as cells lacking ABCG2 to the cytotoxic effects of structurally and mechanistically diverse cytotoxic drugs, such as mitoxantrone, topotecan, Bis-POM-PMEA, and cladribine (Figure 4B) suggesting F439A-ABCG2 is incapable of transporting these drugs.

Because binding is the obligate initial step in the transport cycle and the aromatic acid residue is conserved among ABCG family members at positions analogous to F439 (Figure 4A), we tested if the substrate, Hoechst 33342 stabilized either ABCG2 or F439A-ABCG2. Importantly, based on ABCG5,<sup>25</sup> we created F439W-ABCG2 and F439Y-ABCG2 by substituting phenylalanine with either tyrosine or tryptophan in a human ABCG2 retroviral expression plasmids which were then subsequently transduced into *Abcg2*-knockout MEFs. We determined that Hoechst 33342 stabilized WT-ABCG2, F439W-ABCG2,

**TABLE 2** Ligands used to develop the global ABCG2 pharmacophore model

Ko143	Lapatinib	Nilotinib	Vemurafenib	Imatinib
Sunitinib	Sorafenib	Gefitinib	Tariquidar	Curcumin
AST-1306	WHI-P154	Apigenin	Pelitinib	Nobiletin
Hoechst 33342	Pheophorbide a	Mitoxantrone	Topotecan	4-[(2-phenylquinazolin-4-yl)amino]benzonitrile



**FIGURE 4** Mutation of the conserved binding residues rescues ABCG2 ligand-binding and transport. A, The human ABCG2 sequence containing the residues N436 and F439 was aligned with the corresponding sequences in its ABCG2 homologs (left) and orthologs (right). The residue F439 in ABCG2 are either conserved or substituted by other hydrophobic residue Y (ABCG5) in the other G-subfamily ABC transporters. The residue N436 in ABCG2 is not conserved in the other G-subfamily ABC transporters (left). The amino acid F439 is conserved in ABCG2 orthologs from frogs to humans (right). B, The cytotoxic effects ( $EC_{50}$  values) were determined by CellTiter-Glo luminescent cell viability assay of cytotoxic drugs Mitoxantrone, Topotecan, Bis-POM-PMEA, and Cladribine using the mAbcg2-KO MEFs expressing human ABCG2-WT (black), ABCG2-F439A (red), or vector control (green). Each bar represents the mean  $\pm$  SE ( $n = 5-10$ ). C, MEF-hABCG2-WT, F439A, F439Y, or F439W membrane vesicles were incubated with 10  $\mu\text{M}$  of Hoechst 33342 for 60 min at 37°C. Representative western blot is shown on the left side and the densitometry quantification data from four independent experiments is shown on the right side. D, 0.5  $\mu\text{M}$  of Hoechst 33342 was used as ABCG2 substrate and 10  $\mu\text{M}$  of Lapatinib was used as ABCG2 inhibitor to measure transport activity of ABCG2-WT, F439A, F439Y, and F439W. The cells were incubated with Hoechst 33342 (and Lapatinib) for 120 min. Each bar represents the mean  $\pm$  SE ( $n = 8$ )

and F439Y-ABCG2 but not F439A-ABCG2 (Figure 4C). We then investigated the role of an aromatic residue at this position in Hoechst 33342 transport. The uptake of Hoechst 33342 was determined for WT-ABCG2 (ie, F439), F439A-ABCG2, F439W-ABCG2, and F439Y-ABCG2 in the absence (black bar) or presence (red bar) of Lapatinib with the difference in uptake between the two conditions representing ABCG2 mediated transport (green bar) (Figure 4D). As expected, F439A-ABCG2 displayed no net ABCG2 mediated transport, whereas F439W-ABCG2 and F439Y-ABCG2 had transport activity similar to WT-ABCG2. As F439 is near the dimer interface, we tested the possibility that F439A-ABCG2 might be impaired in dimerization and compared it to WT-ABCG2, F439W-ABCG2, and F439Y-ABCG2. Non-denaturing gel electrophoresis revealed that F439A-ABCG2 formed a dimer like WT-ABCG2, F439W-ABCG2, and F439Y-ABCG2 (Figure S3A).

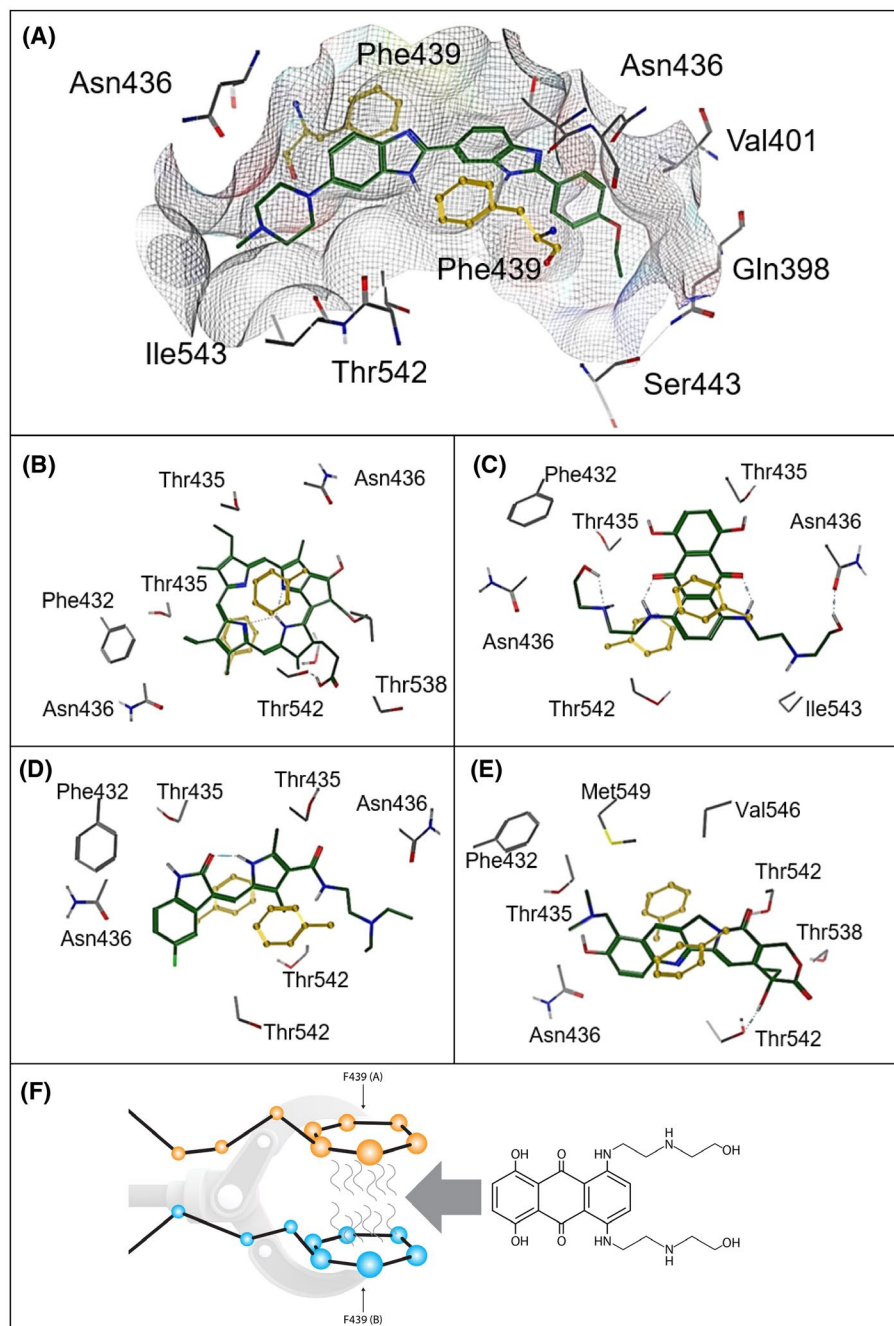
### 3.5 | Discussion

For the drug and xenobiotic transporting ABC transporter, ABCG2, recent cryo-electron microscopy structures have provided valuable snapshots of some substrates and inhibitors in the binding pocket (also termed cavity one), that apparently use the same residues to interact with ABCG2. It was unknown if one or more residues were required for binding to ABCG2. We selected some of these binding pocket residues for further studies and demonstrated, surprisingly, that a single highly conserved amino acid, F439 was essential for binding all the ligands tested. Importantly, F439 was also crucial for ABCG2 mediated substrate transport. Based on the structures of the ligands requiring F439 a unique pharmacophore model was developed that has the capability of predicting numerous drugs that interact with ABCG2. Further, in silico docking of ligands to a structure of the ABCG2

homodimer (Figure 5A-E), suggested that the two F439 aromatic rings “clamp” ligands in the binding pocket (Figure 5F). Importantly, conversion of this phenylalanine to amino acids with similar planar, aromatic side-chain rings, that is, tryptophan or tyrosine (which is also found in ABCG5 and essential for cholesterol transport<sup>25</sup>), retained binding, and transport activity comparable to wildtype ABCG2 (Figure 4C,D) which suggests that this aromatic “clamp” may be a

key feature for ligand engagement by other ABCG family members.

Our studies do not exclude the possibility that other residues in the binding pocket impact how ligands interact with ABCG2. Instead, we show that many structurally diverse ligands require F439 for optimal binding to ABCG2. This is consistent with the findings of Manolaridis et al<sup>11</sup> demonstrating F439A could not transport estrone-3-sulfate.



**FIGURE 5** Diagrammatic representation of structurally different ligands in the ABCG2 binding cavity and a depiction of the F439 interaction. A, Hoeschst 33342. B, Pheophorbide a. C, Mitoxantrone. D, Topotecan. E, Sunitinib. F439 (yellow) provides an anchoring point for ABCG2 ligands (A-E). F, “A” and “B” indicate F439 has been donated from separate ABCG2 monomers. The molecule that will interdigitate between the two F439 is the ABCG2 substrate and chemotherapeutic agent, mitoxantrone

However, our findings contrast with their findings, where N436A which was shown incapable of estrone-3-sulfate transport. Our studies demonstrate that N436A is nearly as capable of transporting PhA and Hoechst 33342 as WT-ABCG2. Thus, while N436A appears to have impaired interaction with some ligands, it is unlike F439A which has no demonstrable binding or transport activity which we speculate might be due to impaired interaction with a membrane constituent like cholesterol,<sup>26,27</sup> an important modulator of ABCG2 activity.<sup>28</sup> Perhaps the increased thermal stabilization by the alanine substitution at N436 alters ABCG2 conformation to preserve the interaction with certain ligands and not others. Based on the findings from G-protein coupled receptors (GPCRs), transmembrane alanine substitutions generally produced thermostabilization.<sup>29,30</sup> Indeed, this is the basis of a structural approach that has resulted in the structures of many GPCRs.

Studies of ABCG2 ligand binding have suggested that more than one drug/ligand binding pocket exists<sup>31,32</sup> and analysis of the outward- and inward-facing conformations of the ABCG2 homodimer supports this interpretation, with both an internal binding site and an extracellular surface binding site that also spans the two-fold symmetry axis, but residues near the extracellular surface.<sup>10,26</sup> Multiple binding pockets may occur in WT-ABCG2 and in the drug-selected, gain-of-function variant, R482G-ABCG2, that reportedly has more than one drug binding site.<sup>33</sup> Multiple binding pockets may be a general property of membrane transporters as multidrug pumps interact with ligands through both surface binding sites and buried sites (eg, AcrAB/TolC).<sup>34,35</sup> Consistent with the structural model, cellular studies have predicted that mitoxantrone binds ABCG2 through two asymmetric binding sites, a surface exposed site and a buried, second binding site.<sup>31,33,36</sup> Other studies, have also proposed that ABCG2 drug binding occurs through at least two binding sites.<sup>32</sup> From the ABCG2 structure, one might infer that these two binding sites are so called “cavity one” and “cavity two”.<sup>26</sup> It is likely that cavity one and two have different affinity for ligands. Importantly, our studies do not preclude the existence of two distinct ligand binding sites in ABCG2, but instead suggest that the ABCG2 binding site in cavity one (harboring F439) is the preferred site, and the one required for maximal ABCG2 interaction with a broad array of ligands, both substrates and inhibitors. Furthermore, based upon our transport studies with Hoechst 33342 (and the drug resistance studies (Figure 4B) as a proxy for transport), we propose that the substrate translocation pathway first requires binding to cavity one and that this is mediated by an obligate interaction with F439 which is sequentially followed by the opening of the “leucine plug”,<sup>26,37</sup> and substrate movement into cavity two, ultimately leading to substrate release. It is notable that inhibitors also require F439A for binding to cavity one. This suggests that inhibitors, such as Ko143 and tariquidar

compete directly with substrates to prevent their initial binding, regardless of whether they ultimately become “wedged” by additional contacts, as was proposed,<sup>11</sup> will require future kinetic and affinity binding studies to elucidate such an important relationship.

The identification of a key phylogenetically conserved amino acid, essential to ABCG2 interaction with a diverse array ligands was indispensable to the creation of structural models depicting the interaction of multiple substrates and inhibitors (Figure 5). This empirical knowledge was essential to the creation of unique pharmacophore and docking models. Our global pharmacophore model can be utilized for virtual screens to identify new and novel substrates and inhibitors of ABCG2 which will be invaluable in prioritizing the development of drugs that avoid an interaction with ABCG2. Indeed, although defects in ABCG2 are important causes of gout, due to defective uric acid transport, the ability of ABCG2 to transport uric acid<sup>4-6,38,39</sup> seems counterintuitive as it is a small, soluble molecule compared to the large hydrophobic drugs such as the KIs that seem more frequently thought of as ABCG2 substrates. However, when tested, our docking and pharmacophore model suggest that the F439 clamp can  $\pi$ - $\pi$  stack with the uric acid heterocyclic rings (Figure S6). Indeed, uric acid was predicted by the top poses to interact with ABCG2 in a location between the two F439 residues. In total, these findings not only provide a novel basis for ABCG2 interacting with multiple ligands (one that was not obvious or easily inferred from the recent structural findings), but also provides mechanistic insights that suggest that in a native cellular environment, substrates must first interact with a single-amino acid, F439, which appears to be a pre-requisite to initiating transport. This is in contrast to N436A, which is still capable of transport. These findings have therapeutic important implications toward the efforts to predict and identify drugs that can evade ABCG2 which are particularly important for drugs attempting to penetrate into tumors or cross the blood-brain barrier.<sup>3,40-42</sup>

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## ACKNOWLEDGMENTS

This work was supported by NIH grants R01CA194057 (JS), R35-GM118041 (TC), P30CA21745, CA21865, CA96832, and by ALSAC. We thank Diana Zhou for her excellent assistance with some of the experiments contained herein. We thank the members of the Schuetz labs for constructive comments and edits.

## AUTHOR CONTRIBUTIONS

J. D. Schuetz was the principal investigator of this study and conceived the research together with T. Gose, T. Shafi, Y. Fukuda, S. Das, A. Shelat, and R. C. Ford; T. Gose performed



a majority of these studies with advice and assistance from J. D. Schuetz, T. Shafi, Y. Fukuda, S. Das, Y. Wang, A. Allcock, A. Gavan McHarg, J. Lynch, A. Shelat, and R. C. Ford; T. Shafi and R. C. Ford performed in silico molecular docking experiments; S. Das and A. Shelat created a pharmacophore model; J. D. Schuetz wrote the manuscript along with T. Gose, with input from T. Shafi, Y. Fukuda, S. Das, A. Shelat, I. Tamai, and R. C. Ford; All authors discussed the results and contributed to the final manuscript.

## REFERENCES

- Zhou S, Schuetz JD, Bunting KD, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med*. 2001;7:1028-1034.
- Zhou S, Morris JJ, Barnes Y, Lan L, Schuetz JD, Sorrentino BP. Bcrp1 gene expression is required for normal numbers of side population stem cells in mice, and confers relative protection to mitoxantrone in hematopoietic cells in vivo. *Proc Natl Acad Sci*. 2002;99:12339-12344.
- Enokizono J, Kusuhara H, Ose A, Schinkel AH, Sugiyama Y. Quantitative investigation of the role of breast cancer resistance protein (Bcrp/Abcg2) in limiting brain and testis penetration of xenobiotic compounds. *Drug Metab Dispos*. 2008;36:995-1002.
- Ichida K, Matsuo H, Takada T, et al. Decreased extra-renal urate excretion is a common cause of hyperuricemia. *Nat Commun*. 2012;3:764-767.
- Dehghan A, Köttgen A, Yang Q, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet*. 2008;372:1953-1961.
- Hosomi A, Nakanishi T, Fujita T, Tamai I. Extra-renal elimination of uric acid via intestinal efflux transporter BCRP/ABCG2. *PLoS One*. 2012;7:e30456.
- Fukuda Y, Wang Y, Lian S, et al. Upregulated heme biosynthesis, an exploitable vulnerability in MYCN-driven leukemogenesis. *JCI Insight*. 2017;2:e92409.
- Shukla S, Kouanda A, Silverton L, Talele TT, Ambudkar SV. Pharmacophore modeling of nilotinib as an inhibitor of ATP-binding cassette drug transporters and BCR-ABL kinase using a three-dimensional quantitative structure-activity relationship approach. *Mol Pharm*. 2014;11:2313-2322.
- Hussaarts KGAM, Veerman GDM, Jansman FGA, van Gelder T, Mathijssen RHJ, van Leeuwen RWF. Clinically relevant drug interactions with multikinase inhibitors: a review. *Ther Adv Med Oncol*. 2019;11:1758835918818347.
- Jackson SM, Manolaridis I, Kowal J, et al. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol*. 2018;25:333-340.
- Manolaridis I, Jackson SM, Taylor NMI, Kowal J, Stahlberg H, Locher KP. Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature*. 2018;563:426-430.
- Robey RW, Steadman K, Polgar O, et al. Pheophorbide a is a specific probe for ABCG2 function and inhibition. *Cancer Res*. 2004;64:1242-1246.
- Yasuda K, Ganguly S, Schuetz EG. Pheophorbide A: fluorescent bcrp substrate to measure oral drug-drug interactions in real-time in vivo. *Drug Metab Dispos*. 2018;46:1725-1733.
- Allen JD, van Loevezijn A, Lakhai JM, et al. Potent and specific inhibition of the breast cancer resistance protein multidrug transporter in vitro and in mouse intestine by a novel analogue of fumitremorgin C. *Mol Cancer Ther*. 2002;1:417-425.
- Reinhard FBM, Eberhard D, Werner T, et al. Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat Methods*. 2015;12:1129-1131.
- Molina DM, Jafari R, Ignatushchenko M, et al. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science*. 2013;341:84-87.
- Martinez Molina D, Nordlund P. The cellular thermal shift assay: a novel biophysical assay for in situ drug target engagement and mechanistic biomarker studies. *Annu Rev Pharmacol Toxicol*. 2015;56:141-161.
- Ashok Y, Nanekar R, Jaakola VP. Defining thermostability of membrane proteins by western blotting. *Protein Eng Des Sel*. 2015;28:539-542.
- Fukuda Y, Cheong PL, Lynch J, et al. The severity of hereditary porphyria is modulated by the porphyrin exporter and Lan antigen ABCB6. *Nat Commun*. 2016;7:12353.
- Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*. 1980;36:3219-3228.
- Jafari R, Almqvist H, Axelsson H, et al. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc*. 2014;9:2100-2122.
- Takenaka K, Morgan JA, Scheffer GL, et al. Substrate overlap between Mrp4 and Abcg2/Bcrp affects purine analogue drug cytotoxicity and tissue distribution. *Cancer Res*. 2007; 67:6965-6972.
- Nagai S, Takenaka K, Nachagari D, et al. Deoxycytidine kinase modulates the impact of the ABC transporter ABCG2 on clofarabine cytotoxicity. *Cancer Res*. 2011;71:1781-1791.
- Fukuda Y, Schuetz JD. ABC transporters and their role in nucleoside and nucleotide drug resistance. *Biochem Pharmacol*. 2012;83:1073-1083.
- Lee JY, Kinch LN, Borek DM, et al. Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature*. 2016;533:561-564.
- Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H, Locher KP. Structure of the human multidrug transporter ABCG2. *Nature*. 2017;546:504-509.
- Jackson SM, Manolaridis I, Kowal J, et al. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol*. 2018;25:333-340.
- Pál A, Méhn D, Molnár E, et al. Cholesterol potentiates ABCG2 activity in a heterologous expression system: improved in vitro model to study function of human ABCG2. *J Pharmacol Exp Ther*. 2007;321:1085-1094.
- Vaidehi N, Grisshammer R, Tate CG. How can mutations thermostabilize G-protein-coupled receptors? *Trends Pharmacol Sci*. 2016;37:37-46.
- Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. Conformational thermostabilization of the 1-adrenergic receptor in a detergent-resistant form; 2008;105:877-882.
- Cox MH, Kapoor P, Briggs DA, Kerr ID. Residues contributing to drug transport by ABCG2 are localised to multiple drug-binding pockets. *Biochem J*. 2018;105:877-882.
- Ferreira RJ, Bonito CA, Cordeiro MNDS, Ferreira MJU, Dos Santos DJVA. Structure-function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. *Sci Rep*. 2017;7:1-17.

33. Clark R, Kerr ID, Callaghan R. Multiple drugbinding sites on the R482G isoform of the ABCG2 transporter. *Br J Pharmacol*. 2006;149:506-515.
34. Elkins CA, Mullis LB. Substrate competition studies using whole-cell accumulation assays with the major tripartite multidrug efflux pumps of *Escherichia coli*. *Antimicrob Agents Chemother*; 2007;51:923-929.
35. Wang Z, Fan G, Hryc CF, et al. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife*. 2017;6:e24905.
36. Haider AJ, Cox MH, Jones N, et al. Identification of residues in ABCG2 affecting protein trafficking and drug transport, using co-evolutionary analysis of ABCG sequences. *Biosci Rep*. 2015;6:e24905.
37. Khunweeraphong N, Stockner T, Kuchler K. The structure of the human ABC transporter ABCG2 reveals a novel mechanism for drug extrusion. *Sci Rep*. 2017;7:1-15.
38. So A, Thorens B. Uric acid transport and disease. *J Clin Invest*. 2010;120:1791-1799.
39. Matsuo H, Takada T, Ichida K, et al. Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. *Sci. Transl. Med*. 2009;120:1791-1799
40. Morfouace M, Cheepala S, Jackson S, et al. ABCG2 transporter expression impacts group 3 medulloblastoma response to chemotherapy. *Cancer Res*. 2015; 75:3879-3889.
41. Robey RW, Pluchino KM, Hall MD, Fojo AT, Bates SE, Gottesman MM. Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat Rev Cancer*. 2018;18:452-464.
42. Wijaya J, Fukuda Y, Schuetz JD. Obstacles to brain tumor therapy: Key ABC transporters. *Int J Mol Sci*. 2017;18:2544.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Gose T, Shafi T, Fukuda Y, et al. ABCG2 requires a single aromatic amino acid to “clamp” substrates and inhibitors into the binding pocket. *The FASEB Journal*. 2020;34:4890–4903. <https://doi.org/10.1096/fj.201902338RR>