Bioinformatics-guided approaches to membrane protein structure determination as applied to ABC transporters

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

2022

Iqra Younus

School of Biological Science

List of Contents

Word count: 64510

List of Figures

List of Tables

List of Abbreviations

Å	Angstrom
ABC	ATP binding cassette
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
BSA	bovine serum albumin
bbABCG1	Little brown bat ABCG1
CFTR	Cystic fibrosis transmembrane conductance regulator
CV	column volume
2D	Two dimensional
3D	Three dimensoinal
DDM	n-Dodecyl β-D-maltoside
EM	electron microscopy
$\bar{E}$	Electrons
gpABCB6	Giant panda ABCB6
GFP	green fluorescent protein
g	gravitational force
His	Histidine
IMAC	immobilized metal affinity chromatography
K_d	dissociation constant
kDa	kilo Dalton
M	Molar
mABCB5	Mouse ABCB5
mABCG4	Mouse ABCG4
mM	milliMolar
NBD	nucleotide biding domain
OD	optical density
PAGE	polyacrylamide gel electrophoresis
P_i	inorganic Phosphate
RPM	rotation per minute
SDS	sodium dodecyl sulphate
SLS	static light scattering
SOC	Super optimal broth with catabolite repression
T_m	melting temperature
TMD	transmembrane domain
UV	ultra violet
v/v	volume by volume
w/v	weight by Volume
YNB	yeast nitrogenous base
YPD	Yeast peptone dextrose

Abstract

Adenosine triphosphate binding cassette (ABC) transporters form a ubiquitous superfamily of integral membrane proteins involved in the translocation of substrates across membranes. Substrate translocation is powered by ATP. In this study, four medically important classes of ABC transporters were chosen for structural studies: namely, subfamily B, member 5 (ABCB5), subfamily B, member 6 (ABCB6), subfamily G, member 1 (ABCG1), subfamily G, member 4 (ABCG4). A bioinformatics approach was used for ortholog selection of representative ABC transporters to improve the process of structure determination. Selected orthologs were expressed in *Saccharomyces cerevisae* with polyhistidine affinity and Green Fluorescent protein identification tags and purified in the presence of the detergent, n-Dodecyl-β-D-Maltoside. A battery of tests was applied to explore the quality of purified proteins. An assay using a thiol-specific dye, and intrinsic protein fluorescence was used to assess the thermal stability of proteins. ATPase activity, measured in a detergent environment, showed that all of the proteins hydrolysed ATP. A membrane thermal shift experiment demonstrated that mouse ABCG4 was stabilized by tyrosine kinase inhibitors. Mouse ABCG4 was also the subject of cryoEM imaging. Taken together, this project throws light on the use of computational tools for target selection before conducting structural studies. All target proteins were successfully expressed in yeast. Protein purification showed mixed results; however, all proteins could be enriched. All purified protein targets were functionally active. The assays could form a platform for screening compounds that stabilise the proteins. Stabilized proteins could subsequently go into crystal trials.

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

Thank you **Allah Almighty** for maintaining a continuous hope in life. Thanks for unveiling truth and beautiful things you blessed me with. Tremulous venerations are for His **Holy Prophet Hazrat Mohammad (PBUH)**, who is the everlasting torch of guidance and knowledge for humanity.

I want to express gratitude to my supervisors and mentors, **Doctor Stephen M. Prince** and **Professor Robert C. Ford** for continuous support and guidance. Best supervisors I ever met. Indeed, the dedication of **Bob** and **Steve** was a source of inspiration for me throughout this tenure. Thank you for enlightening my thoughts.

I would like to thank the **University of Manchester** for providing me with the platform and facilities to conduct this project, especially, my co-supervisor **Doctor Richard Collins** for his valuable suggestions, discussions, and most importantly for assisting me with cryo-EM experiments. My grateful thank extends to **Aleksandr Mironov** (who helped me in negative staining electron microscopy), Bioimaging facilities, members of the University of Manchester Biomolecular analysis facilities for providing equipments, especially thanks to **Dr. Thomas Jowitt** for giving me advice on protein characterisation, and **Michael Lockhart** for assisting me with SAXS data, Diamond Light Source (DLS) Ltd (Harwell, UK) for providing the opportunity to work on the B21 beamline.

I would also like to thank my colleagues and friends at the Ford and Steve lab especially **Jack Clews, Talha Shafi, Alessandro Barbieri, Tracy Remington, Usamae Aktas, Mohammad, Bowen Chen**, **and Sungo Han Bosco** for providing pieces of advice, kind help and always being supportive. Furthermore, I want to acknowledge the Punjab Education Endowment Fund (PEEF) for financing my study at The University of Manchester. Finally, Bundle of thanks to **my Parents and all my family members** for their love, moral support and continuous backing throughout my study and my life.

Chapter 1

Introduction

Membrane proteins (MPs) are one of the largest protein families that interact with or are a part of, biological membranes. In both prokaryotes and eukaryotes, about 20-30% of all genes encode for MPs (Wagner et al., 2006, Almeida et al., 2017, Wallin and Heijne, 1998), whereas nearly 23% of the human proteome comprises membrane proteins (Uhlén et al., 2015). Typically, membrane proteins are classified into two categories *i.e.* (i) Peripheral membrane proteins (PMPs) and (ii) Integral membrane proteins (IMPs) (Uzman, 2001). PMPs do not integrate into the membrane, they are loosely attached to the phospholipid bilayer or other proteins through non-covalent interactions such as hydrogen bonds or ionic bonds; thus they play roles in transportation, anchoring, and cell signalling. Integral membrane proteins (IMPs), are embedded in the lipid bilayer and mostly serve as carrier proteins, enzymes, and channels. Overall, most significant function of membrane proteins is as receptor. As a receptors, MPs enable communication between the cell and the external environment by interacting with particular ligands (such as hormones and neurotransmitters) (Cuatrecasas, 1974).

Furthermore, the structural diversity of integral membrane proteins is large, ranging from monotopic helical structures to polytopic channels, and they can have both helical structures and β-barrels (Figure 1.1). Because of this diversity in the physical and biochemical properties, membrane proteins have always been a priority for researchers. Moreover, a large number of pharmaceutical drugs (>50%) target membrane proteins (Terstappen and Reggiani, 2001a, Davey, 2004). Therefore, a better understanding of the structure of membrane proteins will help in drug discovery with improved pharmaceutical properties and selectivity.
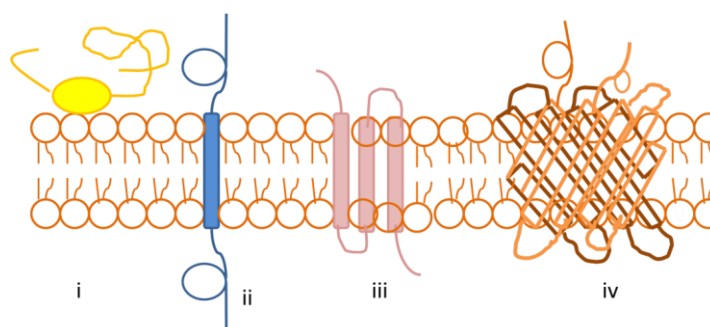
Figure 1.1: Diagrammatic representation of membrane proteins. (i) Peripheral membrane protein (ii) Monotopic IMP (iii) Polytopic IMP alpha helices (iv) Polytopic IMP beta sheets.

1.1 ABC Transporter Family: Structure and Transport Mechanism

ATP binding cassette proteins (ABC) form a ubiquitous and diverse superfamily of membrane proteins, with examples found in all organisms ranging from prokaryotes to higher eukaryotes (Higgins, 1992). The majority of ABC genes encode membrane-bound proteins that contribute to the transportation of different types of molecules across organelle and cytoplasmic membranes (Higgins, 1992). ABC transporters can be classified as exporters or importers, based on the direction of transport relative to the cytoplasm (Dassa and Bouige, 2001, Saurin et al., 1999). In addition to import and export functions, ABC transporters also function as channels or receptors and contribute to other biological processes as well. In prokaryotes, they are either exporters or importers but in eukaryotes; only exporters are known (Dassa and Bouige, 2001). ABC genes are necessary for several processes in the cell and mutations in these genes contribute to different genetic diseases (Table 1.1) (Dean et al., 2001).

Table 1.1: Importance of ABC family. Diseases marked with * are reported on the gene card. Other diseases are taken from the Uniprot database.

Protein	No of Amino Acids	Mol.wt (kDa)	Functions	Associated Diseases
ABCA1	2261	254.3	Cholesterol efflux onto HDL	• High-density lipoprotein deficiency 1 • High-density lipoprotein deficiency 2
ABCA2	2436	269.8	Macrophage lipid metabolism and neural development.	• Intellectual developmental disorder with poor growth With or without seizures* • Acoustic neuroma*
ABCA3	1704	191.4	Multidrug resistance	• Pulmonary surfactant metabolism dysfunction 3
ABCA4	2273	255.9	N-retinylidene-phosphatidylethanolamine (PE) efflux	• Stargardt disease • Retinitis pigmentosa 19 • Age-related macular degeneration • Fundus flavimaculatus • Cone-rod dystrophy 3

ABCA5	1642	186.5	A urinary diagnostic marker for prostatic intraepithelial neoplasia (PIN)	• Hypertrichosis, Congenital generalized with or without gingival hyperplasia* • Lysosomal diseases*
ABCA6	1617	184.3	Macrophage lipid homeostasis	• Autosomal recessive congenital ichthyosis*
ABCA7	2146	234.4	Cholesterol efflux	• Alzheimer disease 9
ABCA8	1581	179.3	Transports certain lipophilic drugs	• Autosomal recessive congenital ichthyosis*
ABCA9	1624	184.4	Might play a role in monocyte differentiation and macrophage lipid homeostasis	• N/A
ABCA10	1543	175.8	Cholesterol-responsive gene	• Donnai-barrow syndrome *
ABCA12	2595	293.2	Might involve in lipid homeostasis and also has implications for prenatal diagnosis	• Ichthyosis, congenital, autosomal recessive 4A • Ichthyosis, congenital, autosomal recessive 4B
ABCA13	5058	576.2	Inherited disorder affecting the pancreas	• Schizophrenia* • Stargardt disease*
ABCB1	1280	141.5	Multidrug resistance	• Inflammatory bowel disease 13 • Colchicine resistance*
ABCB2/ TAP1	808	87.2	Peptide transport	• Bare lymphocyte syndrome 1
ABCB3/ TAP2	703	75.6	Peptide transport	• Bare lymphocyte syndrome 1
ABCB4	1279	141.5	Phosphatidylcholine (PC) transport	• Cholestasis, progressive familial intrahepatic, 3 • Cholestasis of pregnancy, intrahepatic 3 • Gallbladder disease 1
ABCB5	812	138.6	Melanogenesis	• Borna disease* • Melanoma*
ABCB6	842	93.8	Iron transport/ Heme synthesis	• Dyschromatosis universalis heredetaria (DUH)

				• Microphthalmia • Familial pseudo hyperkalemia (FPH)
ABCB7	753	82.6	Fe/S cluster transport	• Anemia, sideroblastic, spinocerebellar ataxia
ABCB8	718	70.9	Intracellular peptide trafficking across membranes	• Abcd syndrome* • Anemia, sideroblastic, spinocerebellar ataxia*
ABCB9	766	84.5	Located in lysosomes	• Bile acid synthesis defect, Congenital, 5* • Bare lymphocyte syndrome 1*
ABCB10	738	79.2	Export of peptides derived from proteolysis of inner-membrane proteins/ Heme biosynthesis	• Developmental coordination disorder • Stereotypic movement disorder
ABCB11	1321	146.4	Bile salt transport	• Progressive familial intrahepatic cholestasis
ABCC1	1531	171.6	Drug resistance	• Dubin–Johnson syndrome* • Pseudoxanthoma elasticum*
ABCC2	1545	174.2	Organic anion efflux	• Dubin–Johnson syndrome • Bilirubin metabolic disorder*
ABCC3	1527	169.3	Transport of biliary and intestinal excretion of organic anions	• Dubin–Johnson syndrome* • Extrahepatic syndrome*
ABCC4	1325	149.5	Nucleoside transport	• Biliary tract disease* • Dubin-Johnson syndrome*
ABCC5	1437	160.6	Nucleoside transport	• Primary angle-closure glaucoma* • Episodic kinesigenic dyskinesia 1*
ABCC6	1503	164.9	Active transport of drugs directly or indirectly,	• Pseudoxanthoma elasticum

				• Arterial calcification of infancy
ABCC7/ CFTR	1480	168.5	Chloride ion channel (same as CFTR gene in cystic fibrosis)	• Cystic fibrosis • Congenital bilateral absence of vas deferens
ABCC8	1581	176.9	Sulfonylurea receptor	• Familial hyperinsulinemic hypoglycaemia 1 • Leucine-induced hypoglycaemia • Transient neonatal diabetes mellitus 2 • Diabetes mellitus permanent neonatal
ABCC9	1549	174.2	Encodes the regulatory SUR2A subunit of the cardiac $K^+(ATP)$ channel	• Cardiomyopathy, dilated 10 • Atrial fibrillation, familial, 12 • Hypertricotic osteochondrodysplasia
ABCC10	1464	161.6	Multidrug resistance	• Borna disease* • Abcd syndrome*
ABCC11	1382	154.3	Drug resistance in breast cancer	• Apocrine gland secretion variation in* • Lateral sinus thrombosis*
ABCC12	1359	152.3	Multidrug resistance	• Episodic Kinesigenic dyskinesia 1* • Dubin-Johnson syndrome*
ABCD1	745	82.9	Very-long-chain fatty acid (VLCFA) transport	• Adrenoleukodystrophy • Hypoadrenocorticism, Familial*
ABCD2	740	83.2	Major modifier locus for clinical diversity in X-linked ALD (X-ALD)	• Adrenoleukodystrophy • Zellweger syndrome
ABCD3	659	75.5	Involved in import of fatty acids and/or fatty acyl-coenzyme into the peroxisome	• Congenital bile acid synthesis defect 5

ABCD4	606	68.6	Lysosomal transporter	• Methylmalonic aciduria and homocystinuria type cbIJ
ABCG1	678	75.6	Cholesterol transport	• Tangier disease* • Sitosterolaemia*
ABCG2	655	72.3	Toxicant efflux, drug resistance	• Uric acid concentration, serum, quantitative trait locus 1 *
ABCG4	646	71.9	Macrophage lipid homeostasis	• Abcd syndrome* • Sitosterolemia*
ABCG5	651	72.5	Sterol transport	• Sitosterolemia 2
ABCG8	673	75.7	Sterol transport	• Gallbladder disease 4 • Sitosterolemia 1

Up till now, 48 human ABC genes are characterized. Based on amino acid homology and domain organization, these genes are divided into seven subfamilies (A-G) (Dean et al., 2001). ABC proteins consist of the cytosolic nucleotide-binding domain (NBD), which binds and hydrolyses ATP providing energy for transport in most cases. The NBD contains several conserved regions *i.e.* Walker A and B motifs, H and Q loop, and ABC signature motif (Higgins and Linton, 2004). Besides NBDs, ABC transporters are also comprised of low homology transmembrane domains (TMDs), each contains a number of hydrophobic α-helices. The TMD recognizes the substrate and translocate this across the lipid bilayer by forming a translocation pore. All NBDs share the same evolutionary origin and mechanism, while the amino acid sequence of TMDs differs considerably thus potentially exhibiting important differences in molecular mechanism (Higgins and Linton, 2004). ABC transporters are classified into two categories (i) Full transporters: comprised of two TMDs and two NBDs (ii) Half transporters: which contain only one unique TMD and one unique NBD. X-ray structure analysis of bacterial ABC transporters revealed that a minimum of two NBDs are needed for ATP binding and hydrolysis suggesting that the half transporters assemble as either homo or hetero-dimers to form a functional transporter (Smith et al., 2002, Rees et al., 2009, Beis, 2015, Boswell-Casteel et al., 2018).

Various mechanisms have been proposed for ATP-driven import or export through ABC transporters based on the structural data together with biochemical and genetic characterizations. One of the first mechanisms proposed was the sequential binding model,

where, ATP binding to one NBD triggered ATP hydrolysis at the other NBD *i.e.* sequential ATP binding and hydrolysis provides energy to drive substrate transport (Figure 1.2 A) (Senior et al., 1995). This was followed by the "ATP-switch" model, where ATP binding alone (not hydrolysis) provides enough energy to drive transport (Figure 1.2 B) (Higgins and Linton, 2004). Recently, the "ATP-switch" model has been further modified into variations of the "Alternating Access" model. This pathway is initiated with an inward open configuration. The substrates bind to ATP-free transporter. The ATP binding provides the "power stroke" behind translocation and the transporter moves into a short-lived occluded state bound to both ATP and substrate. The first ATP hydrolysis is responsible for outward conformation and substrate expulsion, whereas, the second ATP hydrolysis restores the transporter to its original form (Figure 1.2 C) (Zou et al., 2009, Zou and Mchaourab, 2009). It is still ambiguous whether the ATP molecule binds to NBDs before substrate binding or after substrate binding (Locher, 2016, Xu et al., 2022, Wang et al., 2022).

ABC proteins also have measurable rates of ATP hydrolysis even in a substrate-free basal state (Lerner-Marmarosh et al., 1999). Therefore, it would seem that if ABC proteins could bind ATP in its ground state, then there would be an inappropriately large amount of ATP hydrolysis without any substrate transport, which would be a great loss of cellular energy. Perhaps, there are more complex mechanisms that keep the NBDs in a hydrolytically inactive state until substrate transport is required. ABCB10 crystal structures trapped in both apo- and nucleotide-bound states and recently published data of ABCG1 and ABCD1 support the ATP switch model which provides the most consistent explanation to describe the general mechanisms of the ABC family (Shintre et al., 2013, Xu et al., 2022, Wang et al., 2022). Additionally, it is also possible that the individual subtypes of the ABC family may have adapted mechanisms optimized for the transportation of specific substrates in their niche but further structural and biochemical data are needed to elucidate the mechanistic details.
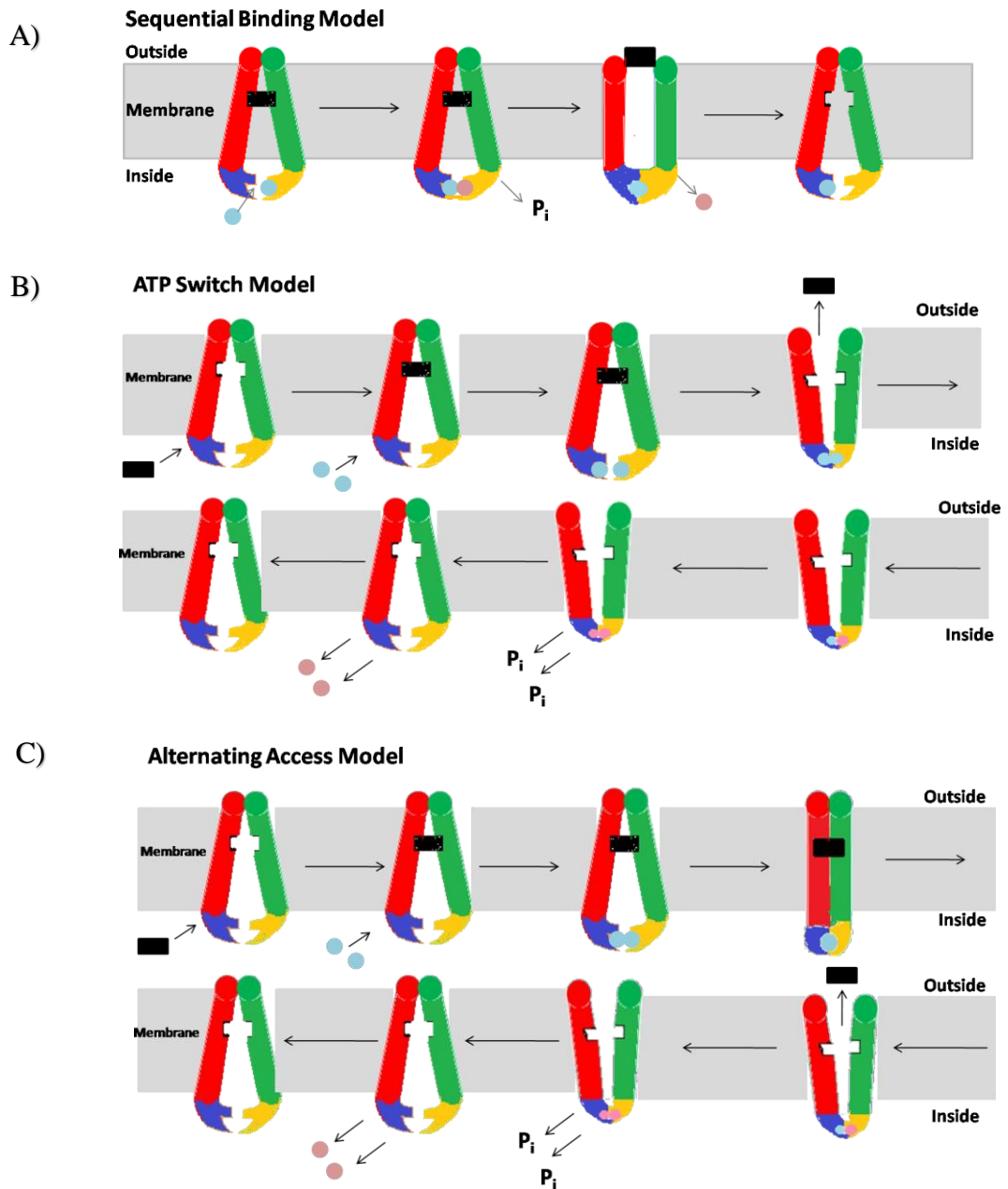
Figure 1.2: Mechanism of ABC translocation. The color scheme is as **TMD1:** Red; **TMD2:** Green, **NBD1:** Blue, **NBD2:** Yellow, **ATP:** Light blue; **ADP:** Pink, **Substrate:** Black**.**
A) Sequential binding model. Catalytic site 1 has bound ATP, while catalytic site 2 is empty. ATP binding at catalytic site 2 allows hydrolysis at site 1. The conformation switch occurs from an inward open to an outward open allowing extracellular substrate release. **B) ATP-switch model:** The transport cycle initiates after binding of substrate to an intracellular site on TMDs that increases the NBDs affinity for ATP and facilitates NBD dimer formation. Two ATP molecules bind cooperatively. A conformational change from closed inward to open outward allows substrate expulsion to an extracellular site. Sequential ATP hydrolysis followed by sequential P_i release resets the protein into an open-inward facing conformation. **C) Alternating access model:** Substrate binds to an inward-open conformation like the ATP-switch model. Two molecules of ATP bind cooperatively. ABC protein enters into an occluded state, sequestering both ATPs and the substrate. Hydrolysis of one ATP drives substrate expulsion while the second hydrolysis restores the protein to basal configuration, ready to accept another transport ligand molecule.

The four ABC transporter classes chosen for study in this research are introduced here.

1.2 ATP Binding Cassette, subfamily B, member 5 (ABCB5)

ATP binding cassette, sub-family B, member 5 (ABCB5), a full-length transporter is one of the 11 members of the B branch of the ABC transporter superfamily (Wilson et al., 2014). According to phylogenetic analysis, ABCB5 is most closely related to the other full transporters of the ABCB subfamily *i.e.* multidrug transporter P-glycoprotein (ABCB1), phosphatidylcholine transporter (ABCB4), and the bile salt exporter pump (ABCB11) (Moitra et al., 2011). The amino acid sequences of ABCB1 and ABCB5 are highly homologous; the amino acid similarity is 89.8% (Kawanobe et al., 2012) and the amino acid identity is 55.0% (Kawanobe et al., 2012). There are four main isoforms of ABCB5 in humans (Moitra et al., 2011) and these distinct ABCB5 isoforms are expressed in several tissue types, including, melanocyte, mammary tissue, retinal pigmented epithelium, and testis (Frank et al., 2003, Frank et al., 2005, Huang et al., 2004, Chen et al., 2005). Full-length isoform ABCB5.ts, which is expressed in the human testis, has 1257 amino acids and a molecular weight of 138.6 kDa. It is considered a functional transporter and like other conventional full-transporters, it contains 2 nucleotide-binding domains (NBDs) and 2 transmembrane domains (TMDs) (Moitra et al., 2011) (Figure 1.3). The ABCB5 full transporter gene is located on chromosome 7p21 and has 28 exons and 141.8 kb of genomic DNA (Moitra et al., 2011).
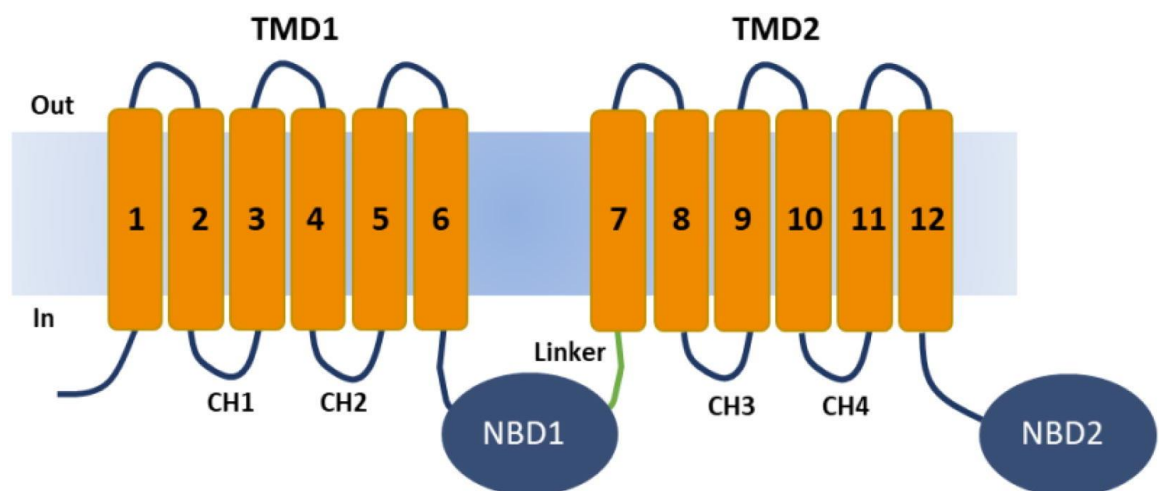
Figure 1.3: Membrane topology of full-length human ABCB5. The full-length ABCB5 is consist of two TMDs and two cytosolic NBDs connected by a linker region. Each TMD consists of six α-helices spanning in and out of the membrane connected by extracellular loops and coupling helices (CH). **NBD** = nucleotide-binding domain; **TMD** = transmembrane domain (Tangella et al., 2021).

ABCB5 is hypothesised to be involved in drug efflux. This was supported by studies determining the intracellular accumulation of doxorubicin or Rhodamine 123 in melanoma (Frank et al., 2003) and hepatocarcinoma cells (Cheung et al., 2011b, Cheung et al., 2011a). Evidence from the literature suggests that blocking ABCB5 at the protein or mRNA level can inhibit the transport of the chemotherapeutic drug doxorubicin and can also make melanoma and liver cancer stem cells more resistant to the treatment (Frank et al., 2005, Arce et al., 2006). It is also suggested that ABCB5 is a marker for the onset of metastatic melanoma and that this transporter may play a role in the regulation of progenitor cell fusion (Frank et al., 2005, Schatton and Frank, 2008). Additionally, ABCB5 has been found to link with leukemia (Lehne et al., 2009) and in several cancer subtypes, such as malignant melanoma, hepatocellular carcinoma, breast cancer, and colorectal cancer (Frank et al., 2005, Schatton et al., 2008, Cheung et al., 2011b, Arce et al., 2006).

There is no structure of ABCB5 on the Protein Data Bank (PDB), yet. However, based on the homology modelling and molecular docking three drug-binding sites have been identified in full-length ABCB5 named site-1, site-2, and site-3 (Tangella et al., 2021). The proposed site-1 in human ABCB5 contains ~20% aromatic amino acids (namely Phe and Tyr), whereas site-2 is comprised of only 4% aromatic amino acids. In site-3, no aromatic residues were reported. Relatively more aliphatic nonpolar residues were identified at site-2. Overall, sites 1 and 2 are primarily made up of positively charged, non-polar, aromatic, and aliphatic residues, while site 3 is primarily made up of negatively charged, and polar residues (Tangella et al., 2021).

To date, no comprehensive studies have been carried out regarding the organization, genetic variations, mechanism of translocation, and evolution of this transporter, leaving a knowledge gap in the field regarding basic questions such as: How is the structure of this transporter laid out? Does ABCB5 function as a full transporter or by dimerization of half-transporter units? Finally, may this transporter's genetic variations have an impact on how it functions?

1.3 ATP Binding Cassette, subfamily B, member 6 (ABCB6)

ABCB6 gene is also termed as P-glycoprotein related protein (PRP) and MTABC3. It is localized on chromosome 2q35.5 (Boswell-Casteel et al., 2018). It encodes an 842 amino acid protein *i.e.* ABCB6. The ATP binding cassette Protein 6 (ABCB6) is a member of the B-subfamily of ABC transporters (Higgins, 1992). It was discovered in 1997 when screening

was performed to identify novel ABC transporter genes involved in drug resistance mechanisms in the liver (Lynch et al., 2009). Two distinct molecular weight forms of ABCB6 are reported (i) High molecular weight (93KDa) (ABCB6-H) and (ii) Low molecular weight (79KDa) (ABCB6-L) (Paterson et al., 2007).

ABCB6 is a half-transporter that forms a homo-dimer to carry out its function (Krishnamurthy et al., 2006, Chavan et al., 2013). ABCB6 consists of the ABC core domain and an additional N-terminal TMD that has been termed TMD0 (Figure 1.4) (Krishnamurthy et al., 2006, Emadi-Konjin et al., 2002). ABC core domain is comprised of TMD1 (206-557 A.A) and NBD1 (558-842 A.A). TMD0 (1-205 A.A) is linked to the ABC core domain through cytosolic linker L_0. This extra N-terminal domain (TMD0) is unique to ABCB6 orthologs as it does not show sequence homology to any other protein (Kiss et al., 2015). The arrangement order of these domains from N to C terminus is as: TMD0-L_0-TMD1-α1-NBD1 (Figure 1.4). Sequence alignment and transmembrane helix prediction algorithm suggest five transmembrane helices in TMD0 (Kiss et al., 2015). ABCB subfamily contains seven half transporters and interestingly all of these half transporters are expressed in intracellular organelles and have an additional, unique, and long N-terminal segment *i.e.* TMD0 (Tusnády et al., 2006). Deletion of TMD0 does not affect folding, membrane insertion, dimerization, ATP binding, and hydrolysis. Thus TMD0 does not play a significant role in folding, membrane insertion, and ATP hydrolysis (Kiss et al., 2015). ABCB6 lacking TMD0 mostly localizes to the plasma membrane and similar results were obtained for N-terminally truncated ABCB9 (Demirel et al., 2010). These results suggest that TMD0 is important for trafficking. TMD0 is also observed among TAP1, TAP2, and SUR1 (Schrodt et al., 2006). TAP1 and TAP2 structures were solved at 6.5Å (Oldham et al., 2016) *via* cryo-electron microscopy (cryo-EM) but resolvable electron density was not found for TMD0s, which suggests that TMD0 is either flexible or unstructured and functions as a docking site for adaptor proteins (Blees et al., 2015). The structure of sulfonylurea receptor 1 (SUR1) was recently determined at 5.6Å and 6Å *via* cryo-electron microscopy (Li et al., 2017, Martin et al., 2017). The SUR1-TMD0 is comprised of 5 helices that are stabilized by its interaction with an inwardly rectifying potassium channel *i.e.* Kir6.2 (Martin et al., 2017, Li et al., 2017). By knowing, that TMD0 of TAP1/2 interacts with adaptor proteins and TMD0 of SUR1 is essential for channel assembly with Kir6.2, one can propose that TMD0 functions as a docking site that is stabilized by protein-protein interactions but it is still unknown if ABCB6 docks with other proteins through TMD0-mediated interactions. To date, there is only one

high-resolution structure of full-length ABCB6 on the protein databank (PDB) which revealed the translocation mechanism (Wang et al., 2020a) (Figure 1.5).
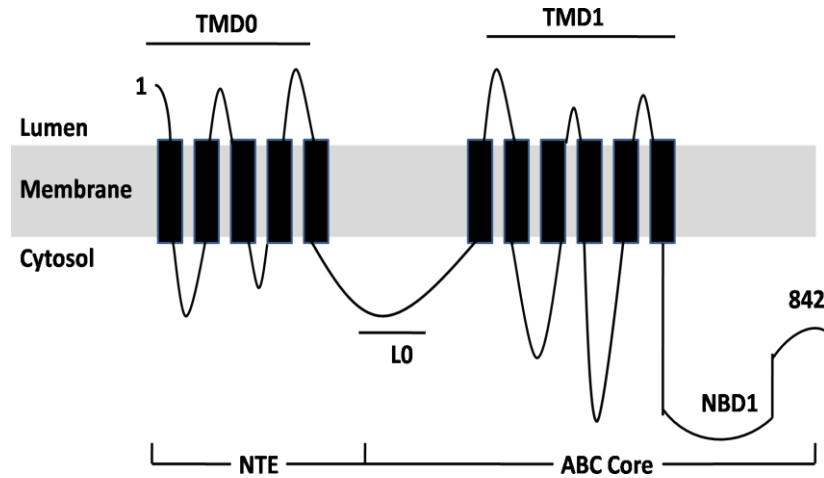
Figure 1.4: Membrane topology of human ABCB6. Transmembrane domain (TMD1) and nucleotide binding domain (NBD1) forms ABC core. TMD0 is an extra, unique and long N terminal extension (NTE). NBD1 is localized to the cytosolic side. L0 is linker which connects TMD0 and TMD1.
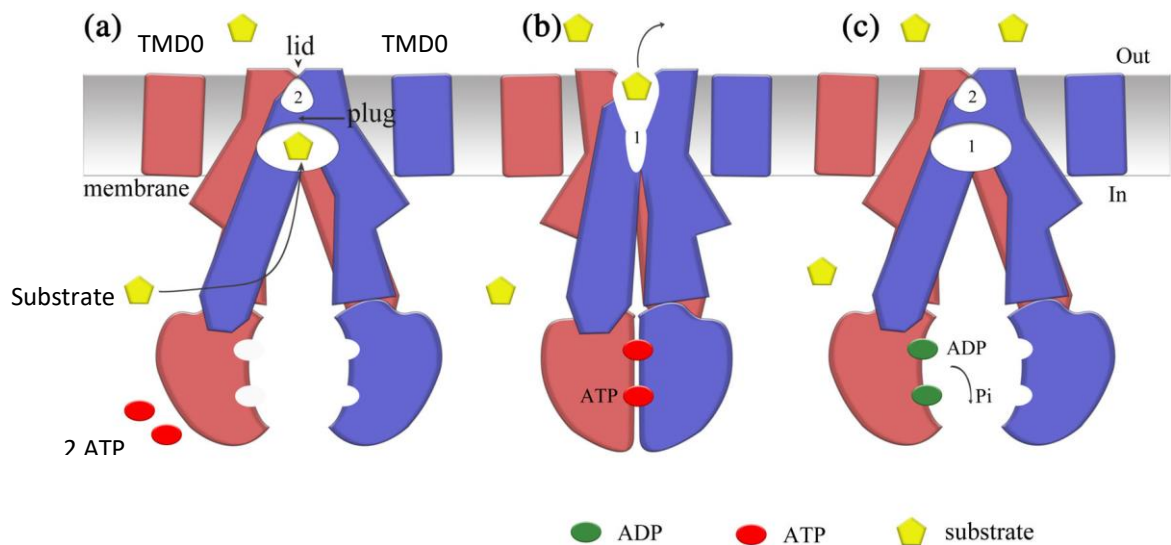
Figure 1.5: The proposed transport mechanism of ATP-binding cassette transporter subfamily B, member 6 (ABCB6). (a) Human ABCB6 protein is inward-facing in apo-state, and cavity-1 is ready to bind substrates from the cytoplasm since cavity-2 is substantially smaller than cavity-1 and is sealed by the lid, substrate binding from the opposing side is avoided. **(b)** Stimulated by substrate binding, two ATP molecules bind to the NBDs, causing the NBD dimer to close and ABCB6 to the outward-facing conformation. After that, the plug and lid are opened, allowing the substrate to move from the smaller cavity-1 to the larger cavity-2. **(c)** When ATP is hydrolyzed, ABCB6 releases substrate from cavity-2 to the outside of the cell. The NBDs release ADPs and inorganic phosphate (P_i) into the cytoplasm which results the protein to restore to the inward state and prepare for the next cycle of substrate transport (Wang et al., 2020a).

Structures of the isolated nucleotide-binding domain (NBD) of ABCB6 have been solved, with X-ray crystallography in the apo- form, ADP bound, ADP+Mg^{+2}, and in complex with ATP so far (Figure 1.6) (Haffke et al., 2010). The greatest detail in current ABCB6 structural knowledge comes from the X-ray structures of its NBDs. The NBD structure of ABCB6 is consistent with the NBD structure of other ABC transporters with sequence similarity between 24-40%. NBD of ABCB6 was compared/aligned with NBD of Sav1866, MDRP2 of P.yoelili (1XCC), P-gp of Mus musculus (5KPI), MsbA, and HylB and observed RMSD 1.6Å, 1.13Å, 1.2Å, 1.9Å and 1.04Å respectively. Like many other ATPase proteins, the ABCB6-NBD has Walker A and Walker B motifs that interact with and hydrolyze nucleotides. There is also a highly conserved signature sequence (LSGGQ) that is unique to NBDs of ABC proteins.

Consistent with the NBD structure of other ABC transporters, ABCB6-NBD also consists of two lobes *i.e.* lobe I and lobe II. Lobe I comprised of parallel *β* sheets (*β3, β6, β9, β10, β11 and β12*), *α* helices (*α2, α3, α8, α9, α10 & α11*) and anti-parallel *β* sheets (*β1, β2, β4 & β5*) (Figure 1.6). Anti-parallel *β* sheets are also known as an anti-parallel sub-domain of lobe I (Jones and George, 2004). Lobe II consists of *α* helices (*α4, α5, α6 & α7*). Another long, N-terminal helix termed as *α1* protrudes out from both lobes and functions as a connecting loop between NBD and TMD of ABCB6. Comparison with full-length murine P-glycoprotein suggested this *α1* helix is highly flexible and might represent a hinge region in full-length ABCB6. Because of the high flexibility of *α1*, their movement may prevent the nucleotide-dependent dimerization of isolated NBD by the blocking monomer-monomer interface (Haffke et al., 2010).

Upon nucleotide binding, several structural changes have been noticed as compared to the apo-form. Inward rotation of lobe II with respect to lobe I, distortion and shortening of an α1 helix, movement of α2 towards β anti-parallel sub-domain, and rearrangement of certain important residues in loop A at the N-terminal region of α2 play a significant role in nucleotide binding and hydrolysis. The observed RMSD for inward movement of lobe II of ABCB6-NBD of ATP, ADP, and ADP+Mg^{+2} are 1.49Å, 1.42Å, and 1.28Å respectively. The overall structural changes in NBD upon nucleotide binding and slight differences in ATP, ADP, and ADP+Mg^{+2} bound structures suggest that hydrolysis of ATP is linked with intramolecular movements and these structural changes are mandatory for the translocation of substrates across the membrane (Dawson et al., 2007, Higgins and Linton, 2004, Locher, 2009). Additionally, Tyr599 of ABCB6 is a highly conserved residue that plays a significant

role. It flips over and forms π-π interactions with adenine. Because of the movement of Tyr599, Thr606 and surrounding backbone atoms of Glu604 and Leu606 reoriented, as a result, the hydrogen bond between *β1* and *β2* breaks. Adenine is further stabilized by van der Waals interactions between the OH group of Tyr640 and N6 of adenine. Hydrogen bonding is also observed between Ser630 and *β*-phosphate moiety and between Thr631 and *α*-phosphate moiety. In ATP-bound structure of NBD Arg634 established a hydrogen bond with *β*-phosphate moiety (2.6Å). In contrast, this Arg634 in ADP and ADP+Mg^{+2} bound structures developed van der Waals forces with N7 of adenine (3.9Å& 3.8Å respectively) and with α-phosphate moiety (3.9Å& 4.0Å respectively).
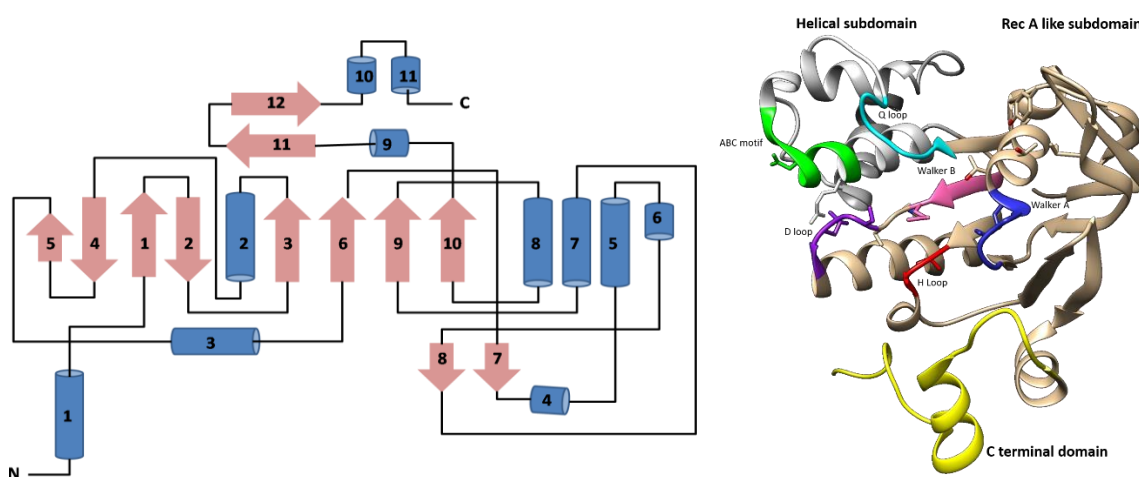
Figure 1.6: Topology of ABCB6 NBD and general NBD. α helices are represented by blue color while β sheets are represented by pink color. Parallel beta strands (β3, β6, β9, β10, β11 and β12) and anti-parallelβ strands (β5, β4, β1& β2) surround the central helix α2. Helix α1 extends out and functions as a connecting link between NBD and TMD of ABCB6.

The localization of ABCB6 remains to be determined. ABCB6 was first discovered as an energy-dependent porphyrin transporter located in the outer mitochondrial membrane (Krishnamurthy et al., 2006) and was later also found in the golgi apparatus (Krishnamurthy et al., 2006), in the endo-lysosomal compartment (Bagshaw et al., 2005, Abdul Jalil et al., 2008), in the plasma membrane (Paterson et al., 2007), in the exosomes released from maturing reticulocytes, and in red blood cells (Kiss et al., 2012). The lack of a specific targeting signal supports the existence of different intracellular localization. The underlying signaling pathway and the biochemical mechanism for cellular and organelle localization of ABCB6 is yet to be determined (Boswell-Casteel et al., 2018). Furthermore, whether the trafficking pattern of ABCB6 is same in all cells or depends on a particular cell type is still unclear. It is supposed that endocytosis takes place which causes the internalization of

ABCB6 from the plasma membrane and then distributed to endosomes and lysosomes (Kiss et al., 2015). In arabidopsis and tobacco, ABCB6 is found to be localized in the plasma membrane (Zhang et al., 2018) while in *C. elegans* it is reported to be a part of early and late endosomes (Kim et al., 2018). In addition to controversial localization data, the function of ABCB6 is also unclear. In mitochondria, ABCB6 plays an important role in the transport of porphyrins for haem biosynthesis. In the plasma membrane, ABCB6 encodes a rare blood group system Langereis (Lan) (Helias et al., 2012). Furthermore, several other functions for ABCB6 have been suggested, it provides protection against phenyl-hydrazine and oxidative stress and resistance to toxic metals (Cd, Cu & As) (Chavan et al., 2011, Lynch et al., 2009, Ulrich et al., 2012). It is found that ABCB6 [alias HMT-1 (heavy metal tolerance factor1)] is involved in heavy metals detoxification in many other species (besides humans) namely *Chlamydomonas reinhardtii, Schizosaccharomyces pombe, Rattus norvegicus, Caenorhabditis elegans,* and *Drosophila melanogaster* (Ortiz et al., 1992, Hanikenne et al., 2005, Vatamaniuk et al., 2005, Paterson et al., 2007, Sooksa-Nguan et al., 2009, Schwartz et al., 2010). However, a direct relation of these phenotypic expressions and functions of ABCB6 is not well studied. Biochemical studies have revealed that ABCB6 affects the antioxidant system by decreasing the levels of reactive oxygen species (ROS). This influence is maintained during haem synthesis. One possible mechanism might be to increase the activity and stability of catalase. Catalase is an enzyme that converts toxic hydrogen peroxide (H_2O_2) to water. It requires a co-factor haem for its function and stability. So, ABCB6 may play a role by providing a co-factor *i.e.* haem to catalase (Boswell-Casteel et al., 2018). Taken together, both the localization and the function of ABCB6 are still debated. A number of studies revealed that upregulation of ABCB6 is linked to multiple chemotherapeutic resistance. Daunorubicin in acute myeloid leukemia, camptothecin in lung cancer, paclitaxel in breast cancer, vincristine, topotecan, paclitaxel, methotrexate, cisplatin, and doxorubicin in ovarian cancer showed the development of resistance and drug inefficacy (Varatharajan et al., 2017, Yasui et al., 2004, Park et al., 2006, Januchowski et al., 2013). However, it is difficult to relate these findings to the mitochondrial porphyrin importing role of ABCB6. The current published data highlighted that the plasma membrane ABCB6 can also export porphyrins proposing that ABCB6 can export chemotherapeutic drugs too, but this has not been directly established. The acquired chemotherapeutic resistance might also be linked to ROS generation. These chemotherapeutics induce the formation of ROS (Brea-Calvo et al., 2009, Focaccetti et al., 2015, Groninger et al., 2002). It is suggested that

mitochondrial ABCB6 affects ROS by regulating porphyrin biosynthesis and limiting the ability of chemotherapeutics to induce mitochondrial ROS formation, adapting alternative mechanisms of drug resistance. However, the underlying mechanisms and potential impact remain to be investigated. A recent study of porphyria patients revealed mutations in ABCB6 as genetic modifiers and are associated with severe symptoms, highlighting the ABCB6 exporting role during porphyrin overload (Fukuda et al., 2016). ABCB6-null mice showed a hematologically normal phenotype at baseline, however, when exposed to porphyrin-induced stress suffered from an enhanced mortality rate (Ulrich et al., 2012). The missense mutations in ABCB6 have been associated with several rare genetic disorders such as ocular coloboma, dyschromatosis universalis hereditaria (DUH), and familial pseudohyperkalemia (FP). Ocular coloboma, is defined as a missing piece of tissue in structures that form the eye such as the retina, iris, optic nerve, or choroid. For example, in iris coloboma, iris is affected. DUH is a disorder of abnormal pigmentation and FP is a high level of potassium in the blood. (Wang et al., 2012, Zhang et al., 2013, Cui et al., 2013, Andolfo et al., 2013, Andolfo et al., 2016).

1.4 ATP Binding Cassette Protein sub-family G member 1 (ABCG1)

ATP binding cassette protein sub-family G member 1 (ABCG1), is a half transporter, with a reverse domain order *i.e.* N-terminal nucleotide-binding domain (NBD) and C-terminal transmembrane domain (TMD) in contrast to subfamilies ABCB-ABCD. The functional unit is a homodimer, similar to other ABC half-transporters, with two transmembrane domains (TMDs) and two NBDs, which bind and hydrolyse ATP for transporting the substrate across the membrane (Theodoulou and Kerr, 2015). Full-length human ABCG1 is composed of 678 amino acid residues and its molecular weight is 75.6 kDa. Human macrophages have been found to contain two isoforms of the ABCG1 protein, isoforms 1 and 4. Isoform 1 and 4 differ in a 12-amino acid stretch which comprises of positively charged residues that is present in isoform 1 but absent in isoform 4 (Gelissen et al., 2010). There is evidence that the longer isoform 1 has a half-life that is two times shorter than that of isoform 4 (Gelissen et al., 2010) most likely because of extra ubiquitination and phosphorylation sites in the 12 amino acid stretch. (Hsieh et al., 2014, Gelissen et al., 2012). ABCG1 shares a 72% sequence identity with ABCG4, a putative sterol transporter (Bojanic et al., 2010, Oldfield et al., 2002). When compared to the other three members ABCG2/G5/G8, ABCG1 has a sequence identity of just about 30% (Xu et al., 2022). While reports of ABCG1/G4 heterodimers *in*

vitro exist (Cserepes et al., 2004a, Hegyi and Homolya, 2016, Vaughan and Oram, 2005). ABCG1 is expressed in a variety of peripheral tissues, but ABCG4 is solely found in the brain and the retina, ABCG1 most likely functions as a homodimer (Oldfield et al., 2002). ABCG1 as a component of the reverse cholesterol transport (RCT) pathway, regulates the homeostasis of sterols in peripheral tissues, preventing the build-up of cholesterol in cells and the development of atherosclerotic lesions (Ranalletta et al., 2006, Yvan-Charvet et al., 2007). Additionally, ABCG1 is also important for the formation of high-density lipoprotein (HDL). However, the mechanism of how ABCG1 is involved in HDL formation remains unclear. Dysfunction of ABCG1 has been linked with many diseases, including diabetes, cardiovascular disease (Sturek et al., 2010), and pulmonary alveolar proteinosis, a lung condition marked by inadequate surfactant secretion (de Aguiar Vallim et al., 2017). ABCG1 is an active lipid transporter and has synergistically linked binding sites for sphingomyelin and cholesterol. Furthermore, ABCG1 clears extra cholesterol from perivascular tissues. The mechanism underlying ABCG1-mediated cholesterol transportation is unknown.

ABCG1 has the same fold as other ABCG subfamily members, with the TMDs consisting of 6 TM helices and 3 extracellular (EC) loops (Taylor et al., 2017, Lee et al., 2016) (Figure 1.7). The NBDs and TMDs are connected by a flexible linker that includes the protein segment differentiating the two ABCG1 isoforms. The N-terminal part of TM1 forms the connecting helix (CnH), while the loop between TM2 and TM3 contacts the NBDs and forms the coupling helix (CpH) (Skarda et al., 2021). TM2 and TM5a line the translocation pathway (Skarda et al., 2021, Xu et al., 2022). ABCG1 has a shorter EC3 compared with ABCG2 or ABCG5/G8, and without N-glycosylation sites. While ABCG1 monomers do not form intermolecular disulphide bonds EC3 contains two cysteines that are in sufficiently close distance to form an intramolecular bond (Skarda et al., 2021).

Like ABCG2, two cavities were also found between the TMDs of ABCG1, a larger cavity (cavity 1) the substrate binding cavity, that opens to the cytosolic side, and a smaller cavity near the extracellular membrane boundary. The two cavities are separated by a hydrophobic gate region (Skarda et al., 2021). Cavity 1 has a narrow, slit-like shape and is laterally open to the plasma membrane. This would allow a flat polycyclic molecule such as cholesterol to enter the cavity, but larger, more globular molecules would be excluded. The proposed mechanism of transport of ABCG1 is shown in Figure 1.8.
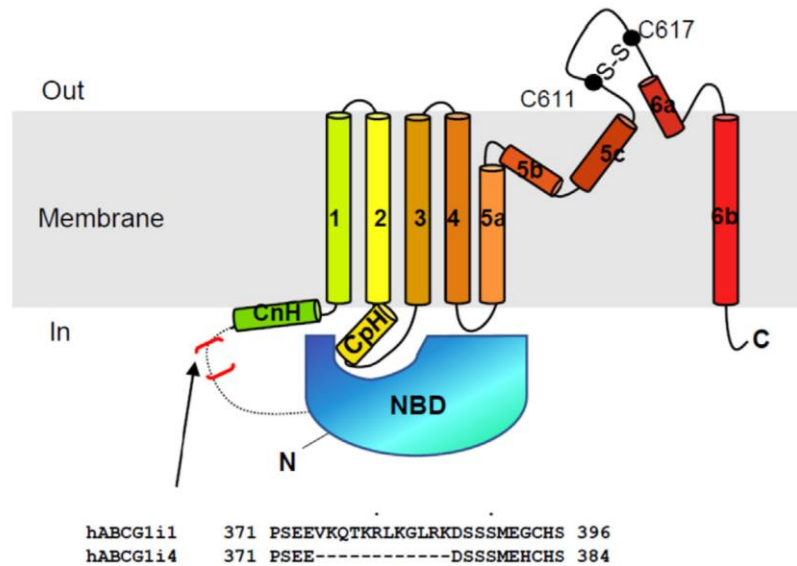
Figure 1.7: Membrane topology of human ABCG1. Showing N- and C-termini, TM helices, coupling helix (CpH), the connecting helix (CnH), and the disulfide bridge in EC3. The bracket indicates the location of a deletion in the linker region for ABCG1 isoform 4 (Skarda et al., 2021).
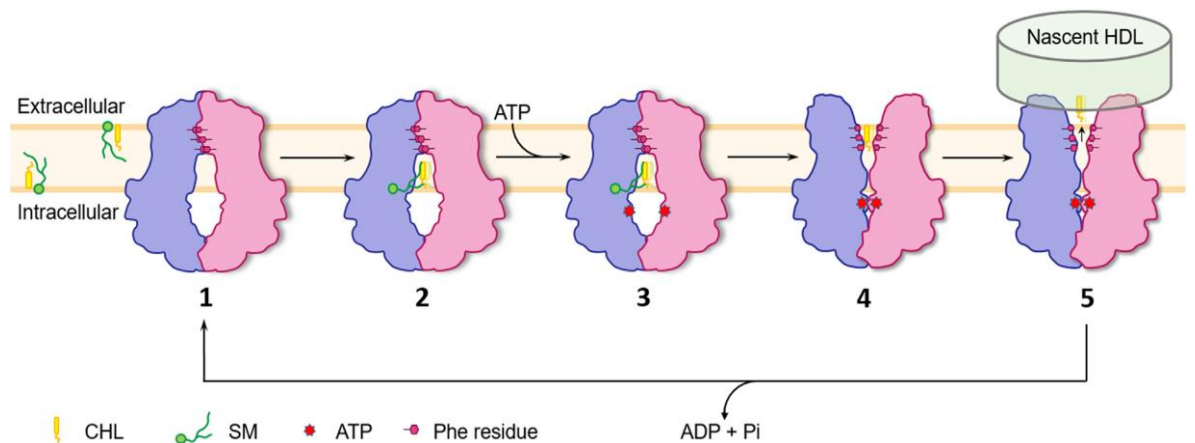
Figure 1.8: The proposed transport mechanism of ATP-binding cassette transporter subfamily G, member 1 (ABCG1). State 1: Apo form a resting state of ABCG1, **State 2:** the inward-facing conformation with simultaneous binding of cholesterol and sphingomyelin, **State 3:** the turnover state upon ATP binding **State 4:** the outward-facing conformation with cholesterol passing the Phe-rich gate, **State 5:** the cholesterol release (Xu et al., 2022).

Like ABCG2, ABCG1 has a hydrophobic gate. However, ABCG2 has leucine amino acids forming the gate, while, ABCG1 has phenylalanine residues (Phe460, Phe582, Phe583) in this region (Xu et al., 2022). In both ABCG1 and ABCG4, the regions forming the binding cavity, the membrane entrance, as well as the hydrophobic gate are identical, suggesting that ABCG4 and ABCG1 share a similar mechanism of cholesterol recognition and transport. To

differentiate between ABCG1 from ABCA1 and ABCG5/G8-mediated cholesterol transport, more research is needed on the architectures of these transporters captured in different states.

1.5 ATP Binding Cassette Protein sub-family G member 4 (ABCG4)

ATP Binding Cassette Protein sub-family G member 4 (ABCG4), is a half-transporter, similar to ABCG1, having an N-terminal NBD and a C-terminal TMD (Figure 1.9). ABCG4 is considered a brain lipid transporter (Yang et al., 2021). In humans, the ABCG4 gene is located on 11q23 (Engel et al., 2001) which contains organizations homologous with regions of mouse chromosome 9A5.3 and rat chromosome 8 (Serikawa et al., 1998). In this context, it is concluded that rat, mouse, and human ABCG4 genes exist in the chromosomal regions where a conserved linkage homology has been identified among these species.
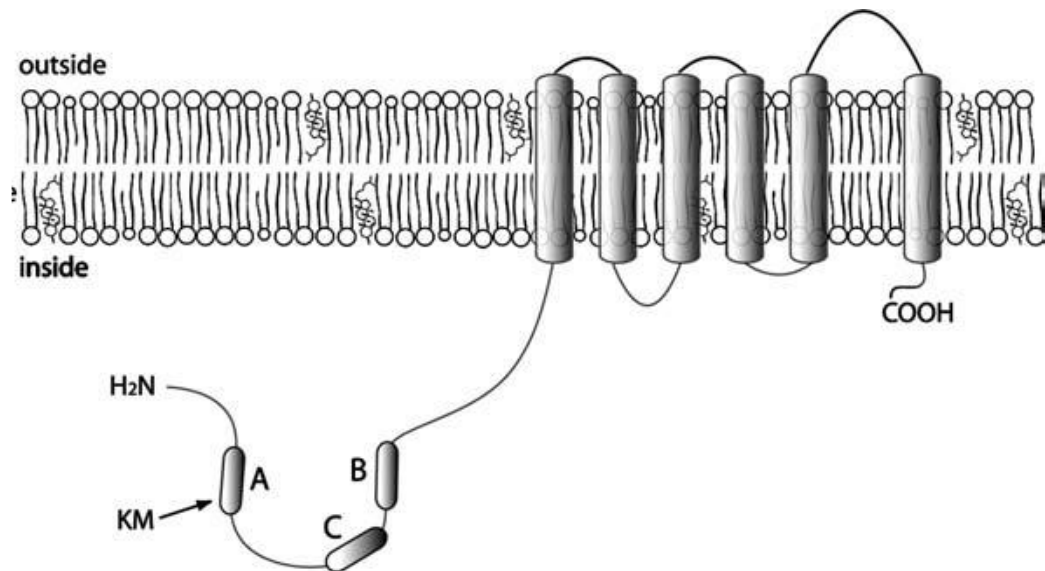
Figure 1.9: Membrane topology of ABCG4 transporter (Cserepes et al., 2004b).

ABCG4 is nearly exclusively expressed in the brain and eye, whereas ABCG1 is widely distributed. (Kerr et al., 2011). The transporter is significant in the context of Alzheimer's disease since it has recently been discovered to be expressed in the blood brain barrier endothelium and implicated in the synthesis and transportation of amyloid β-peptide from the brain (Dodacki et al., 2017, Do et al., 2012, Sano et al., 2016). This emphasises how important it is to comprehend the transcriptional and posttranslational mechanisms controlling this transporter. Understanding the regulation of this transporter is crucial for developing new strategies to upregulate its function in brain cells. This transporter has previously been demonstrated to be significant in sterol transport in the brain, including the

transport of cholesterol and oxysterols. The function, mechanism of action, and localization of ABCG4 are all as yet poorly understood which emphasize to explore ABCG4 in more depth.

1.6 Specialized Hosts and Systems for membrane protein overexpression

To get over-expression of a membrane protein is the first major hindrance in functional and structural studies (Tate, 2001). When the MPs show expression in heterologous systems, the proteins may prove toxic for the host and are expressed at a low level in a spatially delineated membranous environment. They might remain inactive as they are not being folded into functional conformation (Miroux and Walker, 1996). To achieve over-expression of a protein involves three components: a gene of interest, a vector as a vehicle/carrier, and an expression host. The suitable combination of these three elements enhances the quality and quantity of protein produced (Bernaudat et al., 2011).

Recombinant membrane proteins can be expressed *in vivo* using prokaryotic cells (e.g. *E. coli* and *Lactococcus lactis*) and eukaryotic cells (mammalian, yeast, insect cells). In addition, MPs can also be produced *in vitro* using cell-free expression systems (e.g., Easy Xpress® Kits) (Kubicek et al., 2014). However, each system has its own merits and demerits, and the selection of the suitable expression system often remains empirical. Scientists can choose from a range of expression systems to meet their needs. The most widely known expression host is a bacterial expression system, particularly E. coli, which is relatively simple, needs fewer growth resources, cost-effective, availability of an increasingly large number of mutant host strains and cloning vectors, well-characterized genetics, and produces a good yield (Terpe, 2006, Basan et al., 2015, Francis and Page, 2010). On the other hand, there are drawbacks as well. Contrary to eukaryotic systems, transcription and translation in bacterial system are closely connected and occur quickly. Since many eukaryotic proteins need more time and the assistance of folding chaperones to fold into their native state, the increased rate of transcription and translation frequently results in a pool of partially folded, unfolded, or misfolded, insoluble proteins. Furthermore, it is difficult to produce disulfide-containing proteins effectively due to the reducing environment of the bacterial cytoplasm (Oberg et al., 1994). Lastly, bacterial cells lacks the tools necessary to carry out several eukaryotic post-translational changes, like glycosylation, which is vital for the production of folded, active protein (Schmidt, 2004).

Eukaryotic expression systems, like insect (usually Sf9, Sf21) and mammalian cells, are used as an alternative to get around these issues. They can give membrane proteins a setting that more closely resembles their natural environment. However, cell culture of these systems demands advanced tools and abilities in order to monitor and maintain the system efficiently, and slow growth rates and medium costs are possibly troublesome for some specific works (Andréll and Tate, 2013). Therefore, it is believed that protein expression in yeast can address these issues.

Yeast is a unicellular eukaryotic organism with a high growth rate, less nutritional requirements, and facile genetic manipulation, as well as the majority of traits of a higher eukaryotic system, such as post-translational modification and secretory expression (Vieira Gomes et al., 2018). Several yeast genera, including Saccharomyces, Pichia, Yarrowia, Kluyveromyces, Schizosaccharomyces, Zygosaccharomyces, and Hansenula, are frequently used for the expression of recombinant proteins. For the expression procedure to be successful, choosing a suitable yeast strain is crucial. There are two types of yeast: methylotrophic yeasts and non-methylotrophic yeasts (Fernández et al., 2016). It is known that methylotrophic yeasts, notably *Pichia pastoris*, uses methanol as a carbon source and grows to very high density in a simple fermentation process resulting in high cell masses. Due to their strong and tightly controlled promoters, which are a result of their high requirement for methanol oxidising enzymes, methylotrophic yeasts have the benefit of being able to express foreign genes of interest stably and effectively (Minjie and Zhongping, 2013). However, non-methylotrophic yeasts use carbon compounds other than methanol as a carbon source, such as glucose. Strong benefits of this kind of yeast include the depth of understanding of their genetics, metabolism, and stress response, and the familiarity to fermentation process particularly for *S. cerevisiae* that's why non-methylotrophic yeasts have long been used successfully for the expression of recombinant proteins (Britton et al., 2011). In *S. cerevisiae*, human P-gp has been successfully expressed (Mao and Scarborough, 1997). It was demonstrated that the protein was correctly folded and functional. The protein's electrophoretic mobility was said to be the same as P-gp derived from a Sf9 insect expression system, according to the authors (Mao and Scarborough, 1997). Thus, it has been demonstrated that using a yeast expression system to produce P-gp has the potential to be advantageous for subsequent studies in both functional and structural aspects.

Additionally, it has been demonstrated that the ability to produce ABCG2, CFTR, and ABCB5 utilising a yeast expression system is advantageous for later functional and structural studies (Huang et al., 1996, Thonghin et al., 2018a, Keniya et al., 2014). For this

research work, based on benefits of using yeast expression system for membrane protein production, *Saccharomyces cerevisae* was chosen for the expression and purification of selected ATP binding cassette proteins. Comparison of different expression system is given in Table 1.2.

Table 1.2: Comparison of Recombinant Protein Expression system

	E.coli	Yeast	Insect cells	Mammalian cells	Cell-free
Average time for cell division	30 min	90 min	18 hours	24 hours	N/A
Expression cost	low	Low	High	high	high
Expression level	high	Low-high	Low-high	Low-moderate	Low-high
Advantages	Simple, low cost, robust, high yield	Simple, low cost, post-translation modifications	Post-translational modifications	Post-translational modifications, Natural protein configuration	High yield, fast, disulphide bonded and membrane proteins
Disadvantages	No post-translational modification, production of disulphide bonded and membrane proteins is difficult	Glycosylation in yeast cells is different from that in human cells, Differences in lipids composition of membrane	Slow, high cost	Slow, high cost, lower yield	High cost, efficient production requires highly specialized set up

Producing recombinant proteins always involves altering the metabolism of the host cell, which often slows biomass growth rates. This results from the high maintenance needs for the replication of plasmids that have been introduced as well as the resources required for the transcription of the target genes. Especially, plasmids with high copy numbers can overwhelm the cell and cause a metabolic collapse, which leads in reduced product yields or

even cell death. Thus, a good balance of biomass growth and product formation associated with unavoidable cell stress has to be found to achieve high productivity. Thus, it is necessary to optimise a lot of parameters. This includes vector with appropriate promoter system, medium's composition, oxygen availability, growth temperature, and induction conditions. It is important to take the characteristics of the target protein and the intended downstream application into account when choosing a promoter system. Promoter systems with extremely low basal expression levels is useful, if the target protein is toxic for host cell (such as a ribonuclease). Alternatively, for maximal protein yield, a strong promoter should be selected. Promoters are perhaps the best studied and engineered genetic component in many yeast systems due to their vital involvement in the construction of expression cassettes. Well-characterized constitutive or inducible promoters with strong transcriptional activity are used to achieve overproduction of recombinant protein. Constitutive promoters are straightforward and maintain relatively constant levels of expression, whereas inducible promoters allow to control gene expression levels in the presence of an inducer. Galactose-induced GAL1 and GAL10 promoters are frequently utilised in S. cerevisiae; GAL1 promoter is employed in this study.

Using a weak promoter is generally not a good idea because it leads to low levels of the target gene's transcription and, as a result, low levels of the recombinant protein. Similarly, selecting a strong promoter is not always advised because a gene of interest's high transcript levels may stress the cell if its protein product activates the Unfolded Protein Response (UPR), which results in cell death. Hence, understanding protein toxicity for the host, requirements for protein folding, and protein size are critical for selecting the right promoter.

1.7 Purification of membrane proteins

When membrane proteins are expressed at an appropriate level, it is essential to separate them from the lipid bilayer for purification, before structural analysis and this includes typically three steps:

- Solubilisation
- Isolation/ Purification
- Reconstitution

Soluble protein purification is fairly simple and does not require sophisticated techniques. But membrane protein purification is more difficult than soluble protein purification (Smith,

2011). The purification process may be affected by a number of variables, necessitating trial and error to optimise the methodology.

1.7.1 Solubilization

Solubilization is a method in which integral membrane proteins (IMPs) are extracted from the membrane. This can be done with the assistance of amphipathic detergents (Seddon et al., 2004). Detergents are vital during solubilisation, purification, and crystallization processes. Detergents have the ability (due to their amphipathic nature) to extract the IMP, by disturbing the lipid bilayer (Linke, 2009). Detergents imitate the phospholipid bilayer by surrounding the hydrophobic region of integral membrane protein, producing the protein-detergent complex (PDC), which is water-soluble. This avoids the aggregation of protein after its separation from the natural lipid setting (Moraes et al., 2014). Detergents are broadly classified into three categories based on their charge, ionic, non-ionic, and Zwitterionic detergents. The chemical nature of ionic detergents makes them less fit for membrane proteins (MPs) purification as they can interfere with protein-protein interactions causing unwanted denaturation of protein folds (Privé, 2007). Zwitterionic detergents have both negative and positive charges on the same head group and thus have overall zero net charge. Because of this net zero charge, they are generally less harsh than ionic detergents. Nonionic detergents are mostly preferred and widely used for solubilisation. They consist of uncharged hydrophilic head groups and are considered mild because they disrupt the protein-lipid and lipid–lipid bonding instead of disturbing the protein-protein interactions. A general rule is detergents with shorter lengths are suitable for solubilisation and crystallographic studies, whereas detergents having long chains are appropriate for reconstitution methods (Arachea et al., 2012, Moraes et al., 2014).

The efficiency of a specific detergent to solubilize and extract the membrane protein is dependent on its ability to make a micelle. In an aqueous solution, the polar region of the detergent forms hydrogen bonds (H-bonds) with molecules of water, while aggregation of the hydrophobic tail occurs due to hydrophobic interactions and gives rise to a structure termed a micelle. In general, micelles have a molecular mass of less than 100 kDa and are a few nanometers in diameter (Seddon et al., 2004). Critical micelle concentration (CMC) is the minimum amount of detergent at which micelles are produced during the process. A micelle (ordered cluster or aggregate) is made up of several detergent monomers. Micelles have an approximately spherical form, with a hydrocarbon chain inside and amphiphilic

monomer polar/ionic moieties on the outside. Many intriguing features of the micelle are due to its hydrocarbon interior. A micelle is made up of a precise number of surfactant molecules, known as the aggregation number N (Thévenot et al., 2005). The type of surfactants, temperature, and electrolytes in an aqueous solution all affect the value of N (Hoffman, 1978). The quantity of detergent to be used is of prime importance, as lack of detergent can cause aggregation of protein and in excess; detergents can cause protein inactivation because of the elimination of lipids important for function. Too much detergent can also hinder the crystallization method leading to phase separation (Guan et al., 2006).

In membrane protein research, the quality of detergent with high purity is always recommended and cannot be disregarded. Although nowadays most detergents are commercially available and highly pure but researchers should be aware of the purity level. In addition, they should have knowledge of different analytical methods through which the purity has been checked, *e.g.* TLC or HPLC. Impurities, if present in detergents can interact with proteins during the process of extraction/purification can hamper the crystallization procedure. The commonly observed impurities in detergents are peroxides, (α) isomers, and hydrophobic alcohols the greater part of which is water insoluble (Moraes et al., 2014). So, if cloudiness appears in the detergent solution at a concentration at which it is presumed to be soluble, it may be an indication of contaminants. It is recommended to prepare detergent solutions before use (fresh), filter, and protect from direct light (Ashani and Catravas, 1980).

1.7.2 Isolation/Purification of Protein

Isolation of desired protein from cellular proteome is also a crucial step in protein purification. The purification procedure splits the protein and non-protein parts of the mixture and separates the protein of interest from the remaining proteins. Isolation of one specific protein from a remnant is usually the most difficult part of the purification method. The pure result is called protein isolate. Generally, separation steps exploit differences in physico-chemical properties such as size, biological activity, and binding capabilities. Chromatography is a commonly used method for protein purification (Structural Genomics, 2008). The basic procedure is proteins are loaded on the column are separated based on their characteristics. Based on the principle applied to purify the protein, chromatography can be size exclusion (separation based on size differences among proteins), ion exchange

chromatography (based on charge), affinity chromatography, and high-performance liquid chromatography.

Affinity chromatography is a commonly used method for membrane protein purification. In affinity chromatography, the proteins are labelled with affinity tags, bind and elute through the columns. Columns are specifically designed to attach the tag residues. Tagged proteins bind to the matrix while the rest of the untagged proteins are washed out. Then by using a particular elution buffer the desired tagged proteins are eluted out (Labrou, 2014). Proteins are tagged with these residues during the cloning step. These days, commercial vectors are loaded with affinity tags. The desired proteins are cloned adjacent to these tags and expressed in suitable expression systems.

Today, a variety of protein and peptide affinity tags are available to isolate proteins expressed in a heterologous host like E. coli or yeast. In general, an ideal affinity tag has following characteristics:

- can efficiently purify tagged proteins in high yield
- can be placed at any position (N-terminal, middle, C-terminal) in the protein
- can be use with any protein without adversely affecting its function
- can be used to purify protein expressed in any expression system or any host strain
- binds to and elutes from a resin that is inexpensive, can be regenerated, and possesses good flow characteristics.

Luckily, a number of commercially available affinity tags meet most of these requirements, if not all. The selection of an affinity tag clearly depends on the objectives of the specific experiment. The most used affinity tag for protein enrichment is the hexahistidine tag (6 His-tag). The chelated metal ions act as affinity ligands and make it simple to purify His-tagged proteins. The benefits of the 6 His- tag are its reduced size, lack of electric charge, low levels of toxicity, and immunogenicity. The major drawback of using His-tag is non-specific binding. His-tagging produces proteins with good yields from low-cost, high-capacity resins and moderate purity from E. coli extracts, but rather poor purification from mammalian cell extracts (Lichty et al., 2005). Compared to the polyhistidine tag, the maltose-binding protein (MBP) may give a good yield of recombinant protein. However, MBP's drawback is the large size and immunogenicity of the affinity tag, which makes any subsequent application more challenging. The FLAG tag is hydrophilic, small, detectable by antibodies, and unlikely to disrupt protein folding or function. FLAG tag disadvantages include expensive cost, sensitivity to high salt concentration, and a very low binding capacity of resin. Another

well-known affinity tag is the glutathione S-transferase (GST) tag, which is based on GST's high affinity for immobilised glutathione. Because GSTs are a family of multifunctional cytosolic proteins present in eukaryotic organisms but typically absent in bacteria, they are most appropriate for use in prokaryotic expression. The GST tag has long been used to make fusion proteins more soluble in E. coli, just like the MBP tag does (Wang et al., 1999). The Strep-tag is an octapeptide that binds to streptavidin (Schmidt et al., 1996). Protein folding and bioactivity are not hampered by the strept II tag. With high purification and pure yields at a good price, Strep II tag might offer a suitable compromise (Zhao et al., 2013).

In nutshell, experiments needing large amount of partially purified material in high yields at a low cost may find the GST and HIS suitable, whereas FLAG tag may be preferable for those experiments demanding small quantities of the highest purity to outweigh their costs and limited capacity. The Strep II tag is a short tag that generates high purity material in reasonable yields at a reasonable cost, making it a good choice for affinity purification. The Strep II tag, but not the HIS or FLAG tags, was demonstrated to interfere with the crystallisation of one specific enzyme, hence this is one caution when utilising the Strep II tag to purify proteins for protein crystallisation. It is unclear if the Strep II tag has a negative impact on protein crystallisation in general because other proteins have been effectively crystallised with this tag (Lichty et al., 2005).

ABC transporter purification was shown by O'Ryan et al in two steps (O'Ryan et al., 2012b). In the initial step, the protein was purified using Nickel-Nitriloacetic acid (Ni-NTA) affinity chromatography. Next, contaminants and fragmented CFTR were removed using size exclusion chromatography (SEC) to produce full-length pure protein. More than 90% purification quality was demonstrated using the approach. Mouse and human P-glycoprotein was also purified using the same two-step procedure (Thonghin et al., 2018b). Because of this, we find that for the protein purification of this study, a combination of HIS and GFP tag allows quick capture of tagged protein by nickel resin, followed by polishing/enrichment over the size exclusion column.

1.7.3 Reconstitution of membrane proteins

The process of restoring an isolated membrane protein into its original form is termed as reconstitution. *In situ* studies of a membrane protein are difficult in their complex environment. Therefore, they are separated and purified but the actual structure and function cannot be evaluated without a lipid environment (Liao et al., 2006). Therefore, an isolated membrane protein is often incorporated into stable artificial lipid membranes such as

nanodisc, liposomes, *etc*. Different membrane mimectic systems are shown in figure 1.10. Liposomes are artificial vesicles and to prepare them several methods have been used *i.e* mechanical, organic solvents, freeze-thawing, and detergents mediated (Szoka Jr and Papahadjopoulos, 1980). However, the detergent-mediated method is widely used (Rigaud et al., 1995, Seddon et al., 2004). In this method, dried lipids are hydrated in an aqueous buffer followed by agitation which induces vesicle formation. Added detergent is incorporated into the liposome membrane. To reconstitute the isolated protein, the purified protein is mixed with liposome, afterward, the detergent is removed by chemical adsorption or dialysis. The liposome and protein rapidly assemble to form proteoliposomes (Rigaud and Lévy, 2003). However, proteoliposomes are unstable and it's tricky to synthesize them in a defined controlled size and stoichiometry (Bayburt and Sligar, 2010).
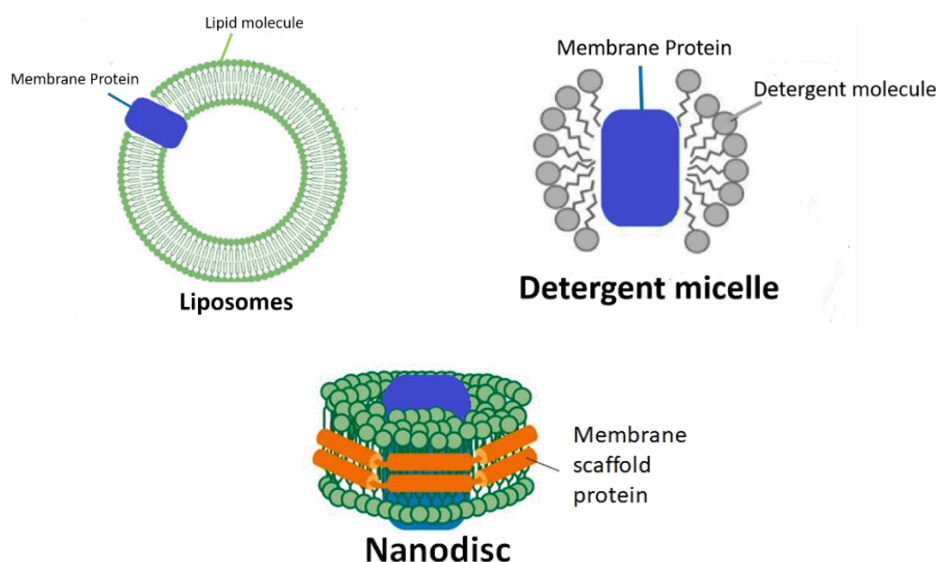
Figure 1.10: Different membrane mimectic system. Membrane protein is symbolised as purple cartoon block.

The de-stability associated with the detergent-mediated method can be overcome through "Nanodisc technology". Nanodisc is a soluble phospholipid bilayer at a nanoscale level that can co-assemble with integral membrane proteins (IMPs) for structural and biophysical studies. Thus, solubilizing membrane proteins (MPs) at the single molecule level offers advantages over proteoliposomes in terms of stability, size, and quick access to both sides of the phospholipid bilayer. It provides a distinct platform to understand and comprehend the function, structure, and pathways of membrane protein (Denisov et al., 2004). In the nanodisc method, solubilisation of desired membrane protein is carried out in the presence of phospholipids and membrane scaffold protein (MSP) (Bayburt et al., 2002). During the detergent removal process, the target membrane protein simultaneously assembles with

phospholipids into a discoidal bilayer, the size of which is determined by MSP length. The nanodiscs produced consequently keep membrane proteins in solution, giving a native-like phospholipid bilayer environment that provides stability to the target protein. Thus, nanodisc provides a cassette, rendering membrane proteins soluble at the single molecule levels that were initially limited to soluble proteins and enzymes only. This technique is useful in obtaining deep insights into structural and functional investigations (Ritchie et al., 2009).

1.8 Structural Analysis

1.8.1 Small-angle X-ray Scattering (SAXS)

The practice of X-ray scattering in solution is termed small angle X-ray scattering. SAXS is an efficient technique for the analysis of the 3D structure of biomolecules in solution. The sample solution should be at μm-mM concentration and monodisperse. The basic principle was developed by Guinier in the 1930s with his studies on metallic alloys (Guinier and Fournet, 1955). Guinier and Fournet described in the first monograph on SAXS that this method besides providing information on size and shape also gives an insight into the internal structure of biomolecules (Guinier and Fournet, 1955). Because of the high speed of data collection and processing, this method is preferred despite having low to medium resolution. Typically, a SAXS experiment is straightforward: X-rays illuminate the sample after passing through a collimating system. An X-ray source can be a laboratory source (by rotating anode) or synchrotron light which provides a higher X-ray flux. Electron density variations in the sample scatter some X-rays away from the primary beam. Scattered radiations fall on a detector and generate a radial pattern. Un-scattered radiations are stopped by a beam stopper/ blocker to prevent damage to the detector and to reduce the noise detection on screen. Scattering of X-rays occurs in an angular isotropic manner and is measured relative to the momentum of transfer given as:

$$q = \frac{2(\pi.sin2\theta)}{\lambda}$$

2θ is the angle between the incident and scattered radiation and λ is the wavelength of incident radiation (Svergun et al., 2013). The scattering profile can then be analysed using various formulae to obtain structural information (Heftberger et al., 2014).

1.8.2 Electron Microscopy

Electron microscope uses the beam of electrons to generate the image of the specimen. There

are two main differences between the use of X-rays and electrons when probing the structure of biomolecules: (i) The scattering cross section is almost 10^5 times greater for electrons as compared to X-rays, so significant scattering is achieved by utilizing electrons for crystals or other objects that are 1-10 nm in thickness, whereas scattering of a similar fraction of X rays needs 100-500 nm thick crystals. (ii) Since electrons are charged particles that can be deflected by a magnetic or electric field which is why it is much easier to focus a beam of electrons than X-rays. Electron beams produced in the high voltage field of an electron microscope have a 0.015-0.04 Ӓ wavelength that is less than that of X-rays. Therefore, theoretically, electrons can achieve resolution far higher than X-rays (Chiu et al., 1999). Transmission electron microscopy (TEM) is a frequently used technique for the structure determination of macromolecules. During this technique, a specimen is mounted on the grid. The grid is then illuminated with the electron beam generated from the electron gun. Electrons interact with the atoms of the specimen and are brought into focus on a detector to diffract which creates an image. Based on the methods of grid preparation TEM can be classified as negative stain electron microscopy (NS-EM) (Brenner and Horne, 1959) and cryo-electron microscopy (cryo EM) (Adrian *et al.*, 1984). In NS-EM the sample is coated with heavy metal. Heavy atoms of the stain diffract the electron readily to create a better contrasting image than the non-stained material (Rubinstein, 2007). However, NS-EM can produce image artifacts, structural deformities, and dehydration. Moreover, the use of stain also usually restricts the resolution to 20Ӓ (Wang and Sigworth, 2006).

Cryo-electron microscopy is another form of TEM in which the sample is protected at cryogenic temperature. Cooling of the specimen decreases the effect of radiation damage. These days, preparatory methods of the desired sample (macromolecules or macromolecular complexes) for cryo EM are well described (Murata and Wolf, 2018). Most such methods involve cooling of samples at a rate fast enough to permit vitrification *i.e.* a solid glass-like state. Several techniques and freezing devices have been developed for producing uniformly thin, vitrified samples. All subsequent steps, up to and including the recording of images in the microscope are carried out in a manner that maintains the sample below -170 °C to avoid devitrification, which occurs at -140 °C. This frozen, hydrated native conformation sample can then be analysed through a temperature-regulated cryo electron microscope (Saibil, 2000). The two-dimensional (2D) images are then analysed using the single particle analysis method (Rubinstein, 2007). In this method, the 2D micrograph containing several images of proteins are classified according to their orientation to the plane. Each class is then

averaged to reduce errors in composite images. Three-dimensional images are constructed using the planar angular information of each class. Recently several high-resolution structures of membrane proteins have been reported using the cryo-EM technique (Jackson et al., 2018a, Taylor et al., 2017, Skarda et al., 2021, Tall and Wang, 2022, Chen et al., 2022).

1.8.3 X-ray Crystallography

In 1901, Rontgen discovered X-rays which led to a new era of crystallography. In the early 1950s, Rosalind Franklin recorded the X-ray fibre diffraction pattern of DNA. In 1952, two scientists James Watson and Francis Crick used that data to determine the double helical structure of DNA (DeMaria, 2003). Later, Kendrew *et al* used crystallography to reveal the structure of myoglobin of sperm whale (Kendrew et al., 1958). In 1985, X-rays were first time used for structure determination of membrane proteins. For many years, this was the primary method for obtaining high-resolution crystal structures of membrane proteins; however, after advancements in transmission electron microscopy, X-ray crystallography has become less popular (Smyth and Martin, 2000, Thonghin et al., 2018c, Goldie et al., 2014).

The purpose of X-ray crystallography is to get a three-dimensional (3D) structure of protein/biomolecules from a crystal. During crystallography, purified proteins are crystallized in a 3D lattice. Before moving to the next step, it should be checked whether crystals contain the target protein or not. After confirmation, X-rays are bombarded on the crystal which is placed on the gonimeter. A specific diffraction pattern is measured. These diffraction patterns create spots of different intensities that are recorded on a detector, which gives information about structure factors for the calculation of an electron density map. The wave that reaches the detector has a particular phase and amplitude that results from the addition of individual scattering factors from all atoms in a unit cell, which each have their own phase and amplitude. Unfortunately, each reflection measured during the diffraction experiment tells us the amplitude but not the phase. There are different ways to solve phase problem *i.e.* direct method, multiple isomorphous replacement (MIR), multiwavelength anomalous dispersion (MAD), and molecular replacement. The direct method is used to solve phase problems of small and medium-sized proteins. Missing phase information is obtained from mathematical relationships between structure factors. MIR and MAD are used for larger proteins while the molecular replacement method is used for a molecule that is similar to another whose structure is known. In MIR, by introducing a known change *e.g.*

heavy atom to unit cell diffraction pattern is recorded. On the other hand, in MAD, X-ray beamline is used at synchrotrons to record altered diffraction intensities at a number of X-ray wavelengths. After computing phases, an electron density map can be calculated. By refinement, the quality of the electron density map can be enhanced. The refined map is used to identify the structure of the protein. The discovered protein structure is built into the map as represented by a PDB (protein data bank) file (Smyth and Martin, 2000).

Getting crystals of suitable quality is a critical rate-limiting factor for membrane proteins and is least well-understood. Because of this, only a few membrane protein structures have been determined (Lacapere et al., 2007). Generally, it is difficult to crystallize a glycosylated protein and proteins having flexible domains (Rossman et al., 1985). It is considered that high purity of proteins increases the chance of crystal growth. Crystals are brittle and vulnerable to radiation damage during data collection because of the solvent molecules that are present in them. Crystal freezing provides a solution to this radiation damage. In this method, protein crystals are dipped in a "cryo-protectant" solution. Vitrified water instead of ice is formed when the crystal is cooled. Cryo-crystallography offers another edge by providing complete data set from a single crystal and results in high-quality resolution diffraction data, thus gives more precise structural information (Ilari and Savino, 2008). A comparison of these three techniques is given in table 1.3.

Table 1.3: Comparison of SAXS, cryo-EM, and X-ray crystallography

Technique	Sample	Advantages	Limitations
SAXS	Solution	Structure analysis at nearly native conditions Few kDa-100 MDa (widerange)	Low resolution
Cryo EM	Vitrified solution	Minimum sample required Direct visualization of particle shape and symmetry	Difficult for MM less than 200 kDa.
X-ray Crystallography	Crystal	High resolution	Crystal formation Flexible protein parts are difficult to resolve. Crystal packing forces may also affect the structure.

1.10 Aims and Objectives of this project

Membrane proteins are recognised as potential future therapeutic targets because they regulate a variety of biological activities. One class of membrane protein is ATP binding cassette transporters (ABC family) which is superfamily of integral membrane protein, however, our understanding of how the ABC transporter family functions is still limited, especially from a structural biology point of view. Thus this study was aiming to obtain high-resolution structures of under-represented ABC transporters of medical interest. Bioinformatics was employed to choose the best orthologs of human ABC transporters as crystallisation target. But by the end of PhD stage, 3D crystallisation was taking a lesser role for selected protein targets than cryo-EM, hence cryo-EM was employed for these aims. Biochemical and biophysical assays were also carried out to add additional information to support structural studies. If the goals are met, ABC protein models could be proposed. Ultimately, future research based upon this finding, for example pharmacotherapy research and structure-based drug design could help in disease treatments.

The explicit objectives of this research project were therefore as follows:

- Screening of orthologs of ABC family members based on the protein sequences to find target proteins with a good chance of obtaining high-resolution structure information.

- Based on bioinformatics assessment: subsequent design synthesis, and cloning of ABC genes which have a high probability of solubilisation, purification, and crystallisation.

- Expression of ABC genes in the *Saccharomyces cerevisiae*, a yeast expression system

- Protein purification from the *Saccharomyces cerevisiae*

- Low-resolution structural studies for basic characterization *i.e.* SAXS & negative stain EM on target proteins

- Biochemical and biophysical assays to validate structural studies

- Crystal trials on target proteins or Cryo-electron microscopy studies of purified target proteins

Chapter 02
Materials and Methods

In this chapter, the materials, methods, and procedures used in this study are all described in detail. Glassware, equipment, and different chemicals were purchased from various manufacturers and suppliers, details of which are provided below.

2.1 Chemicals

Chemicals	Suppliers
Acetic acid	Fischer Scientific
Agar	Formedium
Amino ethyl benzene sulfonyl fluoride (AEBSF)	Sigma-Aldrich
Ammonium persulphate (APS)	Fischer Scientific
Bacto-yeast extract	Formedium
Bacto-peptone	Formedium
Bestatin	Sigma-Aldrich
40% Bis/acrylamide solution	SERVA
Bradford reagent	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Chymostatin	Sigma-Aldrich
Cholestrol hemisuccinate (CHS)	Sigma-Aldrich
Coomassie blue G250	Fischer Scientific
D-Glucose	Fischer scientific
D-Galactose	Fischer scientific
Dithiotheitol	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Fischer Scientific
Epoxysuccinyl-leucylamido-butane	Sigma-Aldrich

Glycerol	Fischer scientific, 99.5 %
Glycine	Sigma-Aldrich
Imidazole	Fischer Scientific
Lithium acetate (LiAc)	Fischer scientific
Magnesium chloride	Sigma-Aldrich
Methanol	Fischer Scientific
n-Dodecyl-β-D-maltoside (DDM)	Anatrace
PageRuler Plus pre-stained Protein ladder	Fermentas
PEG-3500	Sigma-Aldrich
Pepstatin A	Sigma-Aldrich
Phenyl methane sulfonyl fluoride (PMSF)	Sigma-Aldrich
Sodium dodecyle-sulphate	Fischer Scientific
Sodium chloride (NaCl)	Fischer Scientific
Sodium orthovanadate	Sigma-Aldrich
Sucrose	Fischer Scientific
Tetra methyl ethylene diamine (TEMED)	Bio-Rad
Tris-base	Fischer Scientific
TNP-ATP	Tocris Bioscience
Tween-20	Sigma
Uracil dropout supplement	Formedium
Glass beads, 425-600 μm, unwashed	Sigma-Aldrich
Uranyl acetate	Agar Scientific
Yeast Nitrogen Base (YNB)	Sigma-Aldrich

2.2 Laboratory instruments and Consumables

Consumables/ Instruments	Company
AKTA FPLC	GE Healthcare Life Sciences
Avanti J26S XP	Beckman Coulter
Bead beater	Biospec
Bioreactor	Applikon
Carbon coated 300 nickel mesh grid	Agar Scientific
Fluorometer	Cary Eclipse
Glacios cryo-EM	ThermoFisher Scientific
His Trap column	Generon
JLA8.1 8000 rpm rotar	Beckman
K100X glow discharge system	Quorum Technologies
Optima XE90 ultracentrifuge	Beckman Coulter
Optima Max-XP ultracentrifuge	Beckman Coulter
Quantifoil 1.2/1.3 Cu grids	Agar Scientific
SepFast ^{TM} SEC 6-6000KDa	Generon
Ti-45 45000 rpm rotar	Beckman
TLA120.1 rotar	Beckman
Uncle spectroscopic system	Unchained Lab
Viva spin concentrators	Millipore
VitroBot Mark IV	ThermoFisher Scientific
96 well plate reader (flurometer)	Biotek

2.3 Media

In our experimental work, all media were prepared in distilled water and sterilized by heating in an autoclave oven at 121 °C for 15 minutes.

Media	Chemical composition
SOC	2% tryptone, 0.5% yeast extract, 2.5mM KCl, 10mM NaCl, 10mM MgCl2, 10mM MgSO4, 20mM Glucose
LB (Luria broth)	1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl
LB Agar	1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl, 2% Agar
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
YPD Broth	Yeast peptone dextrose: 10g/L yeast extract, 20g/L peptone, 20g/L dextrose or glucose
YPD Agar	YPD broth + 15g/L agar

2.4 Buffers

Microsomes preparation

B6- Lysis buffer	300mM Tris-HCl (pH 7.4), 0.56M sorbitol, 1mM EDTA
B6- Solubilization buffer (without detergent)	50mM Tris-HCl (pH 7.4), 10% glycerol, 1mM DTT
B6- Solubilization buffer (with detergent)	50mM Tris-HCl (pH 7.4), 20% glycerol, 1mM DTT, 2% DDM
B5-Lysis buffer	0.25M Tris-HCl (pH 7.8), 0.25M sucrose, 1mM EDTA, 2mM DTT
B5- High Salt buffer	50mM Tris-HCl (pH 7.8), 500mM NaCl, 10% glycerol
B5-Solubilization buffer (without detergent)	50mM Tris-HCl (pH 8), 150mM NaCl, 10% glycerol

B5- Solubilization buffer (with detergent)	50mM Tris-HCl (pH 7.8), 500mM NaCl, 10% glycerol, 2% DDM, 0.02% CHS
G1/G4-Lysis buffer	0.25M Tris-HCl (pH 7.6), 0.25M sucrose, 1mM EDTA, 2mM DTT
G1/G4- High Salt buffer	50mM Tris-HCl (pH 7.6), 500mM NaCl, 10% glycerol
G1/G4-Solubilization buffer (without detergent)	50mM Tris-HCl (pH 7.6), 150mM NaCl, 10% glycerol
G1/G4- Solubilization buffer (with detergent)	50mM Tris-HCl (pH 7.6), 500mM NaCl, 10% glycerol, 2%DDM, 0.02% CHS

Ni-NTA purification

B6 Purification buffer A	50mM Tris-HCl (pH 8), 20% glycerol, 0.1% DDM, 1mM DTT
B6 Purification buffer B	50mM Tris-HCl (pH 8), 20% glycerol, 0.1% DDM, 1mM DTT, 1M Imidazole
B5 Purification buffer A	50mM Tris-HCl (pH 7.8), 150mM NaCl, 10% glycerol, 0.1% DDM, 0.01% CHS, 1mM beta- Mercaptoethanol
B5 Purification buffer B	50mM Tris-HCl (pH 7.8), 150mM NaCl, 10% glycerol, 0.1% DDM, 0.01% CHS, 1mM beta- Mercaptoethanol, 500mM Imidazole
G1/G4 Purification buffer A	50mM Tris-HCl (pH 7.6), 150mM NaCl, 10% glycerol, 0.1% DDM, 0.01% CHS, 1mM beta- Mercaptoethanol
G1/G4 Purification buffer B	50mM Tris-HCl (pH 7.6), 150mM NaCl, 10% glycerol, 0.1% DDM, 0.01% CHS, 1mM beta- Mercaptoethanol, 500mM Imidazole

Size exclusion column (SEC)

B6 SEC Buffer	50mM Tris-HCl (pH 8), 20% glycerol, 0.1% DDM, 1mM DTT
B5 SEC Buffer	50mM Tris-HCl (pH 7.8), 10% glycerol, 0.1% DDM, 0.01% CHS
G1/G4 SEC Buffer	50mM Tris-HCl (pH 7.6), 4% glycerol, 0.1% DDM, 0.01% CHS
SAXS buffer	50mM Tris-HCl (pH 7.6), 150mM NaCl, 4% glycerol, 0.025% DDM, 0.005% CHS

Agarose gel electrophoresis

1X TAE buffer	40mM Tris-HCl (pH 7.6), 20mM acetic acid, 1mM EDTA
TE buffer	10mM Tris-HCl (pH 8), 1mM EDTA (pH 8)

SDS-PAGE

10X Running buffer	250mM Tris-HCl (pH8), 1.92M glycine, 1% SDS
1X Running buffer	25mM Tris-HCl (pH 8), 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
2X lamelli buffer	0.125M Tris-HCl ((pH 6.8), 4% SDS, 20% glycerol, 0.004% bromophenol blue, 10% 2-mercaptoethanol
6X lamelli buffer	0.375M Tris (pH 6.8), 12% W/V SDS, 60% V/V glycerol, 0.6M DTT, 0.06% bromophenol blue, 0.5ml aliquot, stored at -20ºC

Western blotting

1X Transfer buffer	25mM Tris-HCl (pH 8.3), 192mM glycine, 20% methanol
10X TBS	200mM Tris-HCl (pH 7.8), 1.5M NaCl
TBST	20mM Tris-HCl (pH 7.8), 150mM NaCl, 0.1% tween
Blocking buffer	5% dry milk in TBST buffer
Colloidal coomassie stain	0.02% G250, 5% Aluminium sulfate, 10% ethanol, 2% orthophosphoric acid
Destaining solution	10% ethanol, 2% orthophosphoric acid
Protease Inhibitor Cocktail (100X)	20mM AEBSF, 600μM bestatin, 400mM chymostatin, 700μM E-64, 2mM leupeptin, 1.5mM pepstatin A, 100mM PMSF in dry DMSO plus 300mM benzamidine in d.H2O

ATPase assay

Buffer A	12 % (w/v) SDS
Buffer B	1 % (w/v) ammonium molybdate
Buffer C	1M HCl
Buffer D	2 % (w/v) sodium citrate, 2 % (w/v) sodium metaarsenite, 1 % (v/v) acetic acid in dH_2O
Buffer E	6 % (w/v) ascorbate in 1 M HCl
Buffer F	Freshly prepared (1 : 1 mixture of buffer B and buffer E)
ATPase buffer	50mM Tris-HCl (pH 7.4), 150mM NH_4Cl, 5mM Mg_2SO_4
ATP stock	5mM in ATPase buffer

2.5 Yeast cells

Saccharomyces cerevisiae FGY217 yeast cells [*MATα ura3-52 lys2Δ201 pep4Δ]* were kindly provided by David Drew's lab at Imperial College London. For this study, yeast glycerol stock stored at -80 °C was used for transformation. FGY217 yeast cells cannot synthesize uracil (URA3 knockout) and have a *pep4* deletion which inhibits the expression

of the yeast vacuolar proteinases (Woolford et al., 1986). As a result, decreased degradation and improved expression of membrane proteins have been reported in this strain (Newstead et al., 2007). This yeast cell line has been used to express numerous eukaryotic membrane proteins for purification and crystallization (Drew et al., 2006).

2.6 Expression vector

The p424GAL1 expression vector was kindly provided by David Drew's lab at Imperial College London. For this research work, p424GAL1 plasmid DNA stored at -80 °C was used. This plasmid carries a URA3 gene as a selective marker and the GAL1 promoter (inducible by galactose). There is also a C-terminal tobacco etch virus (TEV) cleavable site, GFP-fusion tag that is codon-optimized for expression in yeast and is followed by an 8-His tag (Drew et al., 2006). The GFP fluorophore has peak excitation at a wavelength of 488 nm and an emission of 512 nm.

2.7 Constructs

Four ATP binding cassette genes have been synthesized.
1. Mouse ABCB5 (mABCB5)
2. Giant panda ABCB6 (gpABCB6)
3. Little brown bat ABCG1 (bbABCG1)
4. Mouse ABCG4 (mABCG4)

ProteoGenix, a life sciences company, synthesized all of the constructs (15 rue de la Haye, 67300, Schiltigheim, France). All of the protein sequences were taken from Uniprot database, reverse translated, and were codon-optimised for expression in *Saccharomyces cerevisiae* (Bai et al., 2011). All genes have a Yeast-like Kozak sequence, AAAACA, inserted before the start codon (ATG). The primary structure of the encoded protein was a determining factor in construct design. In ABCB5 and ABCB6 the transmembrane domain (TMD) is located at the N-terminus of the proteins, while the nucleotide-binding domain is located at the C-terminus, which is why tags were added at C-terminus. However, ABCG1 and ABCG4 have reverse topology, meaning that TMD is at the C-terminus while NBD is at the N-terminus, which is why tags were added at the N-terminus. In summary, the mouse *ABCB5* gene and giant panda *ABCB6* gene were introduced into the p424GAL1 expression vector between 5' BamH1 and 3' Xma1/ Sma1, downstream to galactose (GAL1) inducible promoter and upstream to GFP and 8His tag. Similarly, the *ABCG1* and *ABCG4* genes were

cloned under the galactose promoter using 5' BamH1 and 3' EcoR1 restriction sites. 8His and GFP tags, as well as TEV sites, were added to the ABCG1 and ABCG4 sequences at the N-termini. The TEV cleavage site, GFP, and 8His tags of the p424GAL1 vector were removed by restriction digestion with BamH1 and EcoR1 (Figure 2.1) (Table 2.1).
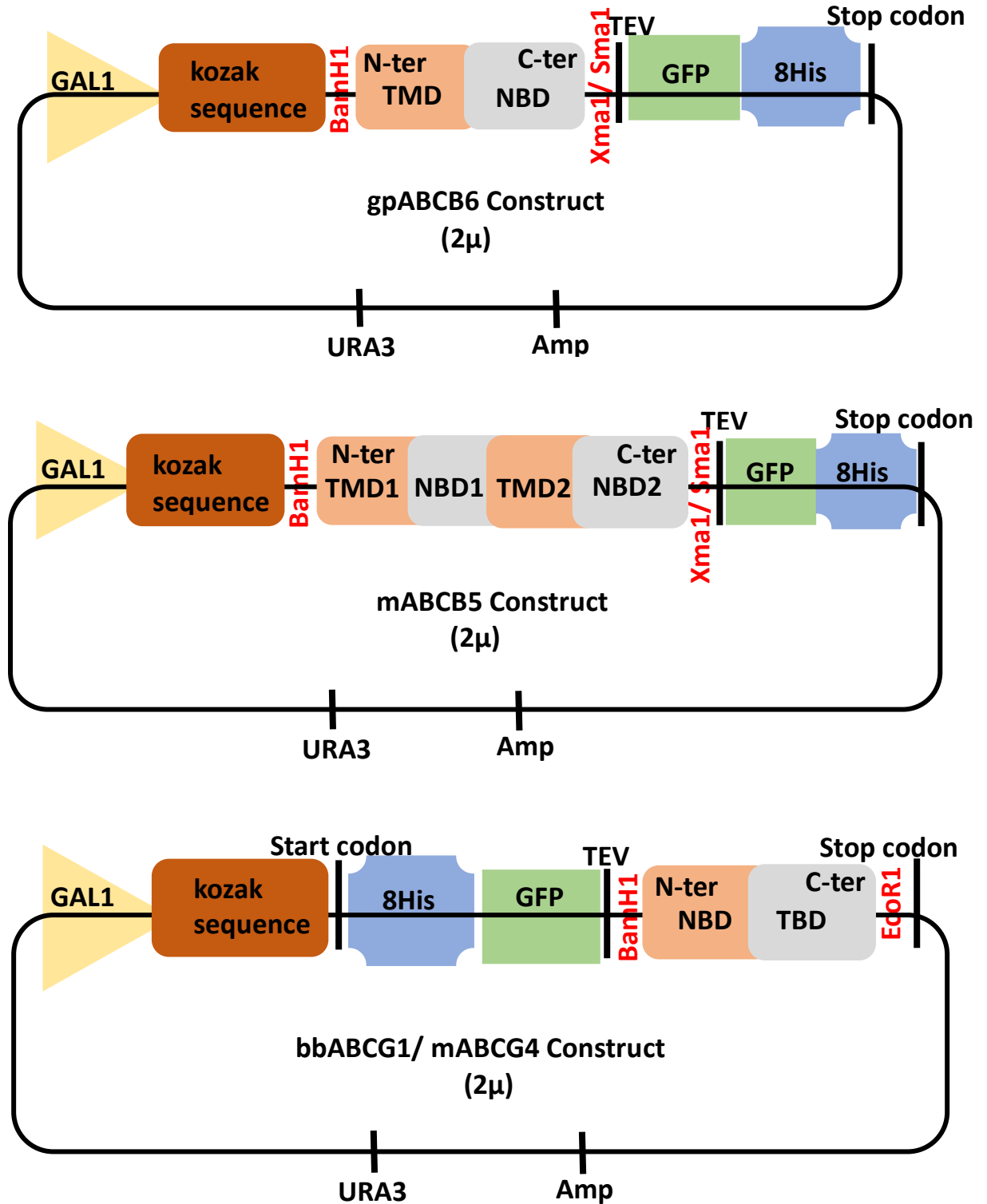
Figure 2.4: Cloning of mABCB5, gpABCB6, bbABCG1, and mABCG4 into p424GAL1 vector, a yeast expression vector. The vector harbors the uracil selection marker (URA3) and ampicillin-resistant gene (Amp) (Adopted from David Drew paper).

2.8 Methods

2.8.1 Bioinformatics

2.8.1.1 TMCRys screening

TMCrys is a freely available artificial intelligence (AI) server that predicts the likelihood of membrane protein crystallisation success (Varga and Tusnády, 2018). A total of 487 orthologs of the ATP binding cassette family (subfamilies ABCBA to ABCBD and subfamily ABCBG) were screened by TMCrys. Protein sequences were obtained from the Uniprot database and uploaded to the TMCrys server. The output of TMCrys was recorded, which was then evaluated and orthologs with high crystallisation score were selected for high-resolution structural studies. Moreover, a correlation graph between purification and crystallisation score was also plotted.

2.8.1.2 Homology Modeling

In this study, MODELLER 9.24 was used for homology modeling of all target proteins (ŠAli and Overington, 1994). The process of homology modelling consists of template selection, target-template alignment, model building, and model evaluation (Eswar et al., 2006). For selection of the templates, PSI-BLAST search against Protein Data Bank (PDB) was performed (Altschul et al., 1997) . The PDB file of each template was downloaded from PDB. The multiple alignment and alignment of target protein sequence to the template sequence was performed using the slign() command in MODELLER. The model of a target sequence was generated based on the alignment against the multiple templates using the "model-mult.py" file and the model was evaluated by Discrete optimized potential energy (DOPE) score (Eswar et al., 2006) using the "evaluate-model.py" file. At the end of the execution, a log file was created. MODELLER always produces a log file. Errors and warnings in log files can be found by searching for the _E> and _W> strings, respectively (Webb and Sali, 2016).

 Basic phython scripts were downloaded from the tutorial and modified according to target proteins.

2.8.1.2.1 Script 1 for Multiple Alignment

```
from modeller import *
log.verbose()
```

```
env = environ()
env.io.atom_files_directory = './:../atom_files/'

aln = alignment(env)
for (code, chain) in (('5do7', 'A'), ('5nj3', 'A'), ('6eti', 'A'), ('6hbu', 'A'),
('6hco', 'A')):
    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)

for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
                                    ((1., 0.5, 1., 1., 1., 0.), False, True),
                                    ((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
              rr_file='$(LIB)/as1.sim.mat', overhang=30,
              gap_penalties_1d=(-450, -50),
              gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,
              dendrogram_file='5do7_5nj3_6eti_6hbu_6hco.tree',
              alignment_type='tree', # If 'progresive', the tree is not
                                     # computed and all structues will be
                                     # aligned sequentially to the first
              feature_weights=weights, # For a multiple sequence alignment only
                                       # the first feature needs to be non-zero
              improve_alignment=True, fit=True, write_fit=write_fit,
              write_whole_pdb=whole, output='ALIGNMENT QUALITY')

aln.write(file='5do7_5nj3_6eti_6hbu_6hco.pap', alignment_format='PAP')
aln.write(file='5do7_5nj3_6eti_6hbu_6hco.ali', alignment_format='PIR')

aln.salign(rms_cutoff=1.0, normalize_pp_scores=False,
          rr_file='$(LIB)/as1.sim.mat', overhang=30,
          gap_penalties_1d=(-450, -50), gap_penalties_3d=(0, 3),
          gap_gap_score=0, gap_residue_score=0, dendrogram_file='1is3A.tree',
          alignment_type='progressive', feature_weights=[0]*6,
          improve_alignment=False, fit=False, write_fit=True,
          write_whole_pdb=False, output='QUALITY')
```

2.8.1.2.2 Script 2 for Alignment of target protein sequence to the template sequence

```
from modeller import *

log.verbose()
env = environ()

env.libs.topology.read(file='$(LIB)/top_heav.lib')

# Read aligned structure(s):
aln = alignment(env)
aln.append(file='5do7_5nj3_6eti_6hbu_6hco.ali', align_codes='all')
aln_block = len(aln)

# Read aligned sequence(s):
aln.append(file='qseq1.ali', align_codes='qseq1')

# Structure sensitive variable gap penalty sequence-sequence alignment:
aln.salign(output='', max_gap_length=20,
          gap_function=True,   # to use structure-dependent gap penalty
          alignment_type='PAIRWISE', align_block=aln_block,
          feature_weights=(1., 0., 0., 0., 0., 0.), overhang=0,
          gap_penalties_1d=(-450, 0),
          gap_penalties_2d=(0.35, 1.2, 0.9, 1.2, 0.6, 8.6, 1.2, 0., 0.),
          similarity_flag=True)

aln.write(file='qseq1-mult.ali', alignment_format='PIR')
aln.write(file='qseq1-mult.pap', alignment_format='PAP')
```

2.8.1.2.3 Script 3 for Model building

```
from modeller import *
from modeller.automodel import *

env = environ()
a = automodel(env, alnfile='qseq1-mult.ali',
              knowns=('5do7A','5nj3A','6etiA','6hbuA','6hcoA'), sequence='qseq1')
a.starting_model = 1
a.ending_model = 5
a.make()
```

2.8.1.2.4 Script 4 for Model evaluation

```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

# read model file
mdl = complete_pdb(env, 'qseq1.B99990001.pdb')

# Assess all atoms with DOPE:
s = selection(mdl)
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='qseq1.profile',
              normalize_profile=True, smoothing_window=15)
```

2.9 Preparation of p424GAL1 vector before cloning

2.9.1 Design of oligonucleotides

A206K mutation was introduced into the GFP gene of the p424GAL1 vector to prevent GFP dimerization (Shaner et al., 2007, Shaner et al., 2005). Primers were designed, sequences are given in table 2.1.

Table 2.1: Primer sequences

Primer Code	Position	Sequence (5' - 3')
A206K-F	Forward	CTCAATCTAAGCTTTCCAAAGATCCAA
A206K-R	Reverse	CTTTGGAAAGCTTAGATTGAGTGGAT

2.9.2 A206K site-directed mutagenesis

A206K site-directed mutagenesis PCR was performed using a high fidelity polymerase enzyme. A PCR reaction was set up as shown in Table 2.2. Conditions used for PCR are given in table 2.3. Cycles of the PCR system were set up to 15 to be able to get the product (Munteanu et al., 2012). Kinase, ligase, and DpnI (KLD) treatment was carried out

after amplification, these enzymes together allowed rapid circularization of PCR product and removal of template DNA (DeCero et al., 2020) (Table 2.5).

Table 2.2: Reaction set up of PCR

Reagents	Initial Concentration	Final Concentration	25 µl Reaction
Q5 hot start high fidelity Master mix	2 X	1 X	12.5 µl
Forward Primer	10 µM	0.5 µM	1.25 µl
Reverse Primer	10 µM	0.5 µM	1.25 µl
Template DNA (50 ng/µl)			1 µl
Nuclease free water			9 µl

Figure 2.3: Conditions of PCR

Steps	Temperature	Time	
Initial denaturation	95 °C	30 sec	
Final denaturation	94 °C	30 sec	15 Cycles
Annealing	55 °C	1 min	
Extension	70 °C	7 min	
Final Extension	70 °C	15 min	
Hold	4-10 °C		

Figure 2.4: Kinase, ligase, and DpnI (KLD) treatment

Reagents	Initial Conc	Final Conc	10µl Reaction
PCR product			1 µl
KLD Enzyme mix	10 X	1 X	1 µl
KLD reaction buffer	2 X	1 X	5 µl
Distilled water			3 µl

These reagents were mixed by pipetting and incubated at room temperature for 5 minutes before the transformation.

2.9.3 Transformation of A206K mutated p424GAL1 expression vector into competent E.Coli cells

The transformation of the PCR product was carried out by the heat shock method. 50 µl of DH5α cells and 5 µl of KLD mixture (section 2.9.2) were mixed in a pre-chilled tube by pipetting 4-5 times and kept on ice for 30 minutes and then heat shocked for 30 seconds at 42 °C in a water bath. The tube was immediately transferred on ice for 5 minutes. 950 µl of room temperature SOC media was added into the tube followed by incubation at 37 °C for 60 minutes with shaking at 250 rpm. Recipe of SOC media is given in section 2.3. Cells

were mixed thoroughly by pipetting and grown on an ampicillin selection plate overnight at 37 °C.

2.9.4 Mini Preparation and sequencing of p424GAL1

A single colony was taken from the plate and inoculated into LB media supplemented with 100 µg/ml ampicillin in a sterile 50 ml falcon tube and grown overnight in a shaker- incubator at 250 rpm at 37 °C. Once OD_{600} reached between 0.7-0.8, a mini preparation of plasmid was carried out using a Qiagen miniprep kit (Vogelstein and Gillespie, 1979). The miniprep plasmid DNA was confirmed on 1 % agarose gel. The concentration and quality were determined by NanoDrop 1000 spectrophotometer.

10 µl of 100 ng/µl plasmid DNA was sent to source bioscience company (1 Orchad Place, Nottingham Business Park, NGX 6P8, United Kingdom) for sequencing. The sequence was analyzed for A206K mutation. This vector was then sent to Proteogenix company for cloning.

2.10 Stock preparation of constructs

5 µg of recombinant DNA of all four constructs (mABCB5, gpABCB6, bbABCG1, mABCG4) were received from ProteoGenix in lyophilized form. 200 ng/µl stock was made by adding freshly prepared TE buffer, pH 8 (see section 2.3) and confirmed on 1 % agarose gel before the transformation.

2.11 Transformation of constructs

2.11.1 Transformation into competent E.Coli cells (DH5α)

The transformation of all of the four constructs was performed into highly competent DH5α cells by the heat shock method as described in section 2.9.3 (Rahimzadeh et al., 2016). Afterwards, mini preparation of recombinant protein containing plasmid was performed using Qiagen miniprep kit (Vogelstein and Gillespie, 1979).

2.11.2 Transformation into FGY217yeast cells

Transformation of all of the four constructs into FGY217 yeast cells was done by the lithium acetate method (Gietz and Woods, 2002). Briefly, FGY217 yeast cells were grown on the YPD agar plates (see section 2.3) to prepare a yeast culture for transformation. A single

colony of the FGY217 yeast strain was grown in 5 ml YPD medium (see section 2.3) in a shaker-incubator at 250 rpm overnight at 30 °C. The overnight culture was diluted to an OD_{600} of 0.1 into a pre-warm YPD media and the cells were grown under the same conditions until an OD_{600} of 0.8-1 was reached. The yeast cells were pelleted by centrifugation at 3,000 g for 10 min at room temperature and the cell pellet was resuspended in 1 ml dH_2O and transferred to a 1.5 ml Eppendorf tube and vortexed briefly. Cells were again centrifuged at maximum speed for 30 seconds at room temperature and the cell pellet was resuspended in 1 ml dH_2O. 100 µl is enough for a single transformation, so, 100 µl of yeast cells were transferred to a sterile 1.5 ml eppendorf tube and centrifuged at maximum speed for 30 seconds at room temperature and the supernatant was discarded. **Transformation mixture:** For each transformation, the transformation mixture was prepared by 240 µl of PEG 3500 50 % (w/w), 36 µl of 1 M LiAc, 50 µl of freshly prepared 2 mg/ml salmon sperm DNA and 1 µl of 100 ng/µl recombinant plasmid DNA (construct) + 33 µl H2O. The yeast pellet was resuspended in the transformation mixture and vortexed followed by heat shock for 25-30 minutes at 42 °C in a water bath. Cells were then micro centrifuged at 12000 g for 30 seconds at room temperature. The supernatant was discarded and the pellet was resuspended in 400 µl dH2O and spread onto on a YNB media plate supplemented with 2 % glucose (w/v) and grown under uracil selection, for 3-4 days at 30 °C. Plates were stored at 4 °C for up to one month.

2.12 SDS-PAGE

SDS-PAGE analysis was used to identify proteins based on their molecular weight throughout the expression and purification process (Shapiro and Maizel Jr, 1969, Smith, 1984). The separating gel was a standard 8 % polyacrylamide gel (Sambrook and Russell, 2006). A BioRad mini gel casting apparatus was used to cast the gels. After diluting 30 % bis-acrylamide (1:29) to 8 % in resolving buffer (0.4 % SDS, 1.5 M Tris-HCl, pH 8.8), the polymerization was initiated by adding 0.1 % APS and 0.2 % TEMED. In stacking buffer (0.4 % SDS, 0.5 M Tris-HCl, pH 6.8,), 4 % acrylamide, 0.1 % APS, and 0.2 % TEMED were used to make the stacking gel. Gels that had been polymerized could be kept in a wet condition for a week at 4 °C.

Protein samples were mixed with either 2X or 6X lamelli buffer to a final 1X concentration and incubated at 4 °C for 1 hour. The ingredients of 2X and 6X lamelli buffer are given in section 2.4. It was critical to prevent denaturation of the GFP tag by not heating the protein

sample and doing the gel electrophoresis slowly. Samples were loaded on 8 % homemade polyacrylamide gels and run at 120V for 1.5 hours (or until the forward dye front was near the edge of the gel) in a Mini-Protein SDS-PAGE BioRad system. 10 µl of pre-stained molecular weight marker was also loaded into separate lane per gel. For GFP detection, gels were scanned under Cys5 (695 nm emission filter), Cys3 (605 nm emission filter), and Alexa488 (530 nm emission filter) using BioRad Chemidoc MP imager before the Coomassie blue staining protocol. After that, Gels were stained using Colloidal Coomassie Blue stain for 3 hours and destained overnight. The ingredients of Colloidal Coomassie Blue stain and destaining solution are given in section 2.4. The white light was used to scan Coomassie-stained gels in the ChemiDoc MP imager.

2.13 Screening of yeast colonies

Transformed FGY217 cells were grown on YNB media and the screening of yeast colonies was performed in a similar way as described by O'Ryan *et al* (O'Ryan et al., 2012a). For each construct, 10-18 well-separated colonies were picked from the transformation plate using a sterile loop and grown overnight in 9 ml YNB media supplemented with 2 % glucose in a shaking incubator at 250 rpm, 30 °C. Glycerol stock of each of the screened colonies was made in 25 % glycerol (v/v) and stored at -80 °C. The overnight culture was used to inoculate 50 ml YNB supplemented with 0.1 % glucose (w/w). The level of glucose was measured using glucose strips (Medi-test Glucose, Macherey-Nagel, Germany) after every 1-2 hours. The culture was allowed to grow until both the glucose is completely consumed and an OD_{600} of 0.8-1.0 was reached. Before galactose induction, glucose consumption is essential for protein expression (Pullikuth and Gill, 2002). When glucose was completely consumed and an OD_{600} of 0.8-1 was reached , protein expression was induced by 2 % galactose and 8 % glycerol (v/v) which had previously been proven to boost the expression levels of other ABC proteins (O'Ryan et al., 2012a) and the culture was grown for 16 hours under the same conditions. Cells were harvested by centrifugation at 3,500 g, 4 °C for 10 minutes. The yeast pellet was re-suspended in lysis buffer supplemented with protease inhibitors and transferred to a 2 ml microfuge tube containing glass beads on ice. Cell lysis was done by mini bead beater for 3 minutes in cold room (4 °C).

Unbroken cells were removed by centrifugation at 3,500 g, 4 °C for 5 minutes. The supernatant was transferred to a clean eppendorf tube and microsomes (inner membrane from yeast cells) were collected by centrifugation at a maximum speed, 4 °C for 2 hours.

The membrane pellet was re-suspended in solubilisation buffer without DDM and mixed with 2X lamelli buffer in equal proportion and incubated at room temperature for 1 hour. Samples were loaded on 8 % polyacrylamide gel along with a pre-stained protein ladder and ran at 120V for 1.5 hours. The gel was blotted on nitrocellulose membrane (section 2.21.1). Before Western blotting, gels were scanned for detection of GFP signal under Cys3, Cys5, and Alexa488 using a BioRad Chemidoc MP imager as described in section 2.12. The highest expressing cells from the glycerol stock was streaked onto a fresh YNB plate and incubated for 3-4 days at 30 °C.

2.14 Time Course experiments

To determine the optimal expression time for each construct in *S cerevisiae*, a time-course study was conducted. A small-scale expression was carried out in a 2 L baffled flask. After glucose consumption, the cells were induced with 2 % galactose (w/v) and 8 % glycerol (v/v) as described in previous section 2.13. At each time point, 50 ml of induced culture was collected in a falcon tube. The cells were harvested by centrifuging them for 10 minutes at 3,500 g, 4 °C. Cell harvesting at each time point must be quick to prevent the cells from continuing to grow and express. The supernatant was discarded and yeast pellet were re-suspended in ice-cold lysis buffer. To limit yeast mobility, 15 μl of harvested cells were mixed with 15 μl of glycerol (50% v/v). The cells were placed on the glass slide and covered with the coverslip. Cells were analysed by a Zeiss Fluorescence microscope using a 60X objective lens. Images were taken with a Cool snap HQ2 CCD camera (Photo metrics) using Micromanager v1.4.23 software. As all constructs have GFP tag, therefore, set of images was taken at 100-200 m/s using a specific band pass GFP filter setting. Image J was used to analyse the images (Abràmoff et al., 2004, Rasband, 2011). The mean intensity of each field of view was calculated.

For each time point, microsomes were also prepared by method as described in section 2.13. SDS-PAGE was run at 120 V, for 1.5 hours and the gel was blotted on a nitrocellulose membrane (section 2.21.1). The densitometry data were generated for western blots by Image J software and a graph was plotted.

2.15 Large scale/ Shake-flask cell culture (12L)

For all four constructs, the same approach was used for large-scale protein expression. The only difference was in timing of yeast cells harvesting. After selecting the best expressing

colony of each construct, *S. cerevisiae* was re-grown from a glycerol stock at 30 °C for 3-4 days on uracil-deficient YNB agar plates. One litre of starting culture was prepared by inoculating a lawn of cells or a single colony from the plate into uracil-deficient YNB media. The culture was grown overnight in a shaking incubator at 230 rpm, 30 °C. The following day, twelve 2-liter baffled flasks, each flask containing 1 L of YNB media were inoculated with 80 ml of starting culture and grown under the same conditions. A spectrophotometer set to 600 nm was used to measure the optical density (OD) of cells periodically. Within 10 hours, cells usually reach an OD_{600} of 0.8-1.0. Glucose level was measured periodically using glucose strips (Medi-test Glucose, Macherey-Nagel, Germany) and protein expression was induced as described in section 2.13. For each construct, cells were continuously grown for a specific time that maximised protein expression (see section 2.14). Cells were harvested by centrifugation at 3,500 g for 10 minutes at 4 °C and then re-suspended in an ice-cold lysis buffer containing a protein inhibitor cocktail. The yeast cell suspension was kept at -80 °C until it was needed.

2.16 Microsomes preparation

Microsomes were prepared by slightly modifying the bead mill process as described by (O'Ryan et al., 2012a, Pollock et al., 2014a). Frozen yeast cells were thawed on ice. For the breakage of yeast cells, acid-washed glass beads (400-600 μm in diameter) were used in a 1:1 ratio (200 g glass beads: 200 g yeast cells). Bead beating was done in the steel chamber of a midi bead beater system (Biospec), in two-minute cycle with one-minute rest on ice for a total of 10-12 minutes of bead beating time. The outer portion of steel chamber was filled with dry ice during the process to maintain the sample as cold as feasible. The slurry mixture was allowed to remain on ice for 5 minutes after bead beating to sediment the glass beads from the slurry. To remove unbroken cells and cellular debris, the suspension was collected carefully and centrifuged at 14000 g for 10 minutes at 4 °C. To collect the microsomes (whole inner membrane fraction), the supernatant was centrifuged at 120,000 g for 90 minutes at 4 °C. The supernatant was then discarded, and the microsome pellet was re-suspended in high salt buffer with the help of a glass homogenizer and centrifuged 100,000 g for 45 minutes. The microsome pellet was re-suspended in solubilisation buffer with no detergent and total protein concentration was measured by Bradford assay. In-gel GFP scanning following electrophoresis (see section 2.12) and Western blotting with anti-His

antibody was used to verified the presence of the protein of interest in microsomes. The microsomes suspension was flash-frozen in liquid nitrogen and kept at -80 °C until used.

2.17 Microsomes solubilisation

To solubilize gpABCB6 microsomes, a stock solution of 10 % DDM (w/v) in solubilisation buffer was made and mixed with microsomes to a final concentration of 2.5 mg/ml total protein and 2 % DDM (w/v). For mABCB5, bbABCG1 and mABCG4 microsomes solubilisation, 2 % DDM: 0.02 % CHS was used. Solubilization was performed at 4 °C with gentle end-to-end rotation for 2 hours. The suspension was centrifuged at 100,000 g for 45 minutes at 4 °C and the supernatant which contains the solubilised microsomes was collected. The solubilized microsomes can be applied directly to a Ni-NTA affinity column for affinity purification or flash frozen and kept at -80 °C for long-term storage.

2.18 Immobilized metal affinity chromatography

A GE healthcare AKTA FPLC system with a Frac950 fraction collection system was used to purify mABCB5, bbABCG1, and mABCG4. Solubilized microsomes were applied to the 5 ml pre-equilibrated Ni-Affinity column with a flow rate of 0.5 ml/min. Flow-through was collected in a falcon tube for further analysis. To eliminate loosely bound contaminants, the column was washed with 10 column volume (CV) of 20-30 mM imidazole elution buffer B (wash 1) and then 8CV 50-80 mM imidazole elution buffer B (wash 2). At 200-250 mM imidazole concentration, the purified protein was eluted. Column was washed with 5CV of 500 mM elution buffer B to remove all bound proteins.

Ni-NTA loose resin (Q1Agen) was used to purify gpABCB6. The column was filled with 10 ml Ni-NTA loose resin and equilibrated with purification buffer A. solubilized microsomes were applied to nickel resin, and the column was incubated at 4 °C for two hours with end-to-end rotation. Flow-through was collected by gravity flow. Column was washed with 10 CV of 20 mM Imidazole elution buffer B (wash 1) and 5 column volume (CV) of 40 mM Imidazole elution buffer B (wash 2). The specifically bound protein was eluted using 150 mM Imidazole elution buffer B. Column was washed with 5 CV of 500 mM imidazole elution buffer B to remove all bound protein from the Ni-resin.

SDS-PAGE was used to analyse the fractions as described in section 2.12. The elution fractions containing purified protein were pooled and concentrated using a 100 kDa cut-off

centrifugal filter (Millipore). Concentrated protein was aliquoted into 500 μl aliquots and snap-frozen for multiple rounds of size exclusion chromatography.

2.19 Size Exclusion Chromatography (SEC)

A GE healthcare AKTA FPLC system with a Frac900 collecting system was used to perform size exclusion chromatography. The column (Generon SepFast™ 6-6000 kDa) was equilibrated with 2 column volume of degassed ultrapure water and followed by a 1.5-column volume of degassed SEC buffer. The sample was ultracentrifuged at 100,000 g at 4 °C for 15 minutes to remove aggregates and injected into a 500 μl super loop, which was then loaded onto the equilibrated column. Gel filtration was carried out at a flow rate of 0.5 ml/min, with 1 mL fractions collected for 1.2 column volume. UV280 absorbance was recorded to assess total proteins and GFP fluorescence was recorded to follow the target protein. SDS-PAGE was used to confirm the location of the target protein in eluted fractions (see section 2.12). Fractions enriched in target protein were pooled and concentrated with a 100 kDa cut-off (Millipore) filter before being aliquoted and snap-frozen for later use.

2.20 TEV Cleavage and negative purification of ABCG4

To improve the quality of purified mABCG4, negative purification was performed after TEV protease treatment. Nickel purified ABCG4 was concentrated and imidazole buffer was exchanged with size exclusion buffer using PD10 column followed by the addition of TEV protease (GenScript). Sephadex G-25 resin, which is found in PD-10 Desalting Columns, enables quick group separation of high molecular weight substances from low molecular weight molecules. The reaction was kept on end-to-end rotation overnight at 4 °C. On the following day reaction mixture was applied to a 5ml Nickel-affinity column with a flow rate of 0.1 ml/min and flow-through was collected, concentrated, and analysed on SDS-PAGE as described in section 2.12. After that, to recover the protein because TEV treatment is not always effective, 5 ml Nickel-affinity column was washed with 500 mM imidazole, 5CV.

2.20.1 Nanodisc preparation of ABCG4

The ABCG4 nanodiscs were made using the ABCG1 nanodisc method (Xu et al., 2022). Briefly, concentrated ABCG4 solubilized in DDM/CHS was reconstituted into lipid nanodiscs by mixing the ABCG4 with purified MSP1D1 scaffold protein and a DDM/CHS (0.5%, w/v) solubilized mixture of 20% 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-

serine (w/w) (POPS, Avanti), 60% 1-palmitoyl-2- oleoyl-glycero-3-phosphocholine (w/w) (POPC, Avanti) and 20 % cholesterol (Sigma Aldrich). The ABCG4-MSPID1-POPC-POPS-Cholestrol mixture was incubated at 4 ℃ for 1 hr, followed by addition of 0.7 g/mL of biobeads SM-2 and the mixture was gently rotated overnight at 14 ℃ to remove detergent. The following day, the biobeads were removed and the collected ABCG1 nanodiscs were injected over a Generon SepFast™ 6-6000 kDa column in cryo-buffer. Peak fractions were pooled and concentrated to 0.5 mg/mL for cryo-EM studies.

2.21 Protein quantification

SDS PAGE as described in section 2.12 was used to run serial dilutions of protein as well as standard dilutions of BSA. Overnight, the gel was stained with Colloidal-Coomassie stain. The stained gel was washed with the de-stain solution until there was no more background noise. The gel densitometry method was used to determine the protein concentration. As described at https://openwetware.org/wiki/Protein Quantification Using Image J, the image was scanned and put into Image J for quantification. In summary, the image was imported into the ImageJ programme, and protein band-containing lanes were chosen. For each lane, an intensity histogram plot was made. The area under the peak provides information about the band's intensity. A standard plot was generated using the calculated intensity of each standard band. The linear slope function of the standard plot was used to determine the enriched protein concentration.

2.22 Protein Characterization

2.22.1 Western blotting

The wet transfer method was used for Western blotting analysis (Mahmood and Yang, 2012, Liu et al., 2014). For SDS-PAGE, the samples were prepared and run as described in section 2.12. Transfer sandwich was made as follows: 2 sponges-2 filter papers-gel-nitrocellulose membrane-2 filter papers-2 sponges. It was made sure there was no air bubble between gel and nitrocellulose membrane and gel should be on the negative electrode while membrane towards the positive electrode. The sandwich was relocated to the transfer tank containing transfer buffer and run for 2 hours in a cold room with a constant current of 150 mA. The membrane was rinsed with TBST for 5 minutes on a shaker and afterward blocked with 5 % skimmed milk in TBST for 30 minutes (see section 2.4). Subsequently, the membrane was incubated with 1:10,000 primary antibody (HRP- conjugated anti-his) overnight at 4 ℃ with

gentle agitation. The membrane was rinsed three times with TBST for ten minutes. After washes, horseradish peroxidase (HRP) substrate (1:1 mixture of luminol solution and peroxide solution) was added for 30-45 seconds. Images were captured using a chemiluminescence filter on a BioRad Chemidoc MP imager system.

2.22.2 Mass spectrometry

Mass spectrometry was used to check the identity and homogeneity of purified proteins. 2 µg protein sample was run on 8% polyacryamide gels and stained with Colloidal Coomassie stain described in section 2.12. Bands of interest were excised and stored in MilliQ water in a micro-centrifuge tube. The MS facility (Michael Smith building, The University of Manchester) processed the samples *i.e.* dissolving the gel fragments and extracting the proteins (Gundry et al., 2010). Data was collected and analysed by MS facility staff. MASCOT Identity scoring was done based on the Swiss-Prot and TrEMBL databases. Further analyses (*e.g.* sequence analysis) was performed using Scaffold4 software (Proteome Software Inc.).

2.22.3 ATPase activity (Chifflet Assay)

This procedure is based on (Chifflet et al., 1988). In the first row of a 96 well plate, phosphate standards with concentrations of 0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 16, 20 nmol were set up to plot a standard curve. In 2nd row 10 µg of purified protein was added to each well to get a good signal. For substrate-stimulated ATPase activity, 10 µM substrate was also added in each well. The reaction was started by adding ATP. The plate was incubated at 25 °C for 25 minutes. Afterwards, in each well 40 µl of buffer A was added, followed by 100 µl of buffer F and incubated at room temperature for 5 minutes. Following incubation 100 µl of buffer D was added to the plate, and the plate was incubated for 15 minutes at 37 °C. The absorbance of the sample was measured in a spectrophotometer at 800 nm. The phosphate standard was plotted and the ATPase activity of the purified protein was calculated using the phosphate standard as a reference point. The components of buffer A-F are given in section 2.4.

2.22.4 Thermal stability assay [Measuring intrinsic tryptophan fluorescence and aggregation temperature (Tagg)]

The melting temperature (T_m), which reflects the temperature at which 50 % of the protein in the solution is denatured, was used to determine the thermal stability of the protein. T_m

was monitored by intrinsic tryptophan fluorescence and the onset of aggregation *i.e.* Tagg was monitored by static light scattering (SLS) (266 nm and 473 nm) using the UNcle instrument (Unchained Labs Inc.). Protein samples were diluted to 10 µg in size exclusion column buffer. Each experimental solution, which contained 10 µg protein and buffer was made in a final volume of 10 µl. 9 µl of each protein sample was loaded in a 16-well Uncle capillary cassette. The UNcle 267 nm and 473 nm laser was employed. The temperature was steadily increased by 1 °C/steps from 20°C to 95°C, with a 30- second hold period at each stage. The raw data was analyzed with MS Excel software, which included initial calculations and buffer subtraction. GradPad Prism 7 was used to plot the graphs.

2.22.5 Thermal stability (CPM) assay

The melting temperature (T_m), as described in section 2.21.3 was also measured by a thio-specific, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide, (CPM) dye. Protein samples were diluted to 10 µg in CPM buffer (100 mM Tris pH 8, 10 % glycerol, 150 mM NaCl, 0.1 % DDM and 0.02 % CHS). 4 mg/ml of CPM stock solution was prepared in DMSO and diluted to 0.025 mg/ml in CPM buffer before use (1:160 ratio). Each experimental solution, which contained 10 µg protein, buffer, 50 ng CPM, and additional additives such as nucleotides, was made in a final volume of 10 µl and incubated on ice for 15 minutes. The experiment was carried out with the UNcle spectroscopic apparatus as described in section 2.22.4. 9 µl of each solution were loaded in a 16-well Uncle capillary cassette. Although peak excitation and emission wavelengths of CPM dye are 384 nm and 470 nm, respectively, it can also be excited in the UV as it also absorbs strongly between 260 nm and 300 nm. Therefore, the CPM fluorescence was excited by using UNcle 266 nm laser (Platt et al.). The laser also activated the tryptophan fluorescence (Ghisaidoobe and Chung, 2014), providing a second readout of the protein state and enabling correction for the thermal quenching of the CPM fluorescence. The temperature was steadily increased by 1 °C/steps from 20 °C to 95 °C, with a 30-second hold period at each stage. The raw data was analysed with MS Excel software, which included initial calculations and buffer subtraction. GradPad Prism 7 was used to plot the unfolding profiles (Swift, 1997). The difference between experimental samples was statistically measured using an unpaired, two-tailed student's T-test. A difference is considered significant when the P-value is less than 0.05.

2.22.6 Dynamic light scattering

The Dynamic light scattering (DLS) was recorded prior to CPM assay at 20 °C using the Uncle spectroscopy apparatus. The duration of data gathering was set to 5 seconds. For each sample, four runs were done and the average was obtained for each sample.

2.22.7 Nucleotide-binding assay

ATP binding was examined using TNP-ATP fluorescence as reported by (LaConte et al., 2017). TNP-ATP is a fluorescent analogue of ATP, this compound hardly emits fluorescence alone, however, strongly fluoresces when it binds to the nucleotide-binding domain with a wavelength of 500-600nm. In a 96-well plate, 10 µg protein in each well was incubated with various TNP-ATP concentrations with a 50 µl final reaction volume. Biotek multi-mode spectrophotometer was used to record fluorescence emission spectra. The sample was excited at 410 nm, and the emission spectra were recorded at 500, 510, 520, 530, 540, 550, and 560 nm. GraphPad Prism 7 was used to analyses and plot the data.

2.22.8 Membrane thermal shift assay on mABCG4

A total reaction of 80 µl of solubilized microsomes (5 mg/ml) was set up. Solubilized microsomes were heated in a PCR machine at various temperatures (20-90 °C) for 3 minutes to establish a thermal denaturation curve to find out the temperature at which ~99% of protein loss occurs. After heating, solubilized microsomes were ultra-centrifuged at 100,000 g for 20 minutes at 4 °C to precipitate the denatured protein during the heating process. The supernatant was collected and run on 8 % polyacrylamide gel, scanned for GFP fluorescence, and subjected to Western blot analysis as described in section 2.12 and 2.21.2. A temperature of 70 °C was chosen for mABCG4 (which is this protein specific) based on an extrapolation from the thermal denaturation curve, which indicated 99% protein loss in the supernatant. To determine the relative thermal stabilization of ABCG4 by ligands, solubilized microsomes were first incubated with the ligands for 60 minutes at 37 °C and then heated at 70 °C for 3 minutes and subjected to western blot analysis. The immunoblot data were analyzed by Image J software and the graph was plotted using GraphPad Prism 7.

2.23 Structural studies of proteins

2.23.1 Negative staining electron microscopy (EM)

Negative stain electron microscopy was used to assess the homogeneity of the purified protein and for initial structural validations. Protein samples were prepared according to previously established protocols with slight modifications (Ohi et al., 2004, Booth et al., 2011). Using a K100X glow discharge device, 300-mesh carbon-coated copper grids were glow discharged at 25 mA for 45 seconds. Purified protein was diluted in SEC buffer to 20-50 µg/ml and applied to the carbon-coated side of the grid. The grid was rinsed three times with ultrapure water and stained with 2 % uranyl acetate (w/v) solution for 30 seconds. The grids were blotted with filter paper between each step. Thermo Fisher Talos L120C transmission electron microscope with Ceta CMOS camera was used to screen the grids (EM facility, FBMH). Until grids are screened by a microscope, grids can be kept at room temperature in a grid box.

2.23.2 Small Angle X-ray Scattering (SAXS) analysis

SAXS analysis was performed *via* SEC-SAXS for mABCB5 and bbABCG1 proteins while *via* batch mode for mABCG4 at the Diamond light source (Didcot, UK). The sample was exposed to $2x10^4$ photons/sec of radiation, and scattering was recorded on an EigerX 4M detector at 13 keV with a fixed camera length of 4.014m. Angular q values of 0.0045 to 0.34 $Å^{-1}$ were used to obtain data. Generic Data Acquisition software was used to record the data in NXS format (Cowieson et al., 2020), which was then converted to 1D (.dat) profiles using Data Analysis Workbench (DAWN).
The data was analysed using ScÅtter (Förster et al., 2010) and ATSAS (Franke et al., 2017). The FoXS algorithm (Schneidman-Duhovny et al., 2010) implemented in the Chimera software (Pettersen et al., 2004) suite was used to compare theoretical SAXS curves to experimental curves.

2.22.3 Cryo-electron microscopy (CryoEM)

2.22.3.1 Grid preparation

Cryo specimens were prepared by freezing biological samples in vitreous ice with liquid ethane as a fast chiller, as described previously (Thonghin et al., 2018a). 200 mesh 1.2/1.3 carbon-coated copper grids were washed in chloroform and allowed to dry overnight. Before

applying protein to the grid, the sample buffer was exchanged with cryo buffer by diafiltration using 100 kDa cut-off concentrators. Grids were glow discharged at 25 mA for 2 minutes using a K100X glow discharge device before loading and freezing. The glow discharged grid was loaded into the FEI vitrobot Mark IV using cryo tweezers. The chamber was kept at a constant humidity of 95% and a temperature of 20°C. 4 µl of 2.5-10mg/ml protein sample was applied on the carbon side of the grid with a range of blotting force 2-8. To achieve even spreading of protein particles on the grids, several blotting times ranging from 3-6 seconds were tested. The grids were immediately plunged into liquid ethane. The grids were kept in grid boxes and stored in liquid nitrogen until they were screened.

2.22.3.2 Data acquisition

Thermo Fisher Glacios Cryo-TEM was used to gather the data. Dr Richard Collins assisted with grid loading and screening (senior experimental scientist, FBMH). The data was collected at a magnification of 120000X, which corresponded to a pixel size of 1.3 Å/pixel. The data was collected using the EPU (FEI firm) automated data collecting programme at a high electron dose of 60 ē and a defocus of -2 µm to -3 µm.

Chapter 3 Results and Discussion: Bioinformatics

In 1970, a Dutch biologist named Paulien Hogewe introduced the term "bioinformatics" to describe the use of information technology to study biological systems (Hesper and Hogeweg, 1970, Hogeweg, 2011). Bioinformatics is a multidisciplinary field that integrates computer science, statistics, and mathematics to record, recover and analyze biological data cost-effectively and quickly (Mehmood et al., 2014). Computational tools are now commonly used to characterize a gene, do phylogenetic analysis, determine physicochemical properties of proteins, predict 3D structures of proteins, run simulations to learn about biomolecule interactions in living cells, and design drugs. The user-friendly interactive automated modeling and SWISS-MODEL server were established about 18 years ago, resulting in rapid expansion in homology modelling (Peitsch, 1996).

For this research work, bioinformatics has been applied to study the ATP binding cassette (ABC) transporter family. Structural details of ABC proteins may facilitate drug discovery by giving diverse targets for ligand screening. However, ABC membrane proteins are found in the cell membrane, making structure determination extremely challenging. The protein data bank (PDB) has a minority of membrane proteins (Varga and Tusnády, 2018). To fully characterize a protein, purified protein is required (Lin and Guidotti, 2009). Obtaining purified protein opens many ways for structural and functional characterization. 3D protein structures are typically determined using X-ray crystallography, Cryoelectron Microscopy (cryo-EM), and Nuclear Magnetic Resonance (NMR). The highest resolution ABC protein structures have been obtained by crystallography, however, protein crystallization is challenging. For example, nucleation of crystal growth can be stochastic and often depends on microscopic impurities in the crystallization droplet. A good-quality protein crystal is obtained after a series of experiments. Initially, the DNA sequence must first be cloned into a suitable expression system and then sufficiently overexpressed. Transmembrane proteins (TMPs) are difficult to express in large quantities, in a functional state and this is the first major stumbling block in functional and structural research (Tate, 2001). Difficulties to express TMPs can arise due to several reasons, such as, when TMPs are expressed, the proteins may be lethal to the host, or the TMPs may remain inactive because they are not folded into a functional conformation (Miroux and Walker, 1996). After the protein has been produced, it must be purified for further analysis. Membrane protein purification is more difficult than soluble protein purification. The purification process can be influenced by a variety of parameters, and trial and error is required to optimize the protocol. Firstly, the TMPs are

solubilized using suitable detergent or detergent mixtures for purification. The optimal detergent should be chosen, which is generally done by screening a detergent library (Linke, 2009). Secondly, the desired protein is then isolated from the cellular proteome by affinity chromatography. The most difficult aspect of the purification method is usually getting a homogeneous and ~95-99% purified protein to create a good-quality crystal. The crystal is then illuminated with X-rays, and diffraction data is recorded. An electron density map can be generated provided that the phase problem can be solved, and then the known sequence of the protein can be built into the density map to generate an atomic model of the structure if the diffraction data resolution is sufficient. An outline of the whole process of structure determination is shown in figure 3.1.
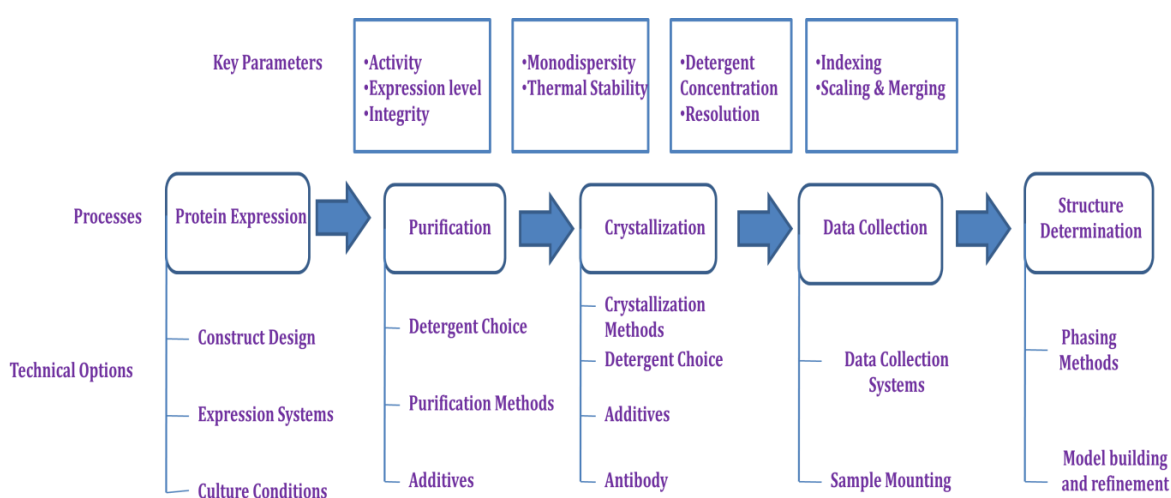
Figure 3.1: An outline of structure determination process with important parameters

Protein crystallography targets must be selected carefully. It is preferable to avoid time-consuming and costly studies using proteins that are unlikely to provide a good quality crystal. *In silico* studies can facilitate the selection of proteins for crystallization. Crystallization prediction tools can be used before starting X-ray crystallography experiments with a given protein to get an idea of crystallization success. Numerous software has been developed in recent years to determine the likelihood of protein crystallization. CRYSTALP2, SECRET, ParCrys, OB-Score, XtalPred algorithm, and TMCrys server are the most well-known tools (Kurgan and Mizianty, 2009, Kurgan et al., 2009, Overton and Barton, 2006, Slabinski et al., 2007, Varga and Tusnády, 2018). CRYSTALP2 and SECRET both accept amino acids with lengths ranging from 46 to 200. Although the OB- score has no restriction on the length of amino acid sequences, it only considers two variables (hydrophobicity

and PI), which restricts the accuracy of its predictions. Both XtalPred and ParCrys are sophisticated servers (Varga and Tusnády, 2018).

The **TMCrys server**, released by the Hungarian Academy of Sciences in October 2018, was selected for this study (Varga and Tusnády, 2018). TMCrys is specifically designed for membrane proteins to help the crystallisation of TMPs by predicting the probability of success of solubilisation, purification, and crystallization of TMPs based on their amino acid sequence, as well as, the success of all three steps taken combined. The TMCrys server takes FASTA sequences with or without predicted topology and provides the output. To avoid overburden, up to ten amino acid sequences can be uploaded to TMCrys as a single file (Varga and Tusnády, 2018). In-house scripts and several software packages are used by TMCRys to identify the various features as follows:

1- **CCTOP:** for predicting membrane protein topology and signal peptide

2- **Amino acid composition:** for estimating the transmembrane and extramembrane regions

3- **ProtParam:** for determining the molecular weight, half-life, and isoelectric point of proteins

4- **NetSurfP:** for predicting the exposed and buried residues, as well as, solvent accessible surface area

5- **Sequence motifs:** for determining the Glycosylation [NX(ST)], and alpha-helical interaction

CCTOP predicts whether a sequence is a transmembrane protein. If not, a yellow panel with the word "notTMP" in the left corner appears. Transmembrane protein panels are either green (when the outcome is predicted to be successful) or red (when the outcome is predicted to be unsuccessful) as shown in figure 3.2.
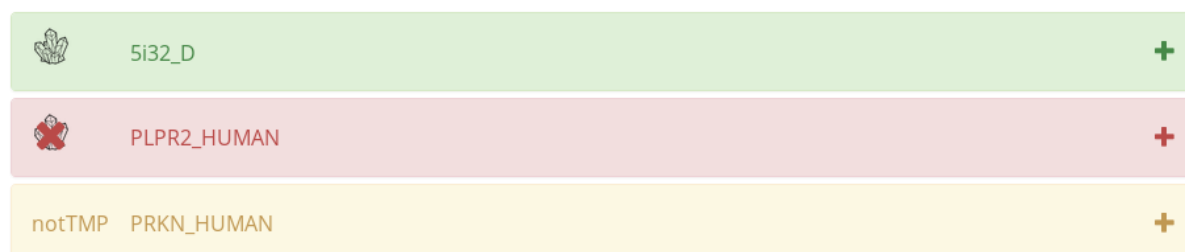
Figure 3.2: TMCrys result panel

The expected values of the various crystallisation steps are shown as slider diagram (Figure 3.3). The range of the values is [0, 1]. A yellow region between the diagram's two sides designates the threshold used for classification. A vertical blue line denotes the value of the actual classification.
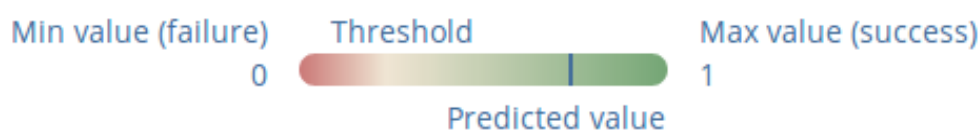
Figure 3.3: TMCrys server slider diagram

Distinct orthologs of the eukaryotic ATP binding cassette (ABC) family (ABCA-ABCBD and ABCG) were screened by the TMCrys server. The orthologs were selected based on their sequence similarity to a human member. The orthologs studied cover a wide range of chordate species, with protein sequence homology to human protein members ranging from 46% to 93%. The rationale behind this approach is that the natural divergence in amino acid sequences could result in an ortholog that is more stable, soluble, and accessible to expression, purification, and structural and functional research than the human ortholog. For example previous data from the Ford lab suggest that possible differences in protein structure affect the stability of different CFTR orthologs (O'Ryan et al., 2012a).

3.1 TMCrys Screening of ABCA Family

An analysis of 131 ABCA orthologs was performed using the TMCrys server. The number of orthologs of each sub-class selected for crystallization prediction is shown in Figure 3.4. The likelihood of getting a purified protein and subsequently good quality crystals were estimated for three sub-families *i.e.* ABCA2, ABCA5, and ABCA12 as shown in table 3.1. The sequence similarity of these orthologs to the human member was determined using NCBI pairwise alignment (Johnson et al., 2008). The appendix has detailed TMCrys results for the ABCA family (Appendix Table 3.1).
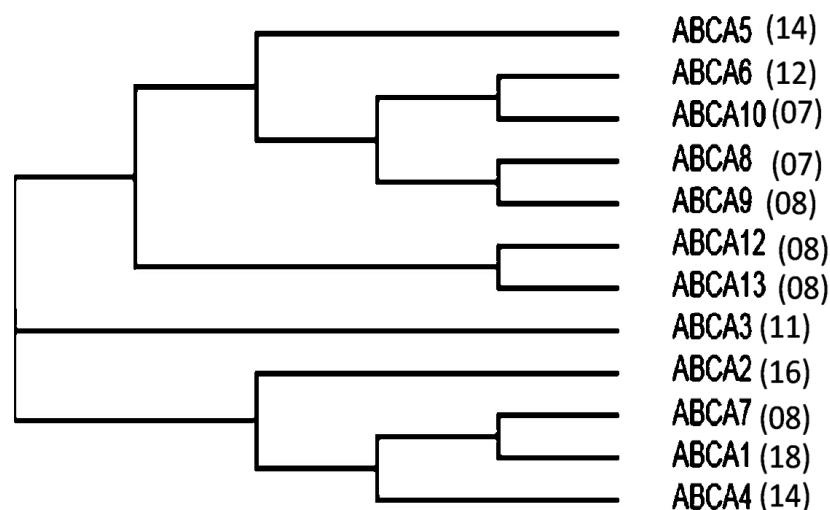
Figure 3.4: TMCrys screening of ABCA family. Number of orthologs of each subclass selected for TMCrys screening is given in brackets. Protein sequences were obtained from the Uniprot database UniProt.

Table 3.1: Representation of successful hits (high crystallization propensity) of the ABCA family. The ortholog protein sequence was compared to human protein using NCBI BLASTp Programme. Protein BLAST: search protein databases using a protein query (nih.gov).

Organisms	Crystallization score	Sequence similarity to human protein
ABCA2 subfamily		
Ictidomys tridecemlineatus (Ground Squirrel)	0.698	92%
Otolemur garnettii (Small-eared galago)	0.695	93%
Anas platyrhynchos (Mallard)	0.562	79%
Taeniopygia guttata (Zebra finch)	0.635	81%
Danio rerio (Zebra fish)	0.6259	77%
ABCA5 subfamily		
Ornithorhynchus anatinus (Duckbill Platypus)	0.716	84%
ABCA12 subfamily		
Ornithorhynchus anatinus (Duckbill Platypus)	0.763	57%
Taeniopygia guttata (Zebra finch)	0.749	46%

3.2 TMCrys Screening of ABCB Family

TMCrys screening was applied to 77 orthologs of ABCB family (Figure 3.5). ABCB1-7 and ABCB11 orthologs with a high crystallization score were identified (Table 3.2). The appendix contains detailed results for the entire ABCB family (Appendix Table 3.2)
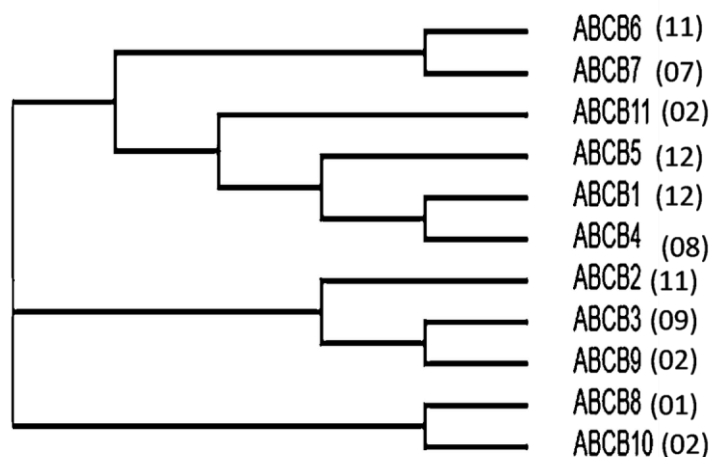
Figure 3.5: TMCrys screening of ABCB family. The number of orthologs of each subclass selected for TMCrys screening is given in brackets. Protein sequences were obtained from the Uniprot database UniProt.

Table 3.2: Representation of successful hits (high crystallization propensity) of the ABCB family. The ortholog protein sequence was compared to human protein using NCBI BLASTp Programme. Protein BLAST: search protein databases using a protein query (nih.gov).

Organisms	Crystallization score	Sequence similarity to human protein
ABCB1 subfamily		
Rattus norvegicuas (Rat 1a)	0.555	87%
Equus caballus (Horse)	0.605	89%
Pan troglodytes (Chimpanzee)	0.573	99%
Ailuropoda melanoleuca (Giant Panda)	0.646	90%
Myotis lucifugus (Little Brown Bat)	0.515	88%
Ictidomys tridecemlineatus (Ground Squirrel)	0.683	85%
ABCB2 subfamily		
Canis lupus familiaris (Dog)	0.685	76%

Oryctolagus cuniculus (Rabbit)	0.5	78%
Gallus gallus (Chicken)	0.543	50%
ABCB3 subfamily		
Gallus gallus (Chicken)	0.659	50%
Canis lupus familiaris (Dog)	0.874	42.44%
Oryctolagus cuniculus (Rabbit)	0.846	42.46%
ABCB4 subfamily		
Rattus norvegicuas (Rat)	0.603	90%
Oryctolagus cuniculus (Rabbit)	0.489	91%
Ovis aries (Sheep)	0.913	88%
ABCB5 subfamily		
Mus musculus (Mouse)	0.591	78%
Oryctolagus cuniculus (Rabbit)	0.5	83%
ABCB6 subfamily		
Ailuropoda melanoleuca (Giant Panda)	0.867	87%
ABCB7 subfamily		
Equus caballus (Horse)	0.88	93%
ABCB11 subfamily		
Rattu snorvegicuas (Rat)	0.4	82%

3.3 TMCrys screening of ABCC Family

TMCrys was used to evaluate a total of 137 orthologs of ABCC family (Figure 3.6). Subfamilies ABCC1, ABCC3, ABCC4, ABCC5, and ABCC8 were found to have good crystallization scores. Table 3.3 shows the crystallization score and sequence similarity to the human protein, whereas the appendix Table 3.3 shows the TMCrys score for the entire ABCC family.
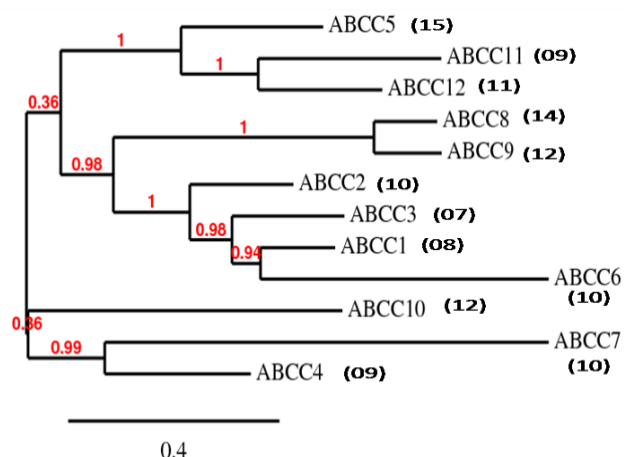
Figure 3.6: TMCrys screening of ABCC family. The number of orthologs of each subclass selected for TMCrys screening is given in brackets. Protein sequences were obtained from the Uniprot database. *UniProt.*

Table 3.3: Representation of successful hits (high crystallization propensity) of the ABCC family. The ortholog protein sequence was compared to human protein using NCBI BLASTp Programme Protein BLAST: search protein databases using a protein query (nih.gov).

Organisms	Crystallization score	Sequence similarity to human protein
ABCC1 subfamily		
Bos Taurus (Bovine)	0.571	90%
ABCC3 subfamily		
*Rattus norvegicuas (*Rat)	0.79	78%
Canis lupus familiaris (Dog)	0.597	84%
Pan troglodytes (Chimpanzee)	0.64	95%
Equuscaballus (Horse)	0.608	81%
ABCC4 subfamily		
Ailuropoda melanoleuca (Giant Panda)	0.667	88%
ABCC5 subfamily		
Myotis lucifugus (Little Brown Bat)	0.657	94%
ABCC8 subfamily		
Equus caballus (Horse)	0.574	96%
Ailuropoda melanoleuca (Giant Panda)	0.608	95%

3.4 TMCrys screening of ABCD Family

TMCrys was used to screen 58 distinct orthologs of ABCD family to predict crystallization scores. The number of orthologs of each sub-family chosen for TMCrys screening is shown in Figure 3.7. Table 3.4 shows the crystallization score and sequence similarity to the human protein. The TMCrys results for the entire ABCD family are given in the appendix Table 3.4.
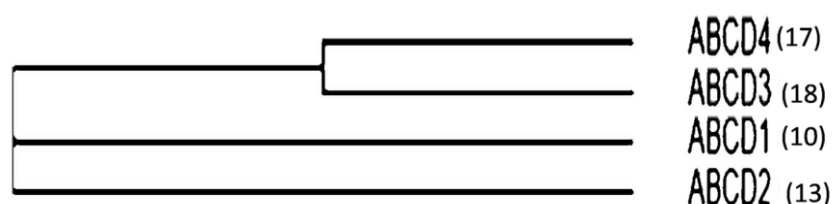
Figure 3.7: TMCrys screening of ABCD family. The number of orthologs of each subclass selected for TMCrys screening is given in brackets. Protein sequences were obtained from the Uniprot database. *UniProt.*

Table 3.4: Representation of successful hits (high crystallization propensity) of the ABCD family. The ortholog protein sequence was compared to human protein using NCBI BLASTp Programme. Protein BLAST: search protein databases using a protein query (nih.gov).

Organisms	Crystallization score	Sequence similarity to human protein
ABCD1 subfamily		
Rattus norvegicuas (Rat)	0.4	91%
ABCD2 subfamily		
Ornitho rhynchusanatinus (Duckbill platypus)	0.652	77%
Oryctolagus cuniculus (Rabbit)	0.5	93%
ABCD3 subfamily		
Equus caballus (Horse)	0.705	96%
Taeniopygia guttata (Zebra finch)	0.85	89%
Ovis aries (Sheep)	0.66	88%
Anas platyrhynchos (Mallard)	0.884	86%

Ornitho rhynchusanatinus (Duckbill platypus)	0.851	94%
ABCD4 subfamily		
Ictidomys tridecemlineatus (Ground Squirrel)	0.732	89%
Taeniopygia guttata (Zebra finch)	0.45	77%
Oryctolagus cuniculus (Rabbit)	0.50	90%
Loxodonta africana (African elephant)	0.52	90%

3.5 TMCrys screening of ABCG Family

Out of 75 ABCG orthologs, successful hits were found in ABCG1, ABCG2, ABCG5, and ABCG8 (Figure 3.8). Orthologs with good crystallization propensity scores are shown in Table 3.5; while TMCrys score of the whole ABCG family is shown in the appendix Table 3.5.
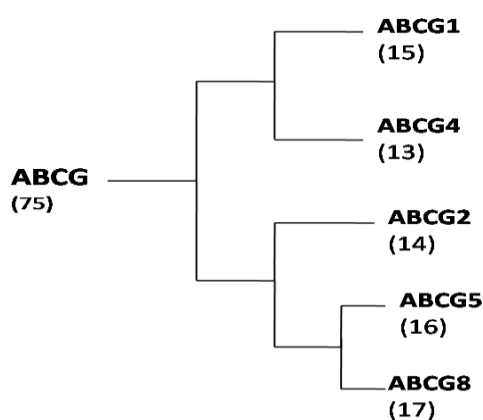
Figure 3.8: TMCrys screening of ABCG family. The digits in brackets show the number of orthologs screened for crystallization by the TMCrys server.

Table 3.5: Representation of successful hits (high crystallization propensity) of the ABCG family. The ortholog protein sequence was compared to human protein using NCBI BLASTp Programme. Protein BLAST: search protein databases using a protein query (nih.gov)

Organisms	Crystallization score	Sequence similarity to human protein
ABCG1 subfamily		

Gallus gallus (Chicken)	0.881	91%
Ornith rhynchus anatinus (Duckbill platypus)	0.877	93%
Myotis lucifugus (Little Brown Bat)	0.897	87%
ABCG2 subfamily		
Canis lupus familiaris (Dog)	0.646	81%
Equus caballus (Horse)	0.682	86%
Anas platyrhynchos (Mallard)	0.674	69%
ABCG4 subfamily		
Mus musculus (Mouse)	0.92	96.6%
ABCG5 subfamily		
Monodelphis domestica (Opossum)	0.815	75%
Taeniopygia guttata (Zebra finch)	0.736	65%
ABCG8 subfamily		
Equus caballus (Horse)	0.741	84%
Monodelphis domestica (Opossum)	0.45	70%
Ornitho rhynchusanatinus (Duckbill platypus)	0.752	62%
Taeniopygia guttata (Zebra finch)	0.649	65%
Anas platyrhynchos (Mallard)	0.578	65%

3.6 TMCRYs screening after Methionine deletion at the first position

TMCrys was used to predict the probability of success of the solubilisation, purification, and crystallisation process of number of orthologs of the entire ABC family, excluding ABCE and ABCF. It was noted that by changing the N-terminal amino acid of proteins, crystallization propensity scores changed substantially. After observing this variation, the TMCrys scoring of the entire family excluding ABCE and ABCF was repeated by removing Methionine (M) from the N-terminus. The average TMCrys score across all screened species is shown on a graph for each ABC transporter family. Within ABCA: The ABCA3 sub-family scored well overall, whereas ABCA13 had the lowest score (Figure 3.9). The percent identity of human ABCA3 and human ABCA13 sequence is 33.95%. In the ABCB family,

the highest-scoring family is ABCB1, while ABCB8 is the least-scoring family (Figure 3.10). The percent identity between human ABCB1 and human ABCB8 is 38.09%. In the ABCC family, both human ABCC4 and human ABCC11 are the highest scoring, on the other hand, ABCC2 and ABCC5 are the least scorings. The percent identity between human ABCC4 and human ABCC11 is 34.49% while the percent identity between ABCC2 and ABCC5 is 42.81% (Figure 3.11). In the ABCD family, ABCD3 appeared as an excellent candidate while ABCD1 least promising candidate (Figure 3.12). In the ABCG family, ABCG1 is a good candidate while ABCG2 is the least-scoring candidate for crystallization (Figure 3.13).
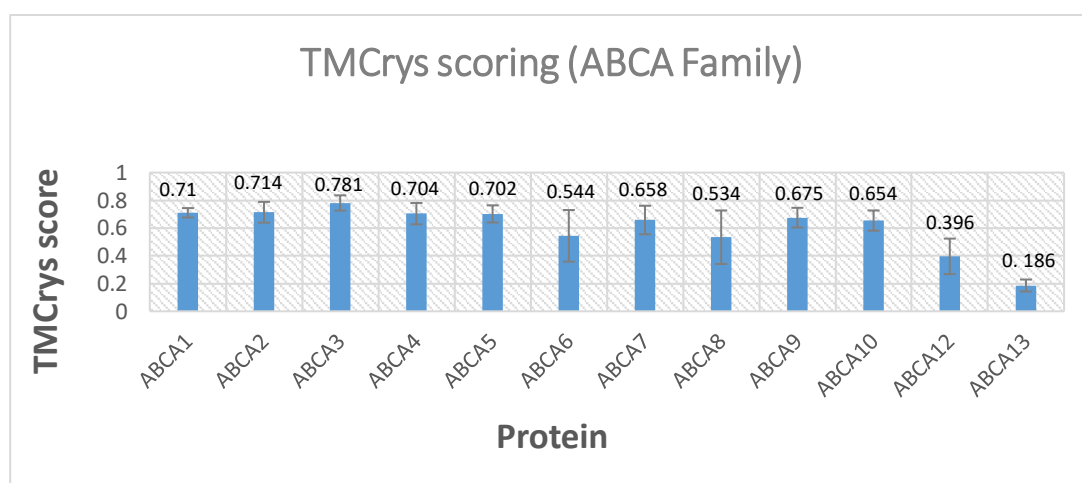
Figure 3.9: Graphical representation of TMCrys scoring of ABCA family after Methionine deletion at first position. Mean crystallisation score of ~20 orthologs of each subfamily with standard deviation is shown.
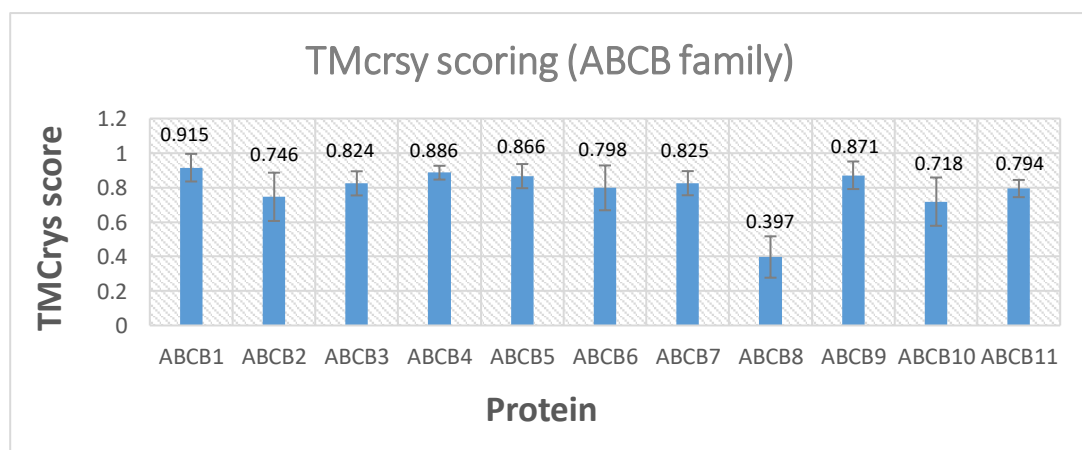
Figure 3.10: Graphical representation of TMCrys scoring of ABCB family after Methionine deletion at first position. Mean crystallisation score of ~20 orthologs of each subfamily with standard deviation is shown.
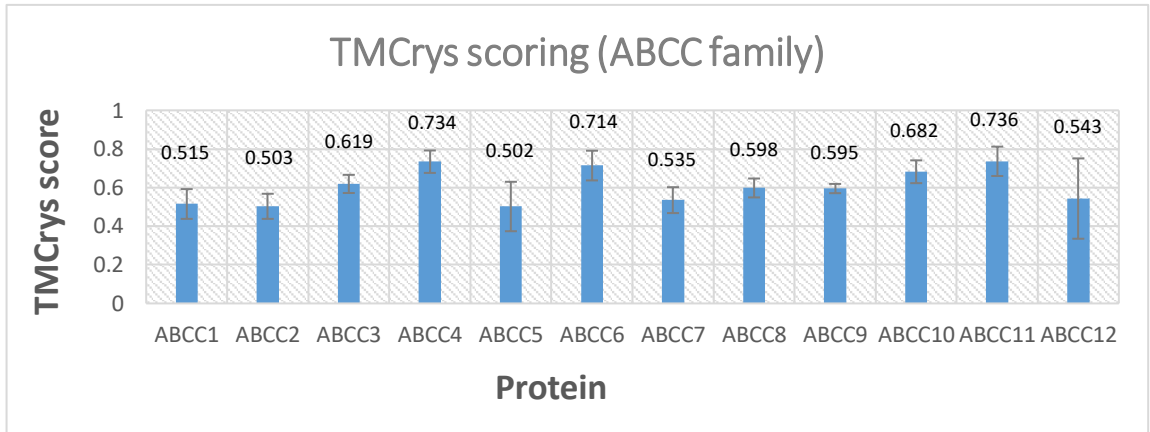
Figure 3.11: Graphical representation of TMCrys scoring of ABCC family after Methionine deletion at first position. Mean crystallisation score of ~20 orthologs of each subfamily with standard deviation is shown.
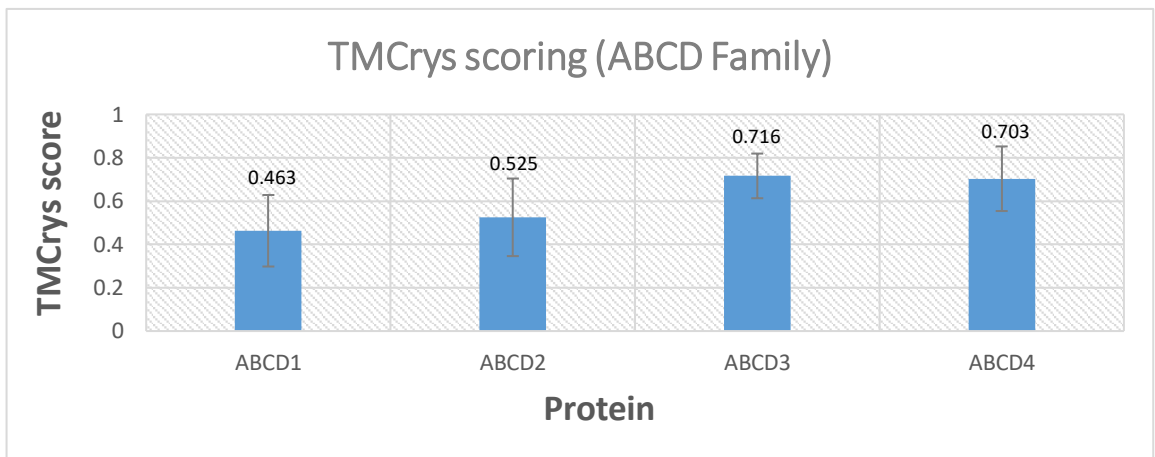
Figure 3.12: Graphical representation of TMCrys scoring of ABCD family after Methionine deletion at first position. Mean crystallisation score of ~20 orthologs of each subfamily with standard deviation is shown.
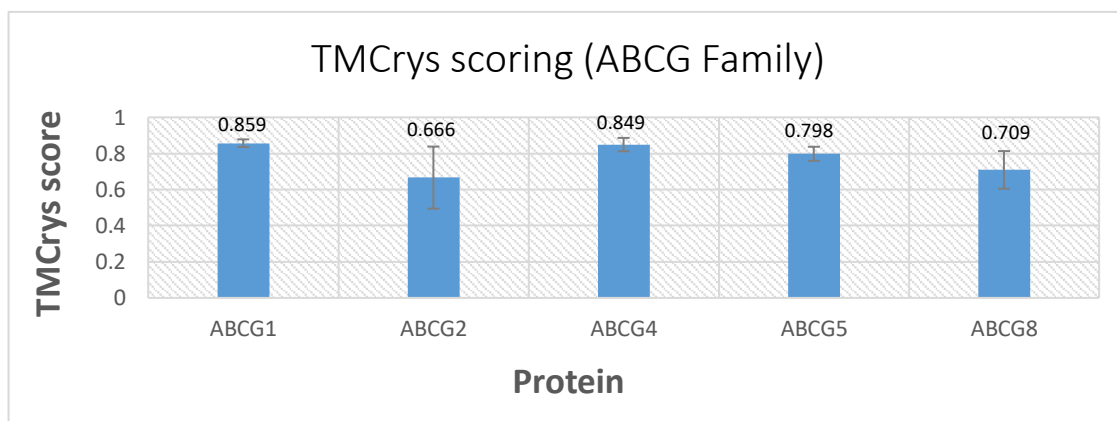
Figure 3.13: Graphical representation of TMCrys scoring of ABCG family after Methionine deletion at first position. Mean crystallisation score of ~20 orthologs of each subfamily with standard deviation is shown.

3.7 Correlation between crystallization and purification score

To explore the correlation between purification and crystallisation scores, a graph was plotted for each ABC sub-family. In the ABCA family, a positive correlation was found between purification and crystallisation scores with exception of ABCA12 and ABC13 orthologs (Figure 3.14). ABCA7 orthologs showed widely spread purification and crystallisation scores but had a positive correlation which means high purification as well as high crystallisation scores. Similar to the ABCA subfamily, a positive correlation was found between the purification and crystallisation score of the ABCB subfamily with exception of ABCB8 orthologs. For ABCB8 sub-family, a high probability of getting purified protein but less probability of getting crystal was predicted by the TMCrys server (Figure 3.15). The crystallisation and purification scores of the ABCC family were concentrated in one area and had a direct relation to scores (Figure 3.16). However, ABCC7 orthologs showed variations of 0.5-0.9 in the purification score. Purification and crystallisation scores of the ABCD family were widely spread and had no correlation between purification and crystallisation scores (Figure 3.17). ABCG family behaved best, purification and crystallisation score was concentrated in one region which is consistent with the ABCC family and a strong positive correlation was found which means a high purification as well as high crystallisation scores.
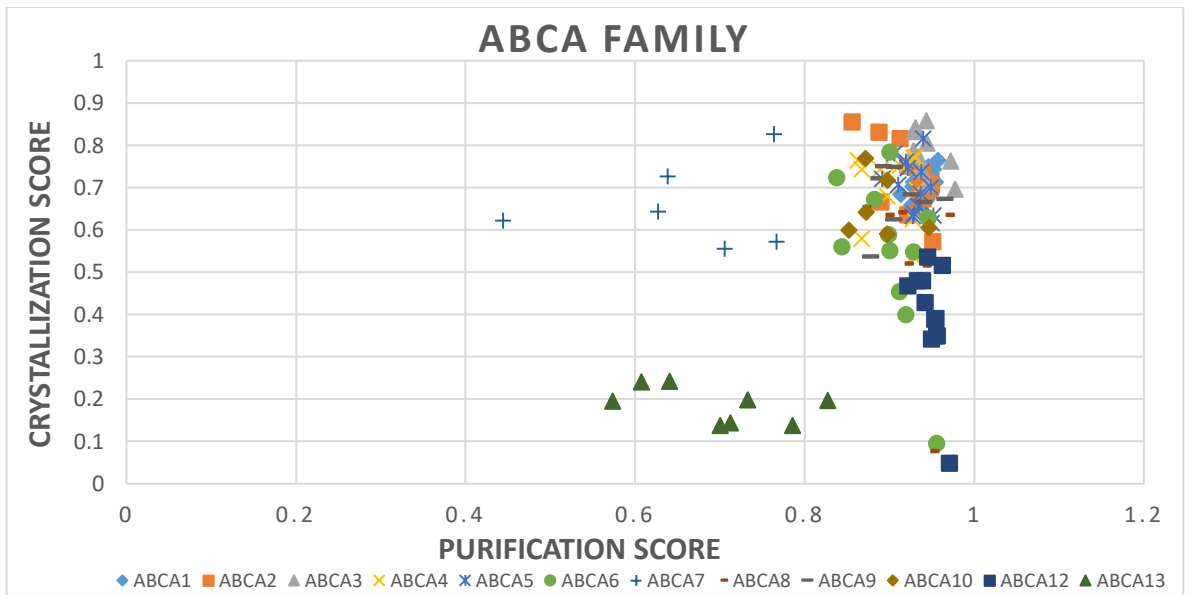
Figure 3.14: Correlation between crystallization and purification score of ABCA family
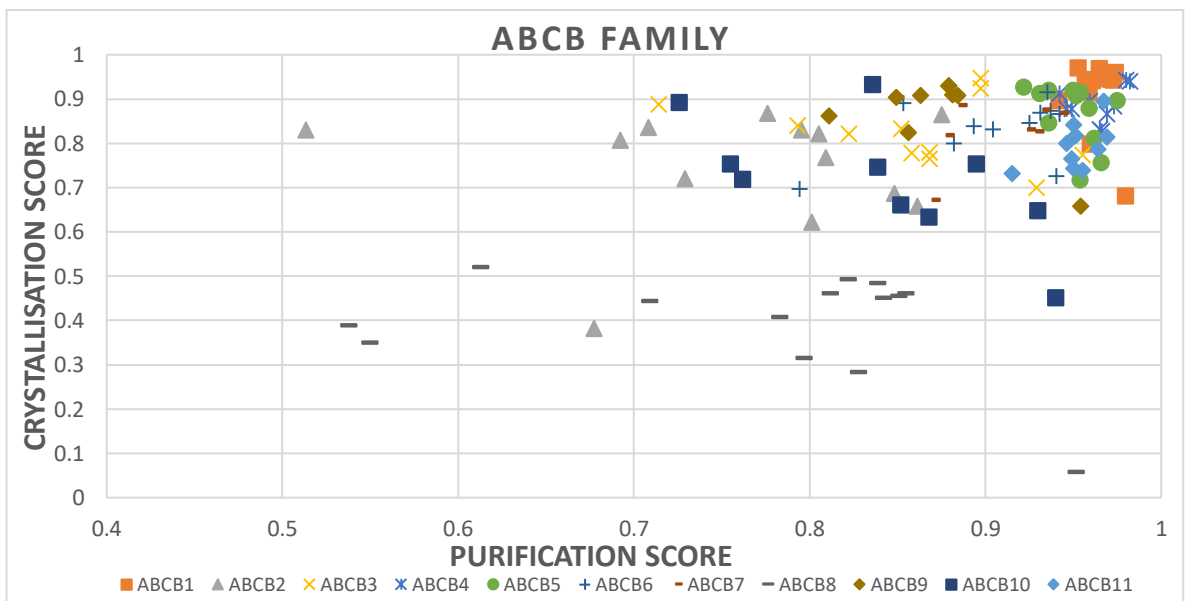
Figure 3.15: Correlation between crystallization and purification score of ABCB family
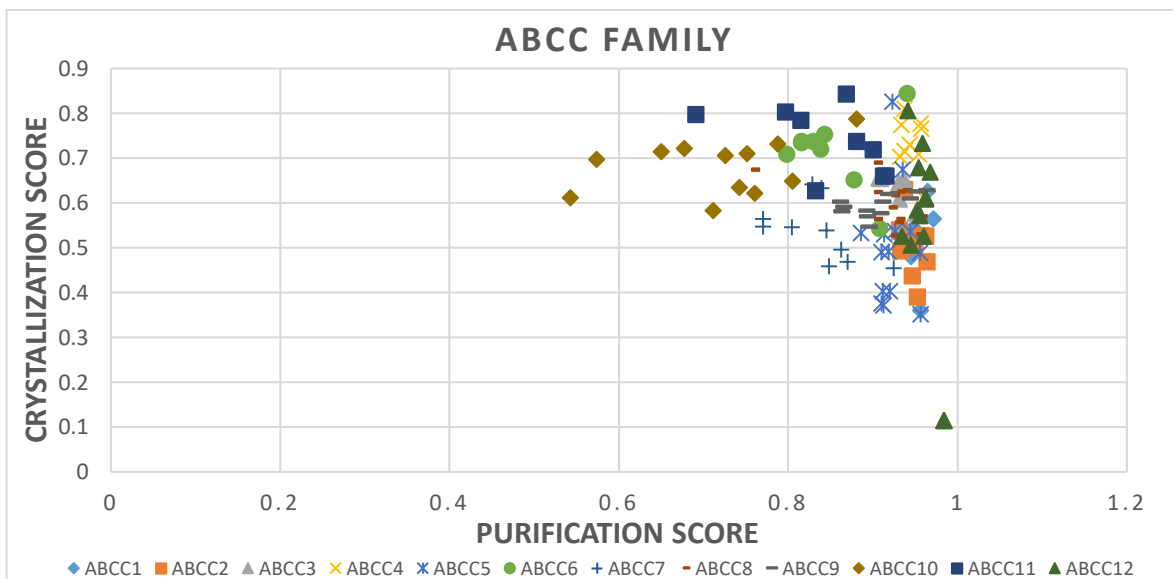
Figure 3.16: Correlation between crystallization and purification score of ABCC family
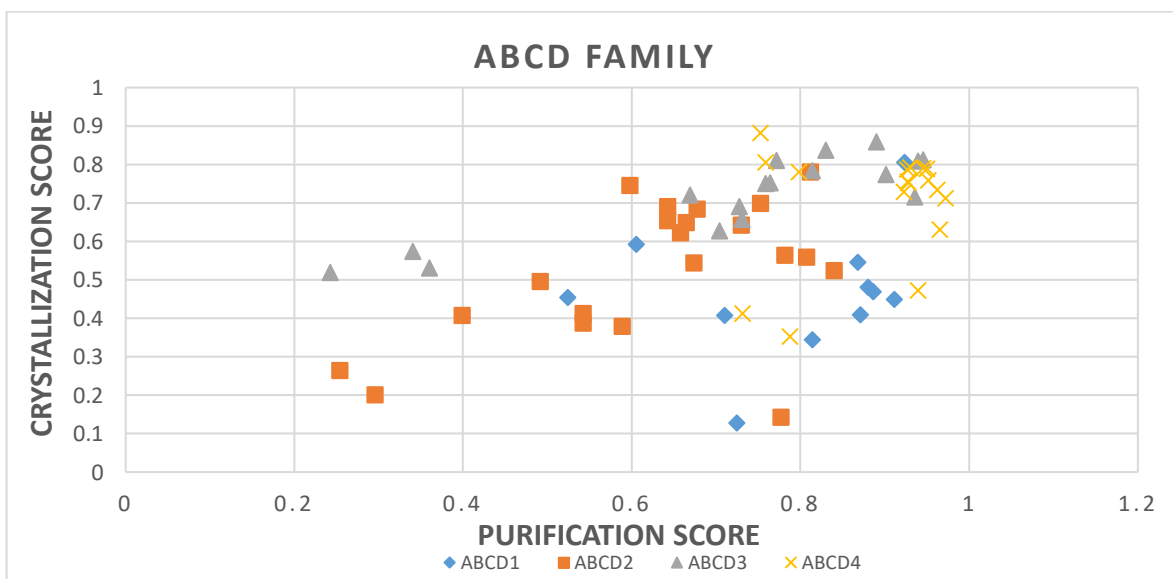
Figure 3.17: Correlation between crystallization and purification score of ABCD family
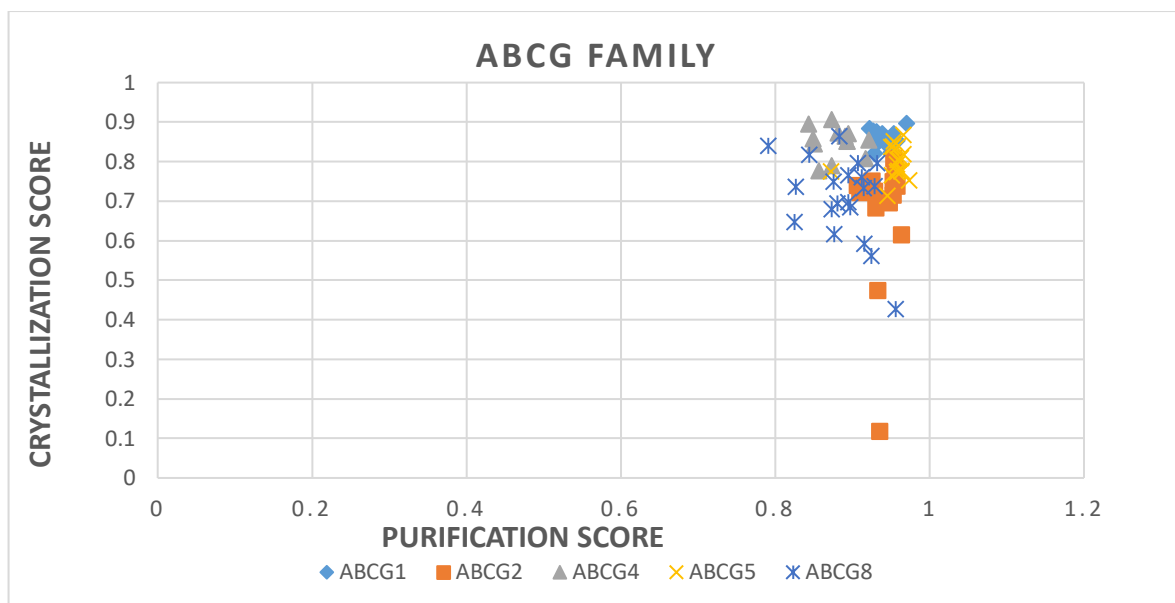
Figure 3.18: Correlation between crystallization and purification score of ABCG family

3.8 Assignment of protein targets for structural studies

The following proteins were chosen for structural research:

1. Mouse ATP binding cassette subfamily B member 5 (mABCB5)
2. Giant Panda ATP binding cassette subfamily B member 6 (gpABCB6)
3. Little brown bat ATP binding cassette subfamily G member 1 (bbABCG1)
4. Mouse ATP binding cassette subfamily G member 4 (mABCG4)

The reasons for selection of above-mentioned target proteins are as follows:

- Human versions of ABCB5, ABCB6, ABCG1, and ABCG4 is quite challenging to crystallize but the chances of crystallization of mABCB5, gpABCB6, bbABCG1, and mABCG4 are predicted to be very high. TMCrys score for mABCB5, gpABCB6, bbABCG1, and mABCG4 are as 0.6, 0.9, 0.6, and 0.6, respectively.

- Sequence similarity of selected orthologs *i.e.* mABCB5, gpABCB6, bbABCG1, and mABCG4 to human protein version are as 78%, 87.6%, 87.8%, and 96.6%, respectively checked by NCBI pairwise alignment.

- **No structural information:** To date, there is no high-resolution X-ray crystal structure of full-length ABCB5, ABCB6, ABCG1, and ABCG4. Only structures of the isolated nucleotide binding domain (NBD) of ABCB6 have been solved, with X-

ray crystallography in the apo- form, ADP bound, ADP+Mg^{2+} and in complex with ATP so far.

- **Clinical importance:** As described in chapter 1 ABCB5, ABCB6, ABCG1, and ABCG4 has clinical significance. A number of studies revealed ABCB5 is linked with different types of cancer and borna disease. Overexpression/upregulation of ABCB6 is linked to multiple chemotherapeutic resistance. Mutations in ABCB6 also cause of several diseases namely; porphyria, ocular coloboma and dyschromatosis universalis hereditaria. ABCG1 is associated with tangier disease and sitosterolemia while ABCG4 plays role in alzheimer disease.

Information of the 3D structure of human ABCB5, ABCB6, ABCG1, and ABCG4 protein is needed to understand its functions at a molecular level. If successful in determining the structure of the chosen orthologs, homology modelling will be used to build an atomic model of human ABCB5, human ABCB6, human ABCG1, and human ABCG4 using the structure of the chosen orthologs as a template.

3.9 Homology Modeling using MODELLER

Homology modeling is also called Comparative modeling. It is the process of constructing an atomic-resolution model of a protein from its amino acid sequence and an experimentally determined 3-dimensional (3D) structure of a relevant homologous protein (template) (Krieger et al., 2003). It is based on the concept that evolutionary related proteins have similar structures (Rost, 1997). If the sequence similarity of a protein having an unknown structure is high enough to a protein of known structure, then its three-dimensional structure can be built (Rost, 1997). Even if the quality of these models cannot compete with the experimentally determined structures, they are extremely cheap to produce and can be applied on a larger scale. The accuracy of homology model depends upon two main factors: 1) Choose the best possible template/templates 2) Optimally align the target sequence onto the template. Optimal template selection is key to generate a reliable model. Template selection can be done by performing a PSI-Blast search against the PDB database (Altschul et al., 1997). When performing a PSI-Blast search, a reliable approach is to identify hits with low E-value (E-value represents the expectation of finding that sequence by random chance) which are considered sufficiently close in evolution to make a reliable homology model. A template with a poor E-value should generally not be chosen, even if it is the only one available, since it may well have a wrong structure, leading to the production of a misguided

model (Webb and Sali, 2016). Beside E-value, a template having a similar function to the target protein can also be used to generate a model. Overview of the homology modeling is given in figure 3.19.

For this study, Modeller 9.24 is used (see section 2.8.1.2) for predicting homology models of all target proteins (ŠAli and Overington, 1994). It is one of the first fully automated command-line-based programs, needs some basic python scripting, relatively fast, and produces reliable results (Kuntal et al., 2010).
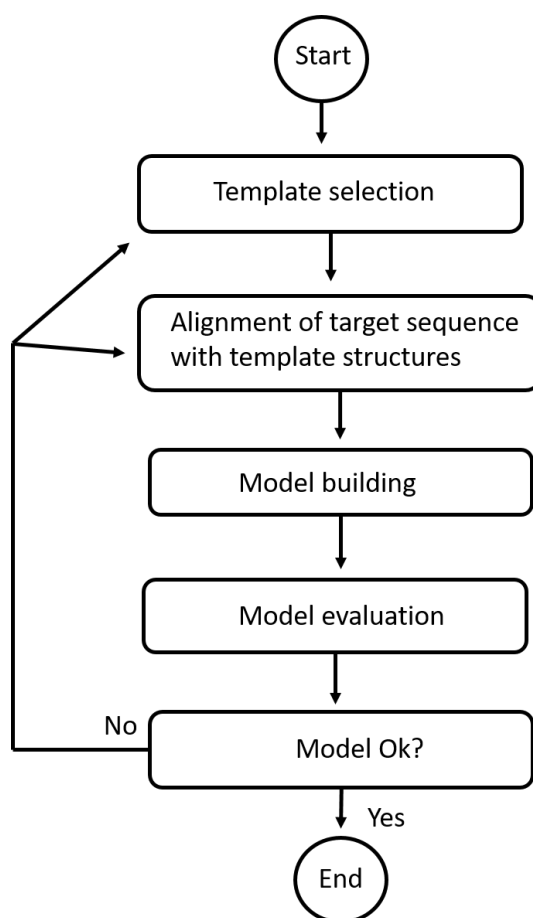
Figure 3.19: Diagrammatic explanation of homology modeling

3.9.1 Homology models of all Target Proteins
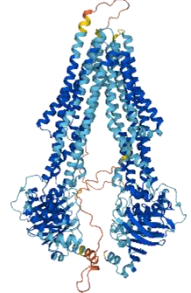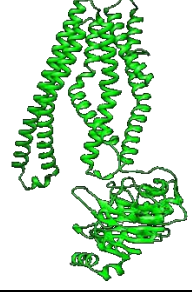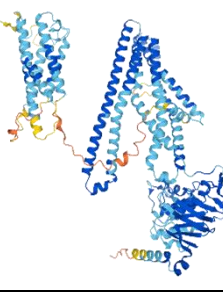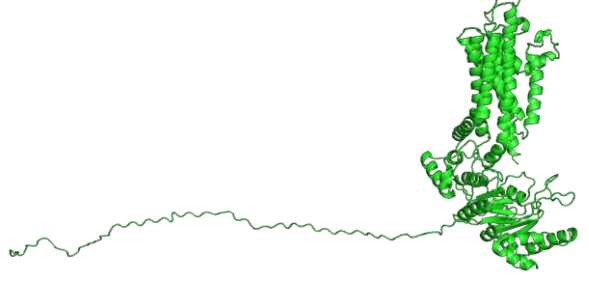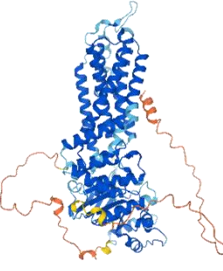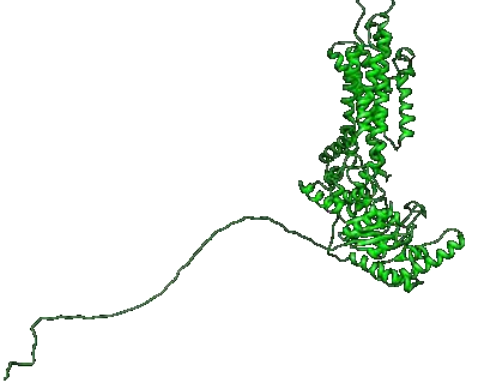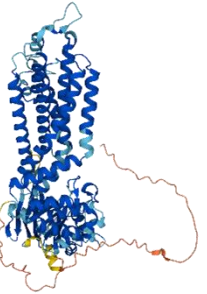
Table 3.6 lists the transmembrane and nucleotide-binding domains of all target proteins, while table 3.7 lists the protein structures. The Alphafold2 structure from the Alphafold2 database is provided for comparison (Jumper et al., 2020). ABCB5 is a full-length transporter (2xTMD and 2x NBD) of the ATP binding cassette protein B family. The first 33 amino acids of mABCB5 remained unstructured in our predicted structure by

MODELLER 9.24, which is also in consistent with Alphafold's predicted structure in which initial 35 amino acids are not solved. Additionally, in our MODELLER predicted mABCB5 structure, transmembrane 2 (TM2) helices have a slight kink as compared to the Alphafold structure. ATP binding cassette protein B member 6 (ABCB6) is a half transporter (1xTMD and 1xNBD, and an additional TMD0 domain). We were unable to get the TMD0 domain of the ABCB6 structure by MODELLER, on the other hand, TMD0 can be seen in the Alphafold structure. This extra TMD0 domain is absent in the template structures that were used for ABCB6 modelling, which would explain why it does not appear in the ABCB6 structure predicted by MODELLER. ATP binding cassette protein G1 (ABCG1) and ATP binding cassette protein G4 (ABCG4) both are half transporters. The first 76 amino acids of ABCG1; while the initial 60 amino acids of ABCG4 have an ambiguous structure. This unstructured region might form an additional domain in ABCG1 and ABCG4 like TMD0. Templates used for homology modelling of ABCG1 and ABCG4 did not contain TMD0, which could justify the initial unstructured amino acids. Root mean square deviation (RMSD) between the recently published structure of ABCB6 on PDB (PDB ID: 7DNY) (Kim et al., 2022) and ABCB6 structure predicted by MODELLER is 1.265Å. RMSD between recently published ABCG1 structure (PDB ID: 7R8C) (Sun et al., 2021) and MODELLER determined structure is 1.361Å.

Table 3.6: Predicted Transmembrane domains (TMDs) and Nucleotide-binding domains of all target proteins are shown.

Protein	Transmembrane domains (TMDs)	Nucleotide-binding domains (NBDs)
mABCB5	46-66 104-124 290-310 314-334 694-714 738-758 814-836 841-863 955-975	422-429 1051-1058
gpABCB6	24-47 59-83 103-124 145-165 185-204 261-284 296-313 381-403 409-430 501-519 525-545	621-628
bbABCG1	428-445 457-478 498-523 535-559 565-587 599-617 649-671	118-125
mABCG4	394-414 426-446 473-493 504-524 533-553 618-638	102-109

Table 3.7: Protein structures determined by MODELLER 9.24 and AlphaFold. RMSD between MODELLER and AlphaFold structures is determined in Chimera. In AlphaFold structures, structure confidence is indicated by different colours as **Dark blue:** >90%, **Light blue:** 70-90%, **yellow:** 50-70%, and **orange:** <50%.

Protein	MODELLER Structure	AlphaFold structure	RMSD (Å)
mABCB5			1.159
gpABCB6			1.018
bbABCG1			1.134
mABCG4			1.160

3.10 Discussion and Conclusion

The primary method for determining the structures of proteins is X-ray crystallography. However, the underlying crystallisation process is expensive, lengthy, and time-consuming as this involves a number of trial and error experiments to make a good diffracting crystal (Wang et al., 2014, Wang et al., 2016, Mizianty and Kurgan, 2011). The success rate of crystallisation experiments can be enhanced by choosing proteins that may crystallise easily. For this purpose, number of bioinformatics tools are designed for determining the probability of protein crystallisation based on the sequences of proteins (Wang et al., 2018). These methods make it easier to prioritise the target proteins that show the greatest promise, look for alternative structural orthologues of the target proteins, and recommend construct designs that would increase the possibility that crystallisation will be effective.

In this chapter, we selected an artificial intelligence AI tool, TMCrys and used it to score various orthologs of human ABC transporters, showing that certain orthologs were much higher scoring than the human versions (and hence better targets for crystallisation). When using the TMCrys server, it was found that the server was oversensitive to changes in a single amino acid at the first position; specifically, by changing the N terminal amino acid, the scores for crystallisation propensity changed. Having any of the three amino acids *i.e.* Glycine (G), Methionine (M), and Valine (V) at the N terminus drastically decreased the crystallization score from ~0.8 to ~0.2. This odd susceptibility of the software to the nature of the N-terminal residue in the amino acid sequence was discussed with the software creators and after some fruitful discussions, we were able to re-score protein targets with much greater confidence and at the same time helped to significantly improve the software package. The reason of oversensitivity of TMCrys server to initial amino acids was found that the TMCrys relies heavily on topology prediction and topology prediction depends on initial amino acids. TMCrys server uses CCTOP to predict the topology of these membrane proteins which incorporates 10 methods and gives a consensus prediction. It turned out that two of these methods (Pro and Prodiv) changes their prediction to give 15 TM helices instead of 17 and these are changing the resulting topology predictions which is extremely unlikely to happen in the living cells. To overcome this computational error completely, TMCrys developers have since rewritten the webserver logic so from now it does not give different results upon changing the first amino acid.

The crystallisation propensity scores for the three subfamilies ABCA12, ABCA13, and ABCB8 were found to be quite low when compared to orthologs of other subfamilies (see

section 3.6). Purification and crystallisation scores were directly correlated, with the best-behaving family being ABCG and the worst-behaving family being ABCD (see section 3.7). The structures of the target proteins as identified by MODELLER 9.24 are similar to those predicted by the most recent online AlphaFold server with slight differences (see table 3.7) (Cramer, 2021).

Conclusion

Bioinformatics tool, TMCrys was used to score various orthologs of human ABC transporters, showing that certain orthologs were much higher scoring than the human versions and hence better targets for crystallisation. The following four protein sequences were chosen for high resolution structural investigation based on the bioinformatics scoring results and their >70% sequence similarity to human protein version:

1. Mouse ABCB5 (mABCB5)
2. giant panda ABCB6 (gpABCB6)
3. little brown bat BACG1 (bbABCG1)
4. mouse ABCG4 (mABCG4)

Chapter 4 Results and Discussion: Expression and Evaluation

The goal of this project was to obtain a structural representation of ATP binding cassette protein subfamily B, member 5 (ABCB5), ATP binding cassette protein subfamily B, member 6 (ABCB6), ATP binding cassette protein subfamily G, member 1 (ABCG1), and ATP binding cassette protein subfamily G, member 4 (ABCG4). In this chapter, expression constructs were designed for orthologs representative of ABCB5, ABCB6, ABCG1, and ABCG4 (see chapter 3), followed by the DNA sequences synthesis (with yeast codon optimisation) and then inserted the constructs into an *E. coli* system that allowed seamless incorporation into our yeast expression system. All constructs were expressed in *S. cerevisiae,* with a GFP fusion tag and octa-His tag (8His) (see chapter 2). GFP tag allowed to track the expression of all of the proteins in real-time during expression, solubilisation, and purification (Misteli and Spector, 1997, Wacker et al., 1997). The optimal expression time for yeast cell harvesting was determined in this chapter using a time course experiment. Abbreviations 'mABCB5' for mouse ABCB5, 'gpABCB6' for giant panda ABCB6, 'bbABCG1' for little brown bat ABCG1, and 'mABCG4' for mouse ABCG4 will be used throughout the thesis.

4.1 Generation and expression of all of the four constructs in *S. cerevisiae*

For this research work, p424GAL1 plasmid vector was used as a cloning vector as well as an expression vector. Target genes were expressed in the p424GAL1 vector under the GAL1 promoter. This vector has been successfully used for cloning and expression of membrane proteins (Drew et al., 2006). To keep GFP from dimerizing site-directed mutagenesis was used to introduce the A206K mutation into the GFP (von Stetten et al., 2012). Primer details can be accessed in Chapter 02 (see Table 2.2). A206K-mutated p424GAL1 vector was sent to Source BioScience and the mutated site was confirmed by sequencing using Universal primer M13R (see appendix Figure 4.1). The A206K mutant p424GAL1 vector, protein sequences (mABCB5, gpABCB6, bbABCG1, and mABCG4), and cloning strategy (section 2.7) were sent to ProteoGenix, a life sciences company in France, for construct synthesis. 5μg of each of these four constructs were received from Proteogenix. For each construct, 200 ng/μl of stock plasmid was prepared by adding TE buffer to the lyophilized powder. On 1% agarose gel electrophoresis, the presence of recombinant plasmid DNA was confirmed (Figure 4.1), and the gene sequence encoding each protein was also confirmed by sanger sequencing. All of the four constructs were successfully transformed into *E. coli* cells first

for plasmid maintenance *i.e.* to increase the plasmid copy number and into *S. cerevisiae* FGY217 cells for expression and subsequent purification of the protein.
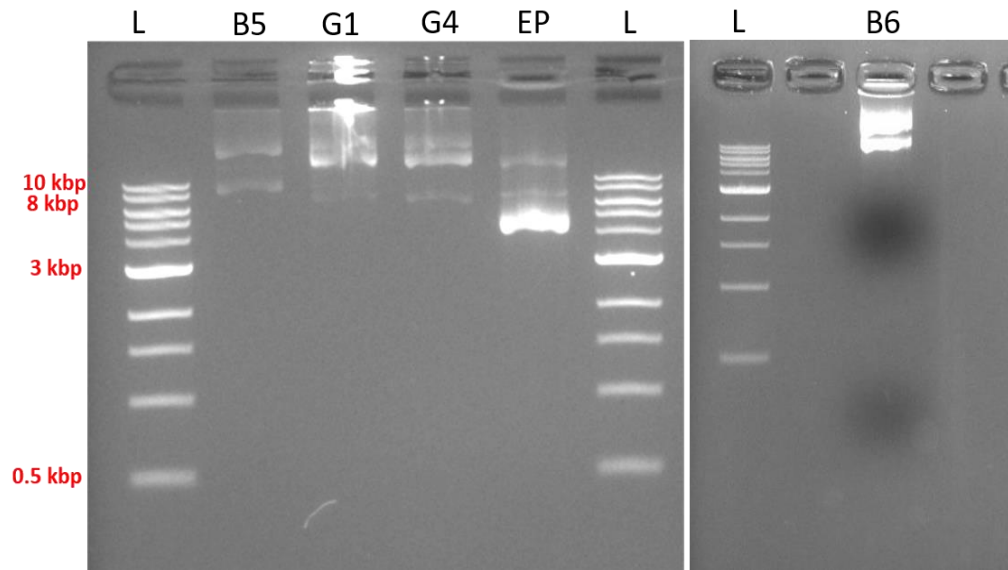
Figure 4.1: **Characterization of mABCB5, gpABCB6, bbABCG1 and mABCG4 on 1% agarose gel under UV-transilluminator.** **Lane 1:** 1 kbp DNA ladder, **Lane 2:** mABCB5, **Lane 3:** bbABCG1, **lane 4:** mABCG4, **lane 5:** Empty plasmid, **lane 9:** gpABCB6, **EP:** Empty plasmid, **L:** DNA ladder. 9 kbp corresponds to mABCB5, while 7 kbp corresponds to bbABCG1, mABCG1, and gpABCB6.

4.2 Analyses of screening of mABCB5, gpABCB6, bbABCG1 and mABCG4 expressing yeast colonies

mABCB5, gpABCB6, bbABCG1, and mABCG4 constructs were successfully transformed into highly competent DH5α cells, as well as, into the yeast cell (see section 2.11). Small scale screening of transformed yeast cells was performed to confirm the protein expression in yeast, as well as, to select the highest-protein expressing yeast colony for large scale expression and protein purification. In order to conduct small-scale screening for each construct, seven to twenty single yeast colonies were picked from the uracil selection plate and grown separately. For each colony microsomes were prepared and combined with 2X SDS loading buffer before being run on SDS-PAGE. SDS-PAGE gel was scanned under Alexa 488, Cyc 5 and Cyc 3 channel settings for GFP detection (as described in section 2.3) and western blotting was carried out using HRP-conjugated anti-His antibody (as described in section 2.3). For a small-scale screen of mABCB5 expression, 7 yeast colonies picked from a transformation experiment. Results of screening of 7 mABCB5 colonies are shown in panel A of figure 4.2. The mABCB5-GFP fusion protein, which generally migrates between the 250 kDa and 150 kDa, is strongly expressed in colony 3, as shown in the figure

4.2. For gpABCB6, 15 out of 20 colonies showed expression between 130 kDa and 95 kDa; the results are depicted in Panel B of figure 4.2. Ten yeast colonies were screened for bbABCG1 and mABCG4, and the findings are displayed in Panels A and B of Fig. 4.3, respectively. For mABCG4, 7 out of 10 colonies showed the expression, compared to 8 out of 10 colonies for bbABCG1. For protein purification on a large scale, mABCB5-colony 2, gpABCB6-colony 8, bbABCG1 colony 8, and mABCG4 colony 2 was used. Expression yield for each transformed constructs is calculated and shown in table 4.1.

A.
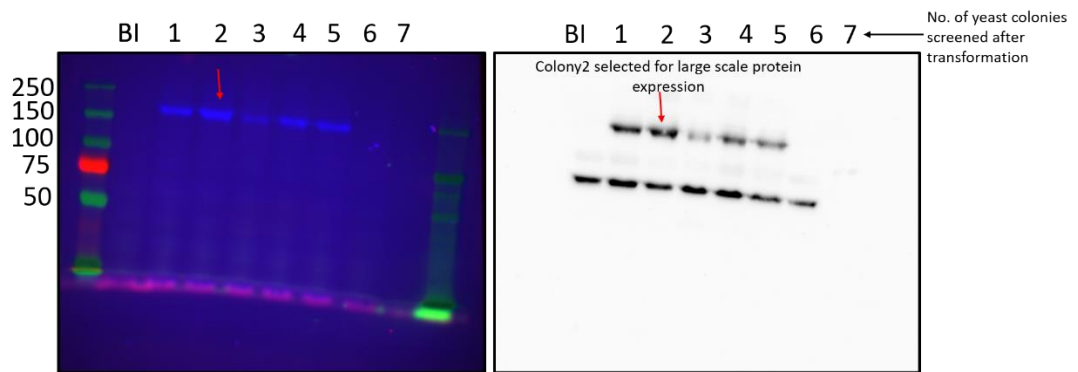
B.
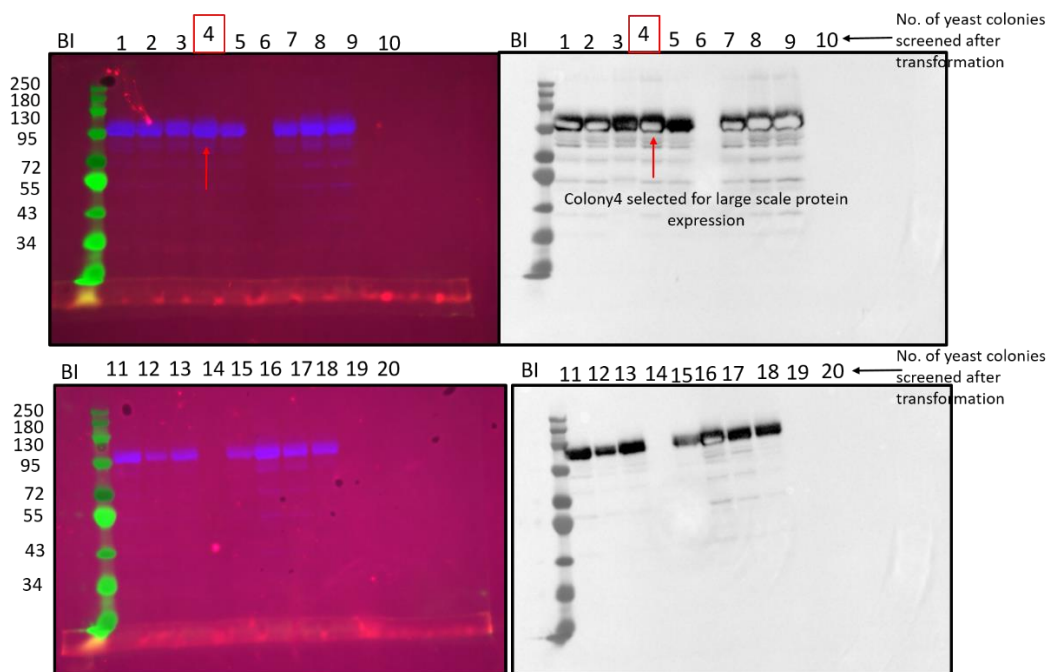
Figure 4.2: Screening of mABCB5 and gpABCB6 transformed yeast colonies. The mABCB5-expressing yeast colonies are shown in **Panel A**. The gpABCB6-expressing yeast colonies are shown in **Panel B.** The western blot results are displayed in the right panel. The SDS-gel scanned for GFP fluorescence is shown in the left panel. For each construct, the colony selected for large-scale expression is indicated by a red arrow. **M:** protein marker in kDa. **BI:** yeast cells before induction. The colonies screened for each construct are represented by numbers.

A.

B.
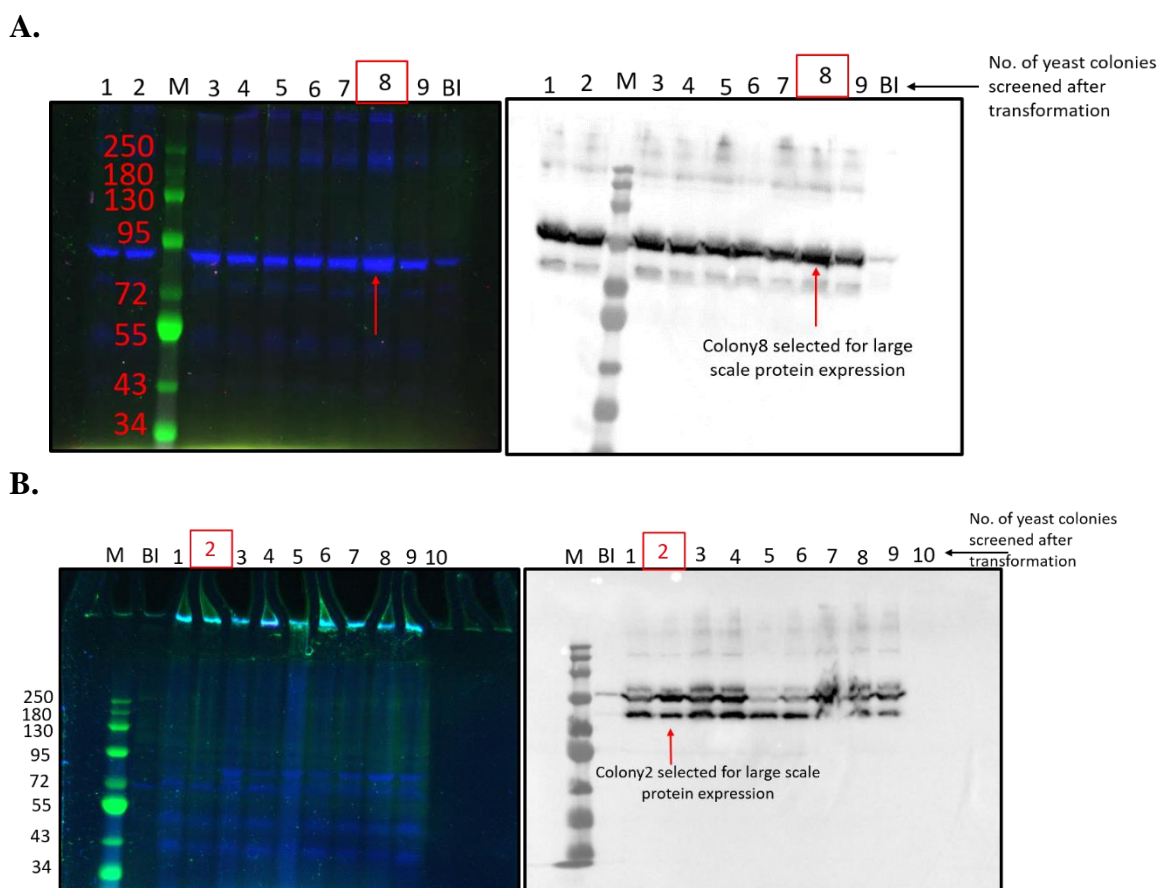
Figure 4.3: Screening of bbABCG1 and mABCG4 transformed yeast colonies. The bbABCG1-expressing yeast colonies are shown in **Panel A.** The mABCG4-expressing yeast colonies are shown in **Panel B.** The western blot results are displayed in the right panel. The SDS-gel is scanned for GFP fluorescence in the left panel. For each construct, the colony selected for large-scale expression is indicated by a red arrow. **M:** protein marker in kDa. The colonies screened for each construct are represented by numbers.

Table 4.1: Represents the expression yield of each transformed construct (Expression yield= No. of yeast colonies showed protein expression/ Total no. of screened colonies X 100)

Construct	Expression yield (%)
mABCB5	71.4
gpABCB6	75
bbABCG1	90
mABCG4	70

4.3 Time Course Analyses of mABCB5, gpABCB6, bbABCG1 and mABCG4 expression

A time course analysis was done to determine the level of protein expression following induction in order to obtain a high yield of mABCB5, gpABCB6, bbABCG1, and mABCG4. In a 2L baffled flask, yeast was grown in YNB media containing 0.2 % glucose (w/v). The media was left at 30 °C overnight while being shaken at 250 rpm. Cells were induced with 2 % galactose (w/v) and 8 % glycerol (v/v) as explained in the methodology section 2.14. 50 ml of yeast culture was collected at each time point, harvested at 3500x g and the yeast cell pellet was re-suspended in the lysis buffer with a protease inhibitor cocktail. A detail of the lysis buffer is given in section 2.4. The harvested yeast cells were fixed on a glass slide using 50 % glycerol (v/v) and analysed with a Zeiss Fluorescence microscope with a GFP filter (Alexa Fluor 488) alone, as well as, using a white light and GFP filter, both channels at the same time to get a better contrast.

Microscopy results for all constructs showed a negligible GFP signal in the first eight hours after galactose induction. However, after eight hours of galactose induction, GFP fluorescence started increasing. The highest intensity of GFP fluorescence of mABCB5, gpABCB6, bbABCG1, and mABCG4 was seen at 14 hours, between 18-20 hours, between 18-22 hours, and 14-18 hours after induction, respectively, and after these time points GFP signal decreased (Figure 4.5-4.8). This decrease in GFP signal continues and after 38 hours and 62 hours of post-induction almost reached the starting levels (Figure 4.5-4.8). The graph of corrected total cell fluorescence (CTCF) was plotted against time and the results are displayed in Figure 4.9. Fluorescence microscopy can also be used to track the protein's location within the cell (Figure 4.4). For mABCB5 and gpABCB6 much of the fluorescence is found around the periphery of the cell as expected (Drew et al., 2008). bbABCG1 and mABCG4 display a punctate localization, either on or just inside the plasma membrane which could be due to protein recycling through a late Golgi/endosomal pathway (Yoo et al., 2002). Subcellular localisation of proteins needs to be confirmed using defined markers for specific organelles.

(mABCB5)   **(gpABCB6)**   **(bbABCG1)**   **(mABCG4)**
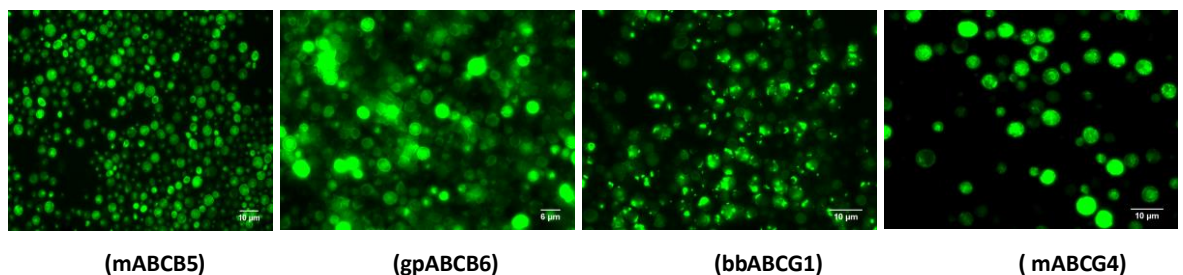
Figure 4.4: Fluorescence microscopy for detection of protein location
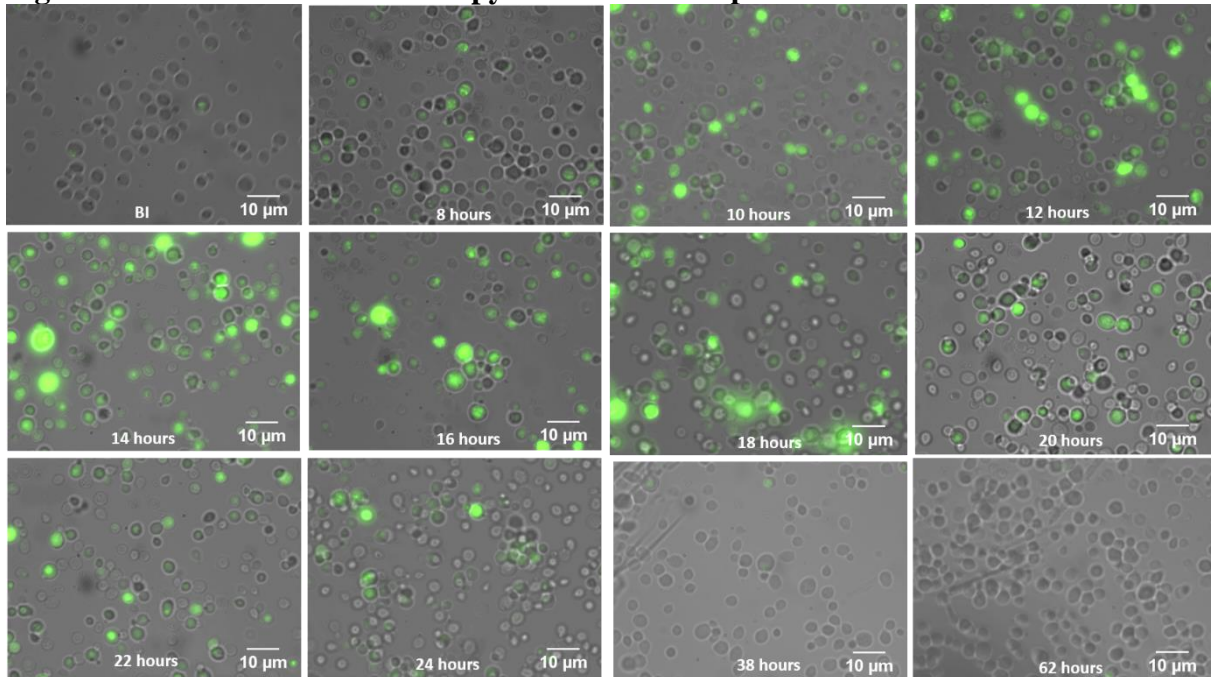
Figure 4.5: Fluorescence microscopy of yeast expressing mouse ATP binding cassette protein subfamily B member 5 (mABCB5) at different time points after induction using Alexa Fluor 488 and a white light together. The highest intensity of GFP fluorescence is seen at 14 hours after galactose induction.
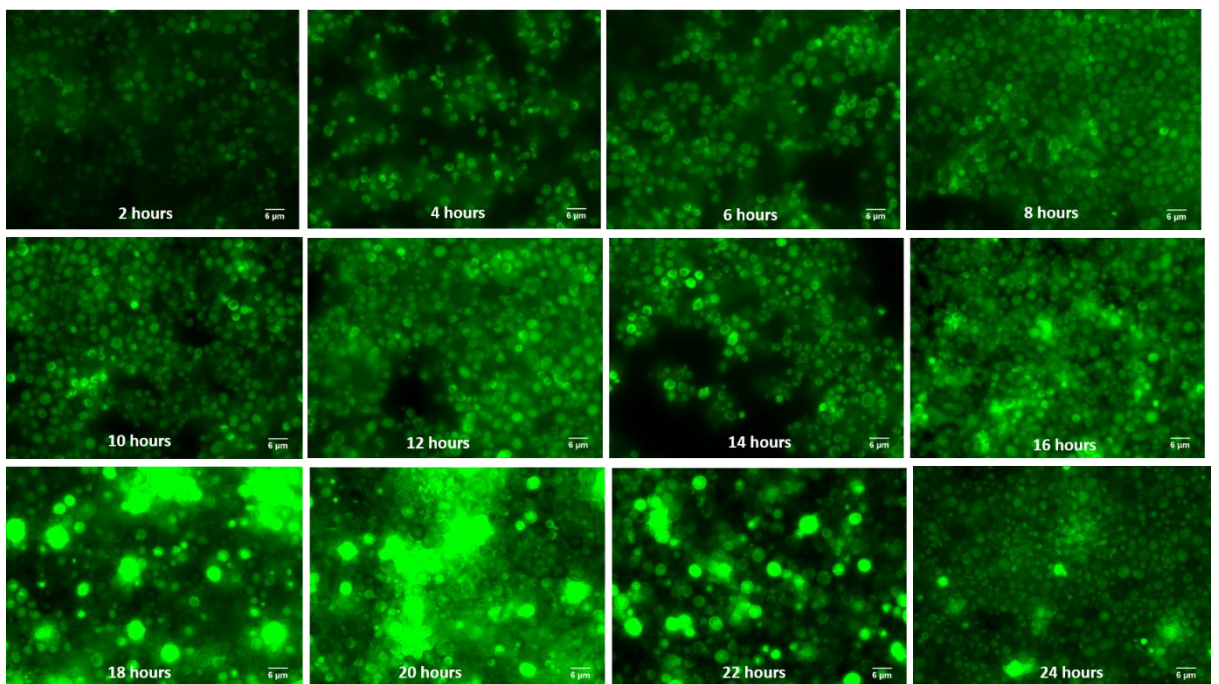
Figure 4.6: Fluorescence microscopy of yeast expressing giant panda ATP binding cassette protein subfamily B member 6 (gpABCB6) at different time points after using Alexa Fluor 488 filter. The highest intensity of GFP fluorescence is seen between 18-20 hours after galactose induction.
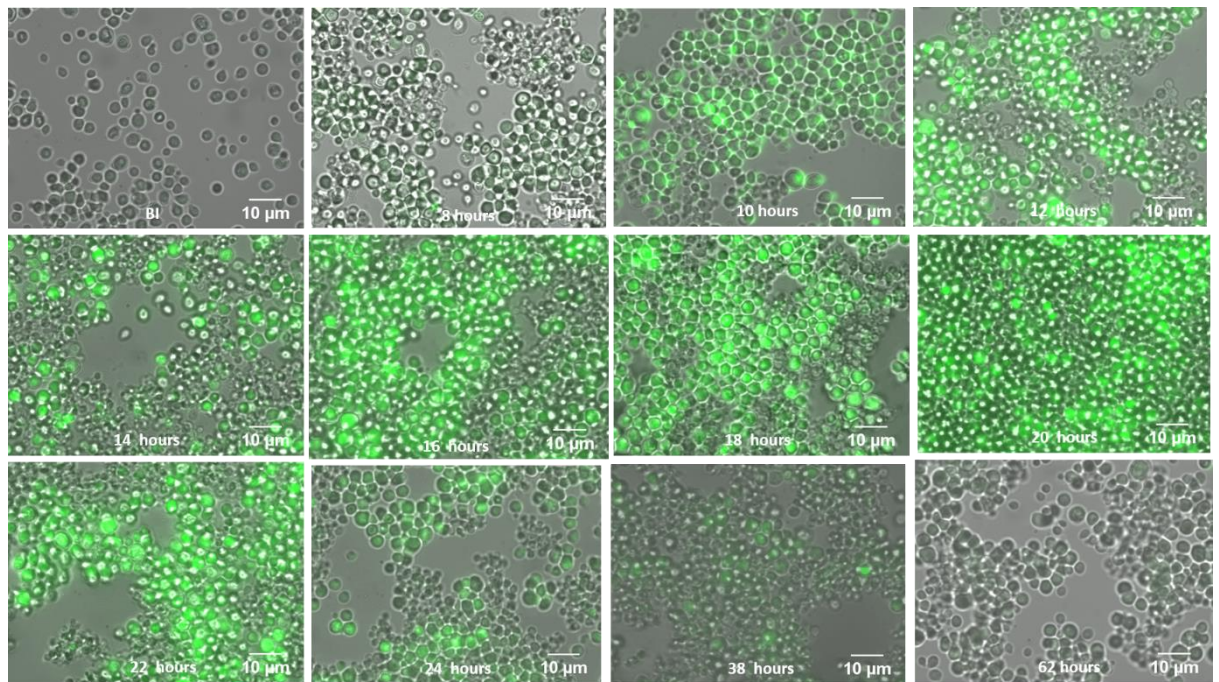
Figure 4.7: Fluorescence microscopy of yeast expressing little brown bat ATP binding cassette protein subfamily G member 1 (mABCG1) at different time points after induction using Alexa Fluor 488 and a white light together. The highest intensity of GFP fluorescence is seen between 18-22 hours after galactose induction.
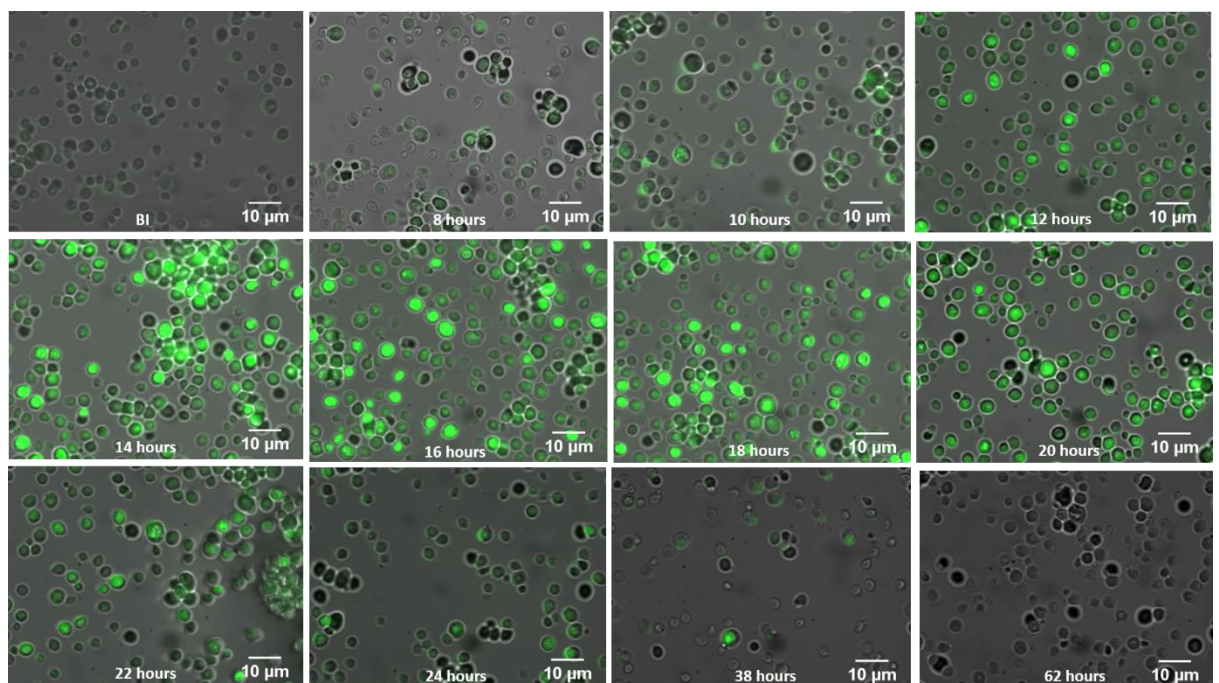
Figure 4.8: Fluorescence microscopy of yeast expressing mouse ATP binding cassette protein subfamily G member 4 (mABCG4) at different time points after induction using Alexa Fluor 488 and a white light together. The highest intensity of GFP fluorescence is seen between 14-18 hours after galactose induction.

(A)

ABCB5

(B)

ABCB6

(C)

ABCG1

(D)
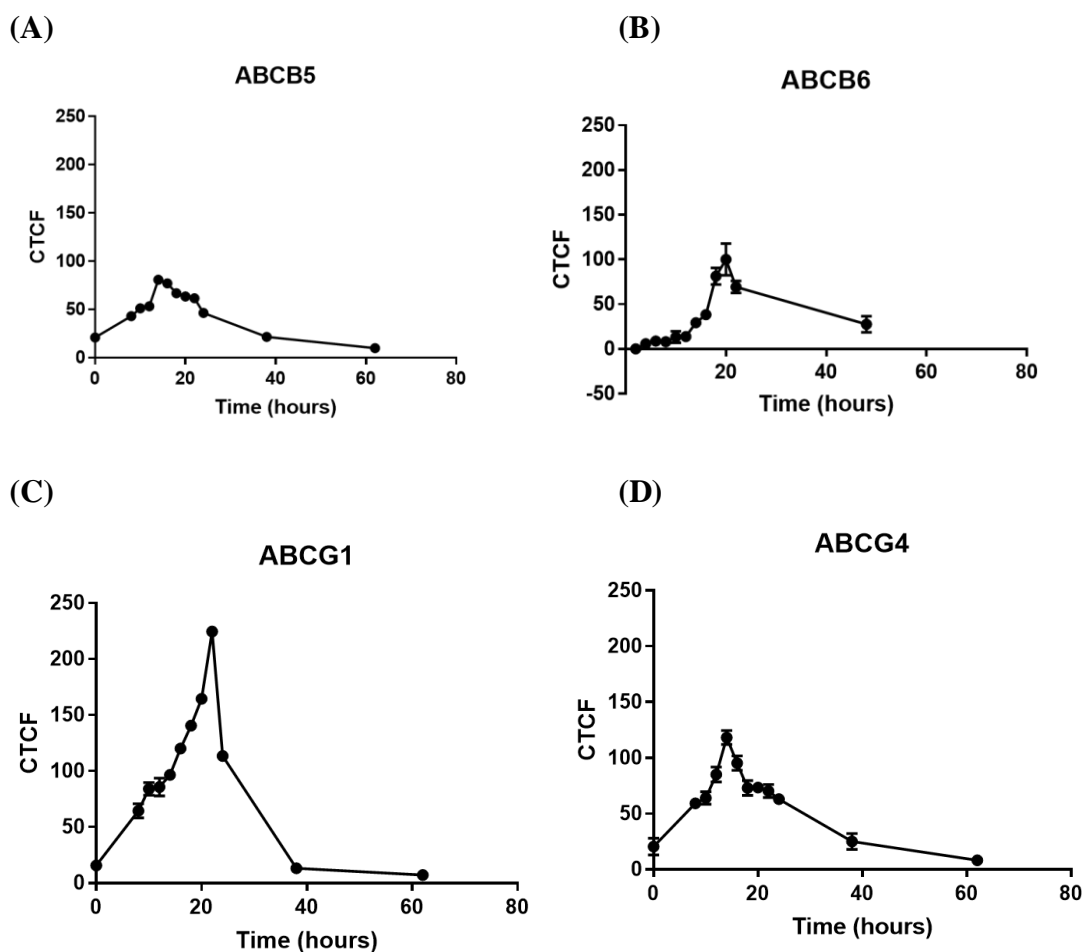
ABCG4

Figure 4.9: Time course experiment graphics for mABCB5, gpABCB6, bbABCG1, and mABCG4. The microscopy images were imported into Image J and calculated for corrected total cell fluorescence (CTCF) *i.e.* Total GFP signal from single cell after subtracting the background noise. mABCB5, gpABCB6, bbABCG1, and mABCG4 maximum expression occurs at 14 hours, 20 hours, 22 hours and 14 hours respectively.

mABCB5, bbABCG1, and mABCG4 expressing yeast cells were taken from different time points after galactose induction and were also processed for crude membranes (microsomes) preparation to monitor the highest protein expression levels by western blot. gpABCB6 western blot data is missing due to a lack of gpABCB6 expressing harvested yeast sample. For mABCB5, bbABCG1, and mABCG4, the highest protein expression was seen at 14 hours, 22 hours, and 14 hours (Figure 4.9).

A.
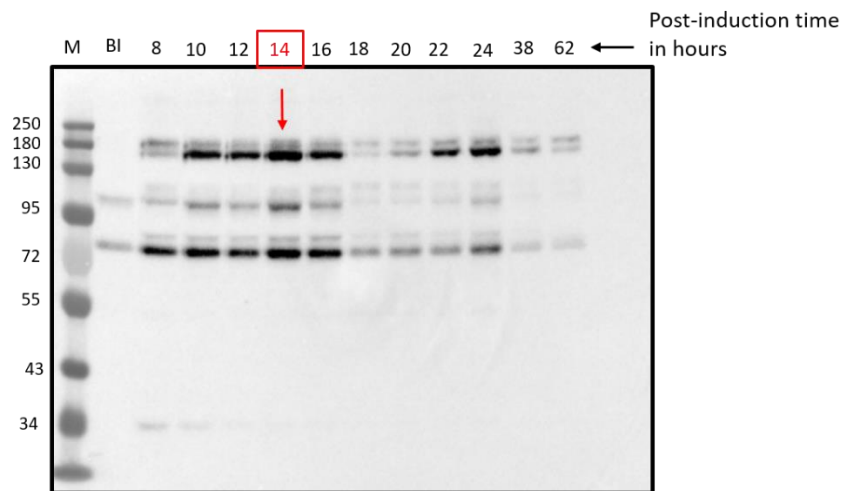
B.
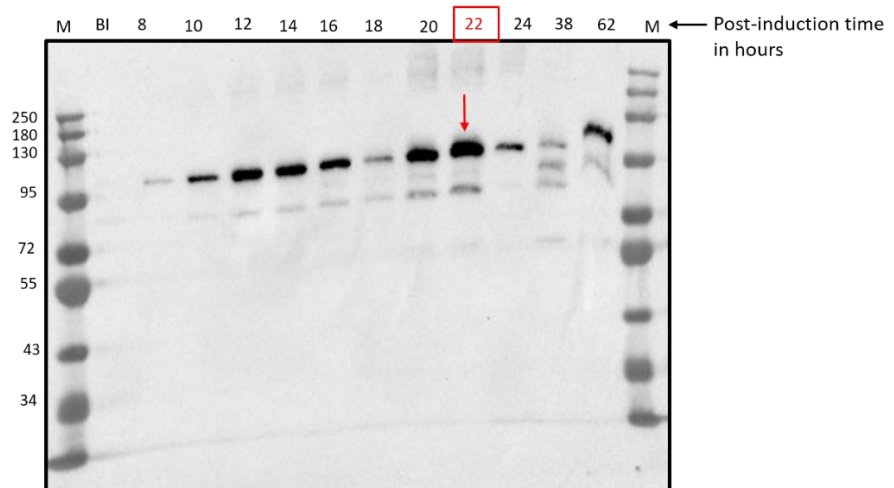
C.
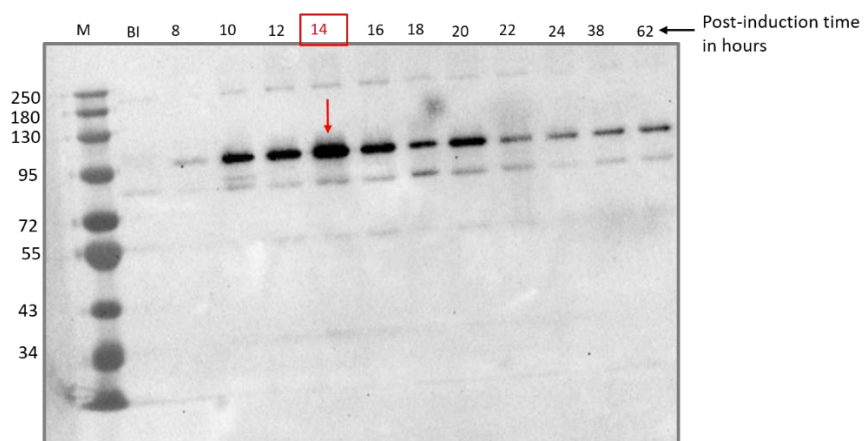
Figure 4.10: Western blot results of time course experiment for mABCB5, bbABCG1, mABCG4 after induction. Panel A shows the mABCB5 western blot. **Panel B** shows the bbABCG1 western blot. **Panel C** shows the mABCBG4 western blot. Red arrow indicates the time of harvesting selected for each construct. M: protein marker in kDa. **BI:** Before induction

4.4 Discussion and Conclusion

mABCB5, gpABCB6, bbABCG1, and mABCG4 constructs were successfully synthesized, cloned, and transformed into highly competent bacterial cells (DH5α) for backup as well as into yeast cells for protein expression and purification. Transformation of the yeast cells with the mABCB5/ gpABCB6/ bbABCG1/ mABCG4-containing vector, is straightforward, however, it is crucial to screen yeast colonies that express a lot of protein. Variable expression levels may result from a variety of reasons, but a large portion of variation is likely caused by the number of plasmid copies present in each cell.

GFP is ~27kDa and emits fluorescence at a wavelength of ~507-509 nm when excited at a wavelength of ~488 nm (Patterson et al., 1997). The use of a green fluorescent protein (GFP) tag enabled quick and simple detection of mABCB5, gpABCB6, bbABCG1, and mABCG4 in living yeast cells and also after extraction from the plasma membrane. It is clear from the microscopy data shown in Figures 4.5-4.8 that the timing of cell harvesting after induction with galactose is an important factor. GFP fluorescence in microscopy images of mABCB5, gpABCB6, bbABCG1, and mABCG4 drop drastically after about 14 hours, 20 hours, 22 hours, and 14 hours of induction respectively, so that there is hardly any detectable protein expression in yeast cells after 24 hours of induction. Western blot data of the time course experiment of mABCB5, gpABCB6, bbABCG1, and mABCG4 is also in agreement with the fluorescence data (see figure 4.10). There could be two possible reasons for decreased protein expression after 24 hours of induction either protein degradation or the likelihood that there is abundant protein expression but GFP has been cleaved off the target protein.

Before induction, yeast cells also showed very less but evenly dispersed internal signals under the blue light filter (Alexa Fluor 488) (Figure 4.5-4.8, pre-induction). This is likely because the yeast cells produce inherently fluorescent proteins such as succinate dehydrogenase and other FAD-containing proteins (Frey et al., 2011). Leaky expression or incomplete promoter repression could also be a reason for the expression in the non-induced cells. Additionally, this time course microscopy images also confirm the correct folding of proteins in the yeast system. Yeast cells possess an intrinsic endoplasmic reticulum-associated degradation (ERAD) pathway for the quality monitoring of membrane proteins. Only properly folded proteins can reach the plasma membrane (PM), whereas misfolded proteins are entrapped and destroyed in the cytosol as the protein destruction part of ERAD occurs in the cytosol (Brodsky and Skach, 2011). Consequently, the production of the GFP

tag signal may serve as a proximate marker for the folding of heterologous-produced membrane proteins (Drew et al., 2001, Newstead et al., 2007).

Taken together, the results of this chapter were important in three aspects:

- To confirm protein expression in yeast
- To determine the time for yeast-cells harvesting to get a high yield of purified protein
- To ensure that microsomes can be made

Conclusion

Yeast expression was induced consistently when OD_{600} of ~1.6-1.8 was reached and glucose was completely used. Optimal yeast harvesting time following galactose induction was determined and this was found within the 24 hours but some variations were observed as 14 hours, 20 hours, 22, and 14 hours, for mABCB5, gpABCB6, bbABCG1, and mABCG4 respectively.

Chapter 05 Results and Discussion: Protein Purification

The downstream structural activities required pure protein and the purification methodology was enabled by the specific design of the expression construct (see section 2.7). The expression of mABCB5, gpABCB6, bbABCG1, and mABCG4 protein in yeast was confirmed by fluorescence microscopy, as well as, optimal yeast cell harvesting times after galactose induction for each protein were also determined as described in chapter 4. In this chapter, we scaled up the yeast culture for protein purification, using a 12L shake-flask-batch method (see chapter 02, section 2.15). Large-scale protein purification was carried out using specially chosen yeast colonies for each design (see section 4.2). Protein expression was induced by the switch of yeast nutrients from glucose to galactose. Once glucose was depleted, 2 % galactose (w/v) and 8 % glycerol (v/v) were used to induce expression. 8 % glycerol (v/v) was introduced as a chemical chaperon that has been reported to significantly enhance membrane protein expression in yeast (Figler et al., 2000). The cells were harvested between 14-22 hours after induction as determined in chapter 4. The harvested cells were re-suspended in lysis buffer supplemented with a protease cocktail to inhibit yeast proteases. The microsomes (crude membranes) were prepared using a bead-beating protocol as explained in chapter 2 (see section 2.16) and dissolved in solubilisation buffer with detergent for protein purification.

5.1 Microsomes solubilisation

DDM is the most commonly used detergent to solubilise membrane proteins and it has been the most successful for structural studies of membrane proteins (Awayn et al., 2005, Anandan and Vrielink, 2016, Parker and Newstead, 2012). The efficiency of microsomes solubilisation in DDM can be influenced by salt concentration and pH value. The efficiency of gpABCB6 was evaluated with 2 % DDM (w/w) solubilisation buffer with pH ranging from 7-9 and different NaCl concentrations (as explained in chapter 02, section 2.16). In addition, the efficiency of solubilizing gpABCB6 microsomes in lysofos glycerol 14 was also assessed. The solubilized material was run on SDS-PAGE gels which were scanned for GFP fluorescence and also processed for Western blot.

Figure 5.1 shows the SDS-PAGE gels and Western blot with soluble and insoluble fractions for 7 different buffer conditions. Results revealed an optimal solubilisation efficiency with 50 mM Tris-buffer containing 2 % n-Dodecyl-β-D-Maltopyranoside (DDM) (w/w), no salt

at pH8 of ~50 % (as measured by the ratio of the gpABCB6 GFP band intensity for soluble and insoluble fractions). Solubilisation efficiency with Lysofos glycerol 14 was comparable with DDM (Figure 5.1), but due to the known harsh character of this detergent, DDM was preferred (Anandan and Vrielink, 2016).

A.
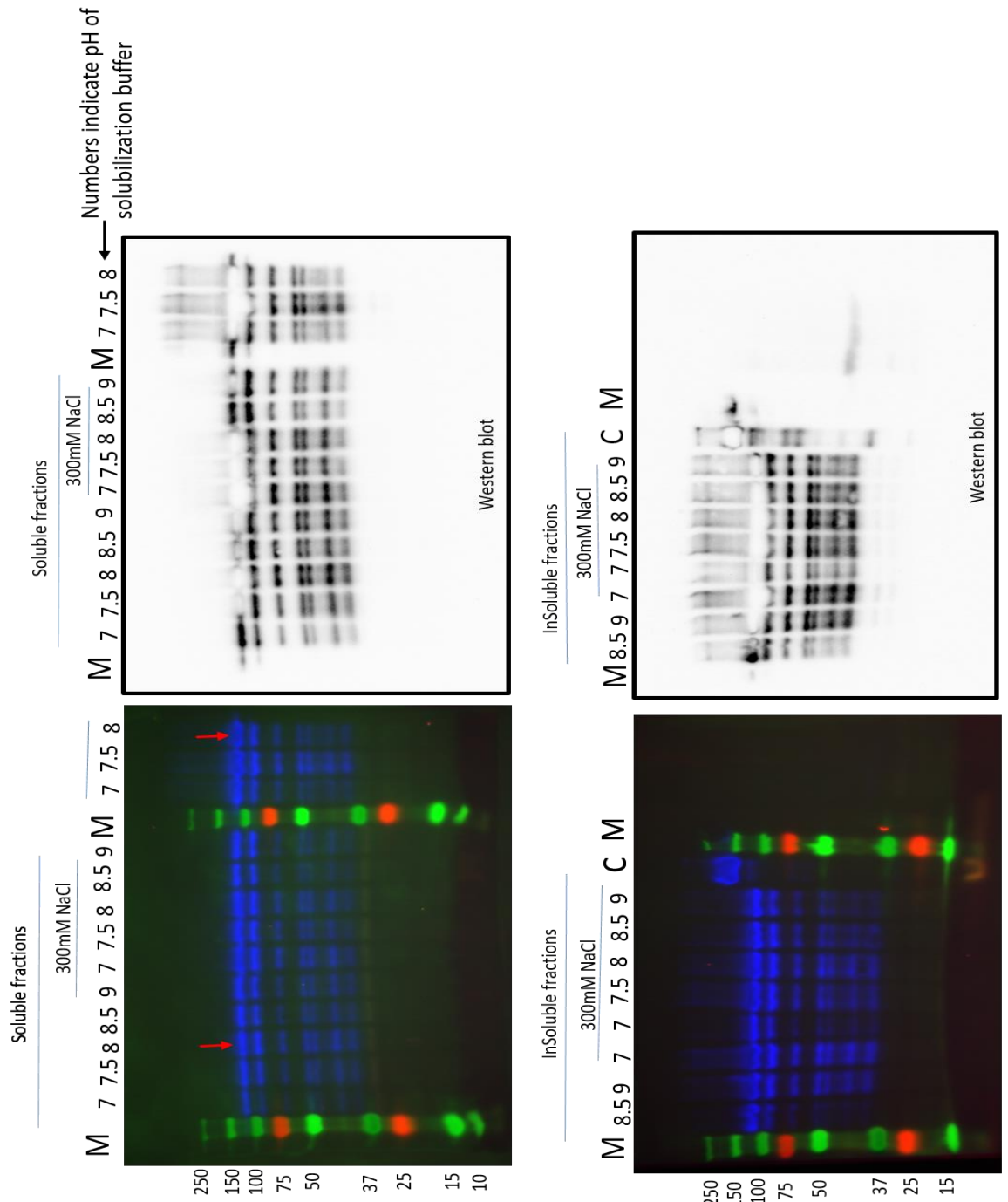
B.
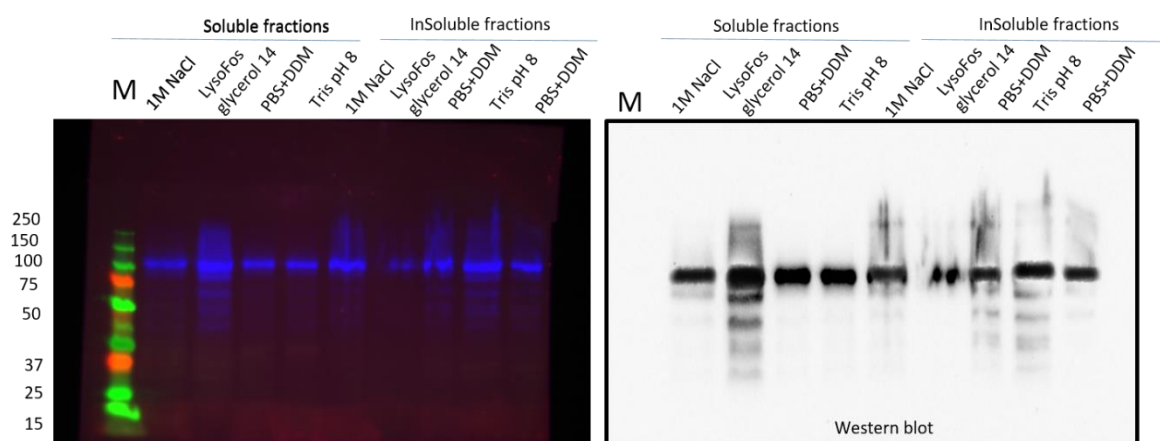
Figure 5.1: Optimization of gpABCB6 microsomes solubilisation with eight different buffer conditions. gpABCB6 containing microsomes at a total protein concentration of 5 mg/ml were incubated with 2 % DDM (w/w) buffer for 2 hours at 4 °C at a detergent to total protein ratio of (8:1). DDM-soluble protein was then separated from DDM insoluble material by ultracentrifugation at 120,000 g and analyzed by SDS-PAGE using GFP fluorescence (left panel) and by western blot using HRP-conjugated anti-His antibody (right panel). **Panel A, B.** 50 mM Tris-buffer containing 2% DDM, 300 mM NaCl at different pH (7, 7.5, 8, 8.5 and 9), 50 mM Tris buffer containing 2% DDM, no NaCl at different pH (7, 7.5, 8, 8.5 and 9); soluble and insoluble fractions. **Panel C.** Soluble and insoluble fractions of **Buffer 1:** 50 mM Tris, 1M NaCl (pH8) **Buffer 2:** 50 mM Tris, 2 % LysoFosglycerol 14, pH8 **Buffer 3:** 50 mM phosphate buffer saline (PBS), 2 % DDM, pH8 **Buffer 4:** 50 mM Tris, pH 8 **M:** Protein marker in kDa, **C:** P-gp as a positive control

For mABCB5, bbABCG1, and mABCG4, the microsomes solubilisation was standardized based on the optimization as described above and the work of Jackson et al 2018b (Jackson et al., 2018b). Microsomes were solubilized using 2 % DDM: 0.02 % CHS with final protein concentrations at 2.5 mg/ml and proceeded for protein purification (Jackson et al., 2018b).

5.2 Ni-NTA purification

The solubilized protein was purified using Ni-NTA chromatography followed by size exclusion (Pollock et al., 2014a). gpABCB6 was purified using Qiagen Ni-NTA loose resin (see chapter 2, section 2.17). The corresponding fractions were run on SDS-PAGE and analysed for GFP fluorescence. In Figure 5.2, a prominent band with a molecular weight of roughly 120 kDa is shown which can be compared to the expressed gpABCB6 construct molecular weight (sequence weight 92 kDa + the additional cleavable GFP and affinity tag = 21 kDa). Additionally, when comparing the quantity of gpABCB6 in the 150 mM fraction with the amount of gpABCB6 in the flow-through collected, Ni-NTA affinity chromatography demonstrates a protein retention efficiency of about 90 %. Imidazole-free

purification buffer was used to dilute the 150 mM elution fraction, bringing the 150 mM imidazole concentration down to less than 5 mM. Using milipore20 concentrators with a 100 kDa cut-off, the diluted fraction of approximately 100 ml was then concentrated to a final volume of 1-2 ml for size exclusion purification or until the absorption of GFP is observable by eye (a greenish yellow colour) in the concentrator.
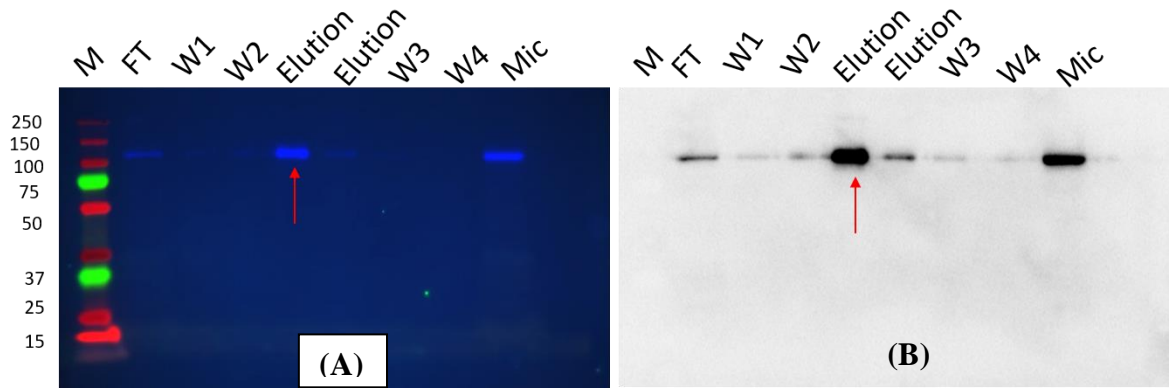
Figure 5.2: Ni-NTA chromatography of gpABCB6. Panel A shows the corresponding purification fractions on SDS gel scanned for GFP-fluorescence under Alexa488, Cyc3, and Cyc5 filters in the chemidoc system. gpABCB6 eluted at 150 mM imidazole concentration. Panel B shows the Western blot using HRP-conjugated anti-His Antibody **M:** The molecular weight markers are indicated on the left in kDa. **FT:** Flow-through, **Mic:** ABCB6 microsomes, **W1:** Wash 1 (20 mM imidazole concentration), **W2:** Wash 2 (40 mM imidazole concentration), **W3:** Wash 3 (500 mM imidazole concentration), **W4:** Wash 4 (500 mM imidazole concentration), **Elution:** 150 mM imidazole concentration. A band at ~120 KDa represents the gpABCB6 indicated by red arrow.

mABCB5, bbABCG1 and mABCG4 were purified using 5ml HiFliQ Ni-NTA FPLC column. All of these three solubilized proteins were purified by one optimized strategy *i.e.* first two washes with 20 mM and 80 mM imidazole to remove unbound protein and non-specific interactions, higher affinity protein elution between 200 mM-250 mM imidazole followed by a last wash with 500 mM imidazole to elute any remaining bound protein from the column. Figure 5.3A depicts mABCB5 fractions on SDS-PAGE scanned for GFP fluorescence; a prominent band with a molecular weight of roughly 150 kDa is shown, which represents the expressed mABCB5 construct molecular weight (sequence weight 137.4 kDa + the additional cleavable GFP and affinity tag = 21 kDa). A very faint band of mABCB5 and other protein contaminants was visible on an SDS gel stained with Coomassie (Figure 5.3B). Comparison of the band intensity of the 200 mM imidazole fraction to the band intensity of the flow through indicates protein retention of about 95 % (Figure 5.3A).
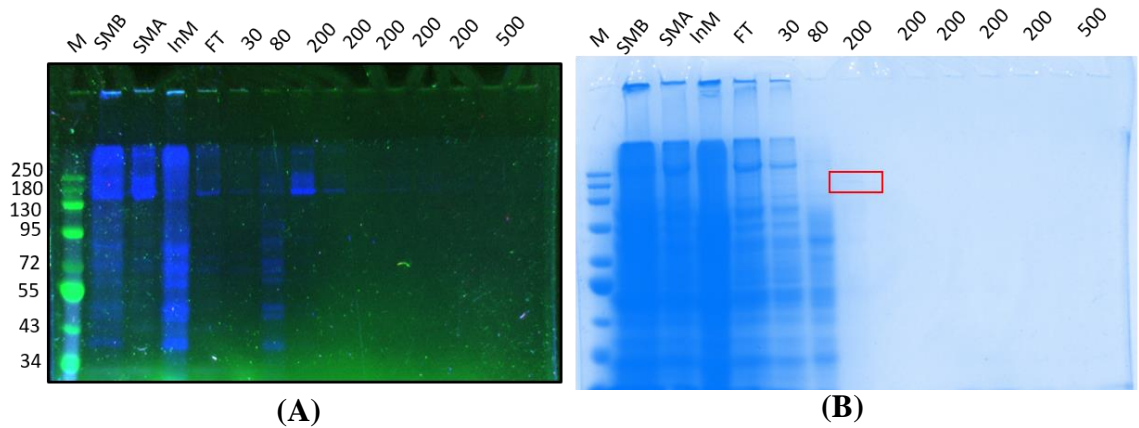
Figure 5.3: Ni-NTA chromatography of mouse ABCB5. **A)** The corresponding purification fractions on SDS gel scanned for GFP-fluorescence under Alexa488, Cyc3, and Cyc5 filters in the chemidoc system. mABCB5 eluted at 200 mM imidazole concentration. A band at ~158.6 KDa corresponds to mABCB5. **B)** The corresponding purification fractions on Coomassie-stained SDS gel **M:** protein marker in kDa **SMB:** Solubilized microsomes before centrifugation **SMA:** Solubilized microsomes after centrifugation **InM:** Insoluble microsomes. **FT:** Flow-through, **Numbers**: indicate imidazole concentration in purification buffer in millimolar (mM).

Figure 5.4A shows a representative chromatogram of bbABCG1 nickel affinity chromatography. In figure 5.4A the highest peak of the chromatogram corresponds to fractions A10 and A11 and showed a band with a molecular weight of ~94 kDa on SDS-PAGE scanned for GFP fluorescence. A band at ~94kDa corresponds to enriched bbABCG1 (sequence weight 75.4 kDa + the additional cleavable GFP and affinity tag = 21 kDa). Comparison of the GFP band intensity of the 200 mM imidazole fraction to the band intensity of the flow through indicates protein retention of about 95%, however, a huge number of protein contaminants were also visible on the Coomassie-stained SDS gel (Figure 5.4B).

A)
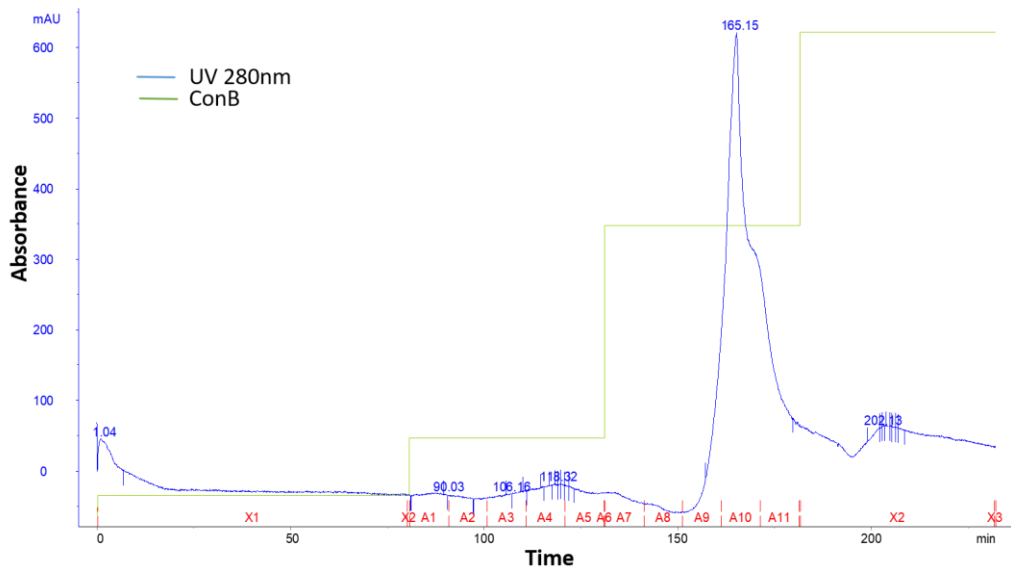
B)
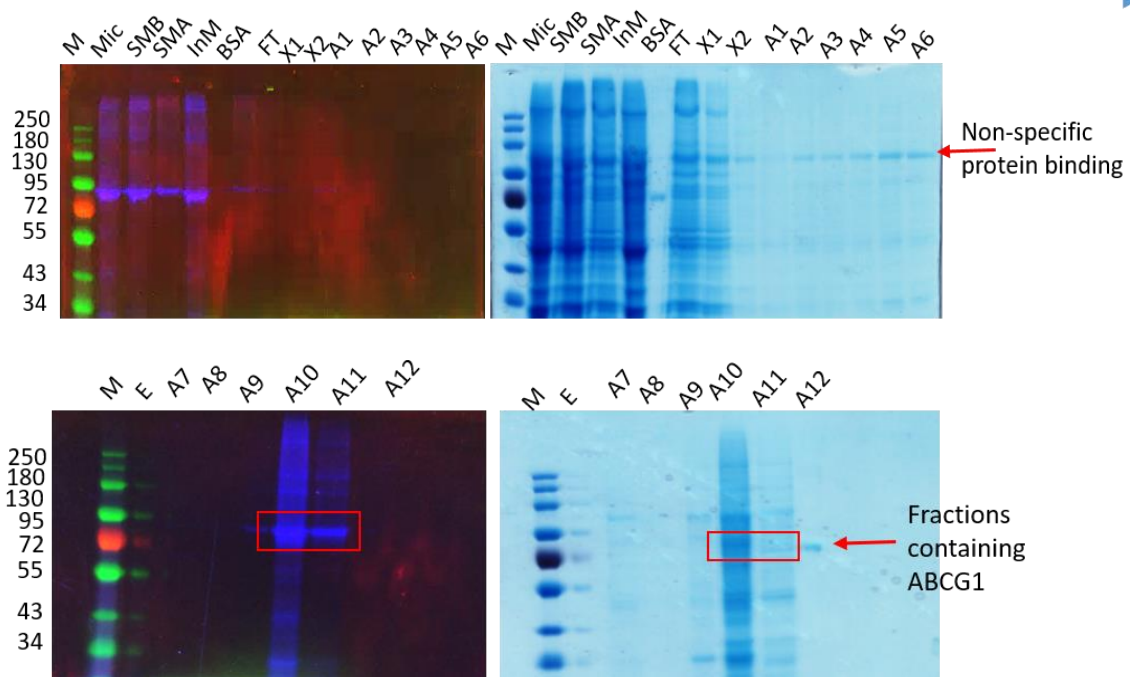
Figure 5.4: Ni-NTA Chromatography of bbABCG1. Panel A represents the chromatogram of nickel affinity chromatography. The x-axis represents time whereas the y-axis represents the absorbance UV280 (green trace represents the Imidazole concentration in buffer B). **Left Panel B** represents the gel analysis of Ni-NTA fractions on SDS-PAGE scanned for GFP fluorescence and right Panel B represents Coomassie stained SDS gel. bbABCG1 is represented by a red box. **X1:** 20mM imidazole, **A1-A6:** 80 mM imidazole concentration, **A7-A12:** 250 mM imidazole concentration, **X2:** 500 mM imidazole concentration, **M:** Protein marker in kDa, **Mic:** bbABCG1 microsomes, **SMB:** Solubilized microsomes before centrifugation **SMA:** Solubilized microsomes after centrifugation, **InM:** Insoluble microsomes, **FT:** Flow through

The purification steps for mABCG4 are shown as an ideal example, here each purification step was examined using SDS-PAGE with Coomassie staining and followed by Western blot. Figure 5.5 represents the ~40% mABCG4 microsomes solubilisation efficiency. Figure 5.6A shows the nickel-purified fractions on SDS gels scanned for GFP fluorescence. Figure 5.6B shows the SDS gel after Coomassie staining and figure 5.6C shows the Western blot. A band at ~93KDa band represents the expressed mABCG4 construct molecular weight (sequence weight 72 kDa + the additional cleavable GFP and affinity tag = 21 kDa). When comparing the amount of protein in the 200mM fraction with the amount in the flow-through, Ni-NTA affinity chromatography shows a protein retention efficiency of about 80-90%, and this efficiency remained consistent in each run. To lower the imidazole concentrations from 200 mM to less than 5 mM, the elution fractions were diluted with purification buffer A. After that, the diluted fractions were concentrated using Millipore20 concentrators with a 100 kDa cut-off for size exclusion purification.
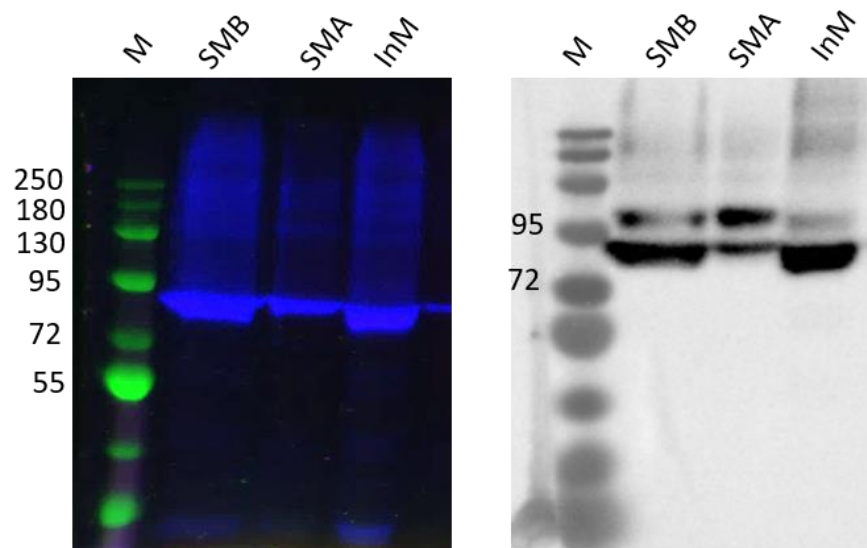
Figure 5.5: mABCG4 microsomes solubilisation. Microsomes were solubilized with solubilized buffer containing 2% DDM (w/v):0.02 % CHS for 2 hours at 4 °C. **M:** Protein marker in kDa, **SMB:** Solubilized microsomes before centrifugation, **SMA:** Solubilized microsomes after centrifugation, **InM:** Insoluble microsomes
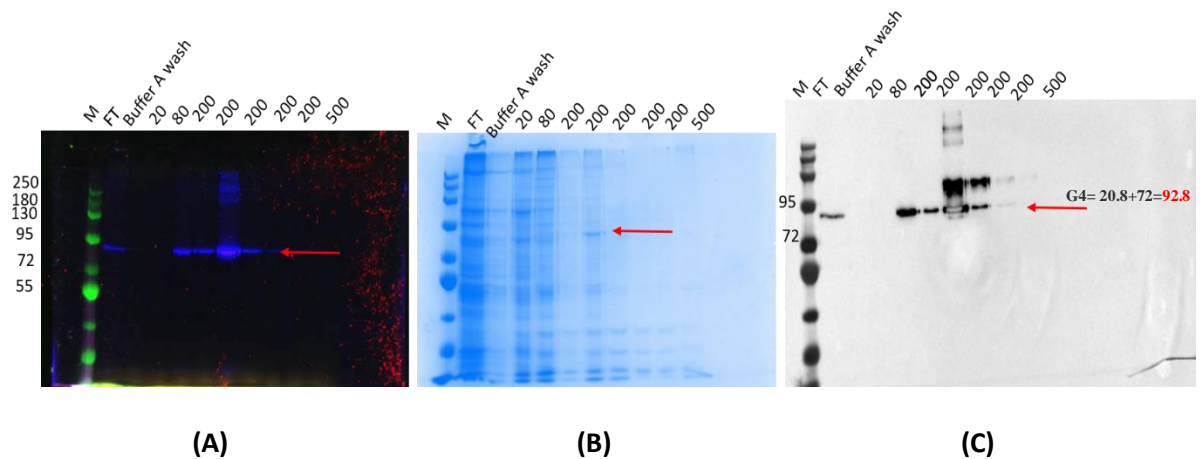
Figure 5.6: Ni-NTA chromatography of mouse ABCG4. Soluble material was applied to the HisTrap 5 ml column. All purification steps were monitored by SDS-PAGE followed by western blotting with a 25 µl sample loaded per well on the gel. **Panel A** represents the SDS-PAGE gel scanned for GFP fluorescence under Alexa488, Cyc3, and Cyc5 filters in the chemidoc system. **Panel B** represents SDS gel after Coomassie staining. **Panel C** represents the western blot. **M:** protein marker in kDa, Numbers indicate imidazole concentration in purification buffer in millimolar (mM)

5.3 Size exclusion Chromatography

A second stage of size exclusion chromatography (SEC) was used to enrich the purified proteins. 500 µl of concentrated protein after affinity chromatography were ultracentrifuged to remove large aggregates, and the protein was then placed onto the Generon SepFast 6-5000 kDa column (see section 2.18). A flow rate of 0.5 ml/minute was used for the purification. The eluent from the column was fractionated into 1ml fractions for 1.2 Colum Volumes (CV) after sample injection. As all 4 proteins have a GFP tag a fluorescence detector was also used to detect the GFP-tagged protein as it was eluted from the SEC column.

Chromatogram of all of the four proteins *i.e.* mouse ABCB5, giant panda ABCB6, little brown bat ABCG1, and mouse ABCG4 chromatograms showed a strong UV280 peak at the void volume of the column corresponding to eluent fractions 9-11 suggesting the presence of large aggregated species. Superimposed on the tail of the void-volume peak was a large shoulder between eluents 11ml to 18ml corresponding to the retention volume expected for the solubilized ABC transporter (Figure 5.7A, 5.8A, 5.9A, 5.10A).

In chromatograms, the reading for UV280 is represented by a blue colour, while the reading for GFP fluorescence is shown by a pink colour. mABCB5 size exclusion chromatogram (Figure 5.7A) UV280 readings show a sharp first peak at the void volume followed by a second, large shoulder peak. The GFP fluorescence reading is consistent with UV280

readings, indicating that the peak at the void volume contains mABCB5 aggregates whereas the shoulder peak corresponds to the monomeric form of the mABCB5 protein.

The gpABCB6 chromatogram (Figure 5.8A) displays only one UV280 peak starting at the void volume. The GFP fluorescence reading reveals a sharp peak at the void volume followed by a shoulder peak. However, the fluorescence signal at the shoulder peak is very low, indicating that only a small amount of monomeric gpABCB6 was eluted and that gpABCB6 was primarily aggregated.

bbABCG1 chromatogram (Figure 5.9A) UV 280 readings show a sharp peak at the void volume followed by a small shoulder peak. The GFP fluorescence reading is consistent with UV280 readings, indicating that the peak at the void volume contains protein aggregates whereas the shoulder peak is the monomeric form of the bbABCG1.

UV 280 measurements of the mABCG4 chromatogram (Figure 5.10A) reveal a strong peak at the void volume followed by a large shoulder peak. In contrast, the GFP fluorescence reading shows a small value corresponding to the initial sharp peak of UV280 at the void volume, but shows a large peak that is consistent with the shoulder peak of UV280 readings. This suggests that mABCG4 is mostly purified in monomeric form and a minor fraction is aggregated.

SEC fractions were further analysed using SDS PAGE. Figure 5.7B depicts the SDS-PAGE analysis of mABCB5; SDS-gel scanned for GFP fluorescence shows the molecular weight corresponding to the tagged mABCB5 construct but Coomassie staining suggests a lot of impurities. Fractions 12-15 with enriched protein were pooled and concentrated using Millipore4 100 kDa cut-off filters. The concentrated mABCB5 protein was analysed for purity and quantification on SDS PAGE relative to serial dilutions of bovine serum albumin (BSA) (Fig 5.7C). Figure 5.8B represents the SDS-PAGE of corresponding peak fractions of gpABCB6. SDS gel was scanned for GFP fluorescence which shows the molecular weight corresponding to the tagged gpABCB6 construct ~120kDa. Fractions 12 and 13 were pooled, concentrated using Millipore4 100 kDa cut-off filters, and analysed for purity on SDS-PAGE relative to BSA (Figure 5.8C). Figure 5.9B represents the corresponding peak fractions of bbABCG1 on SDS-PAGE. Figure 5.10B represents the corresponding peak fractions of mABCB5 on SDS-PAGE. Fractions 12, 13, and 14 were pooled, concentrated using Millipore4 100kDa cut-off filters, and analysed for purity on SDS-PAGE relative to serial dilutions of bovine serum albumin (BSA) (Figure 5.10C).

A)
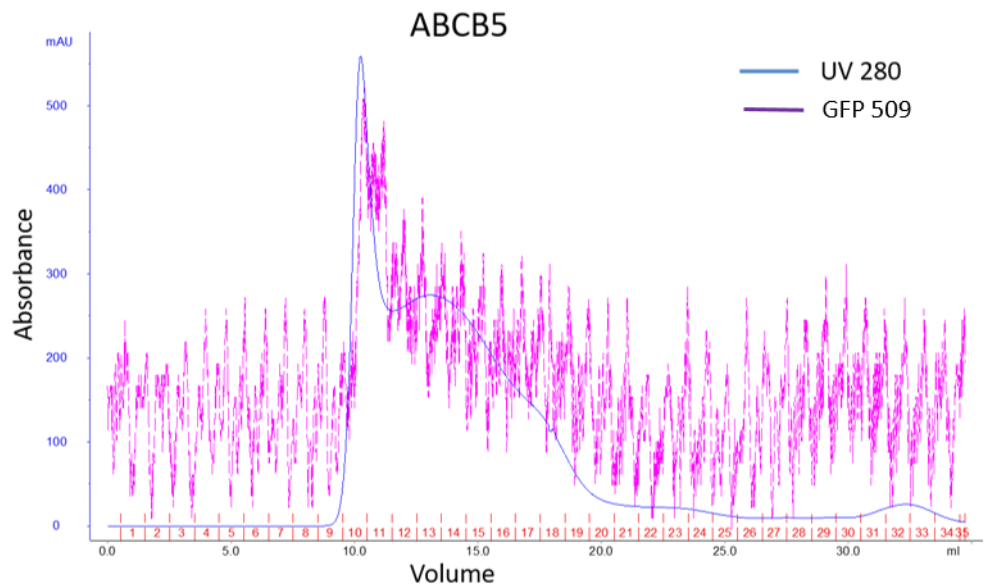
B)
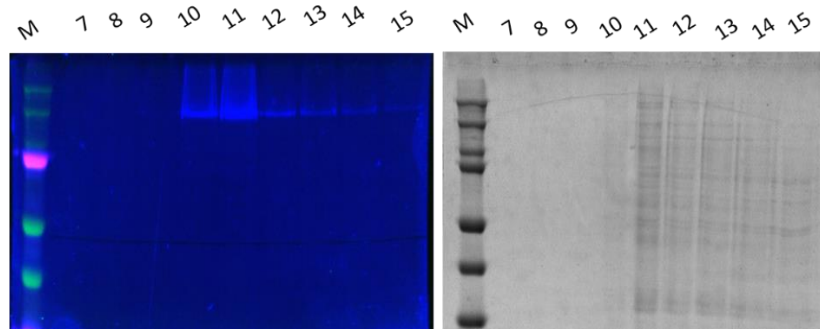
C)
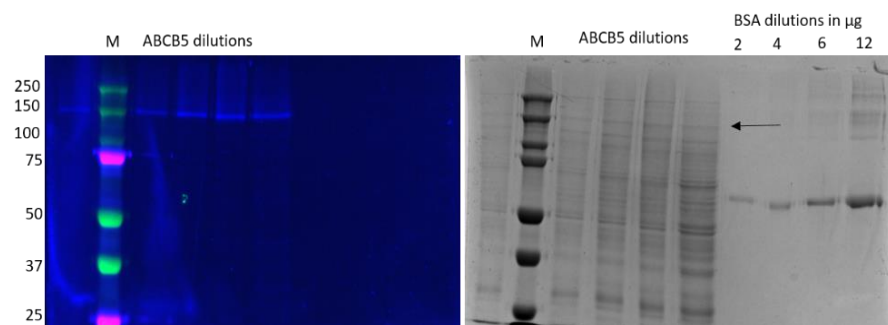
Figure 5.7: mouse ABCB5 size exclusion chromatography (SEC). The chromatogram of mABCB5 is shown in Panel A. Volume is shown on the x-axis, while absorbance at 280 nm is shown on y-axis. GFP fluorescence at 509 nm is shown as a pink trace. Chromatogram illustrates that protein was eluted in fractions 10-18 ml. Panel B represents the corresponding fractions collected during SEC. Concentrated mABCB5 loaded as 1 μl, 2 μl, 3 μl, 4 μl, 5 μl as well as BSA dilutions are shown in panel C. The Coomassie-stained SDS-PAGE gel is shown in the right panel, while the left panel displays the SDS-PAGE for GFP fluorescence.

A)
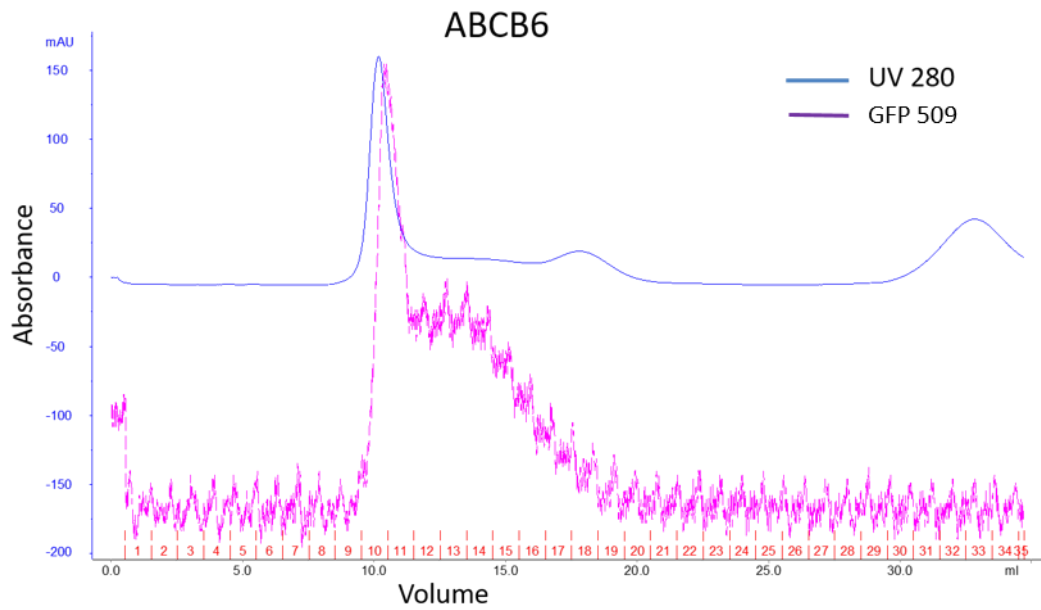
B)
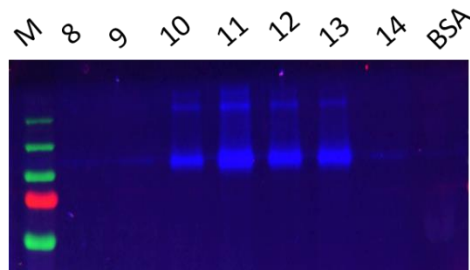
C)
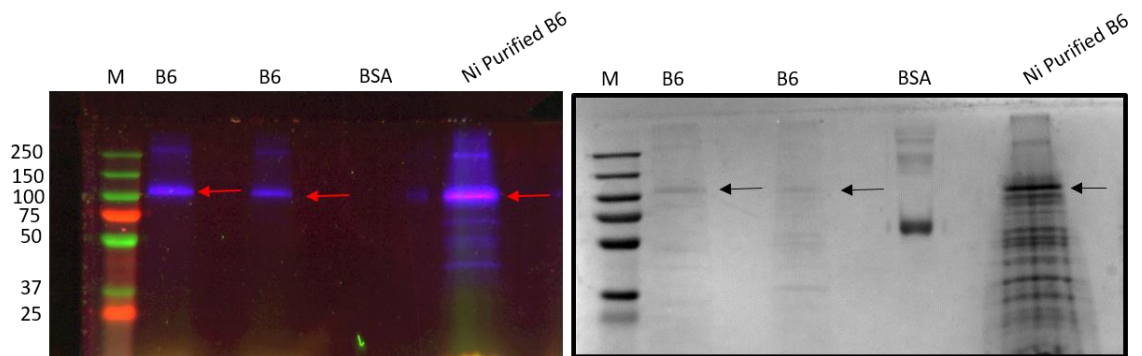
Figure 5.8: Giant panda ABCB6 size exclusion chromatography (SEC). The chromatogram of gpABCB6 is shown in Panel A. Volume is shown on the x-axis, while absorbance at 280 nm and GFP fluorescence are shown on the y-axis. Chromatogram illustrates that protein was eluted in fractions 10-18 ml but the signal at fractions 10 ml and 11 ml suggests protein aggregation. Panel B represents the pre-concentrated fractions 8 to 14 on SD-PAGE gel scanned for GFP fluorescence. Concentrated gpABCB6 loaded as 1 µl and 2 µl as well as BSA are shown in panel C. The Coomassie-stained SDS-PAGE gel is shown in the right panel, while the left panel displays the SDS-PAGE for GFP fluorescence

A)
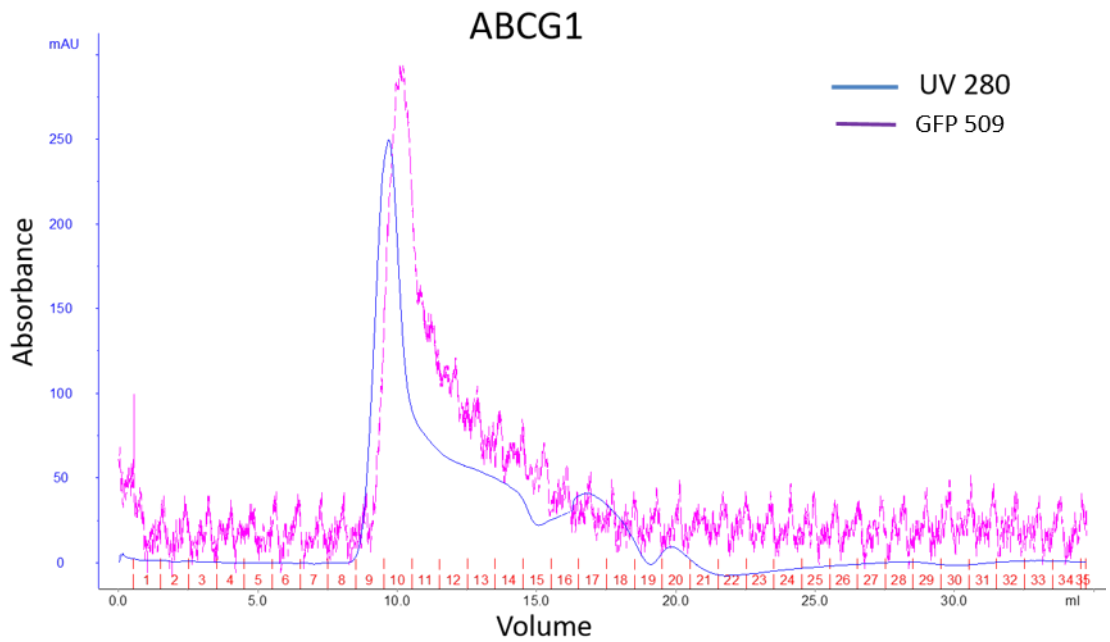
B)
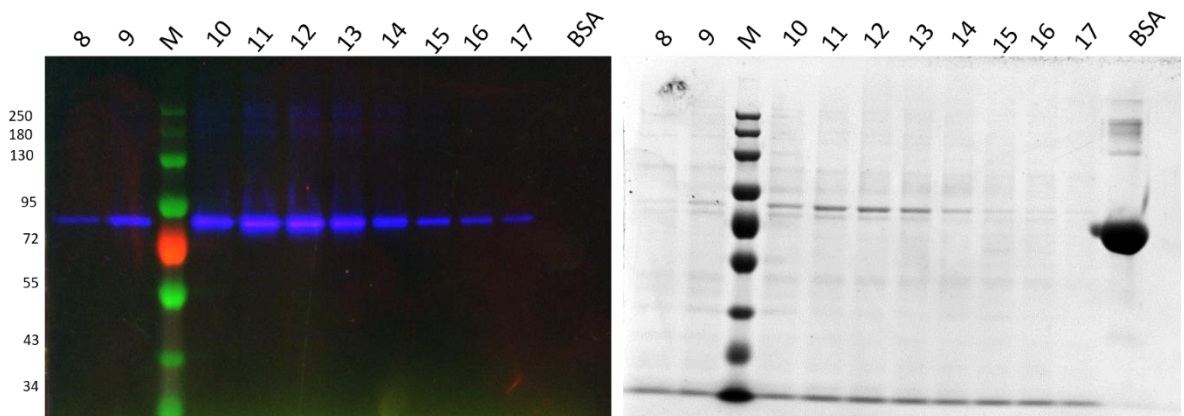
Figure 5.9: Little brown bat ABCG1 size exclusion chromatography (SEC). The chromatogram of bbABCG1 is shown in Panel A. Volume is shown on the x-axis, while absorbance at 280 nm and GFP fluorescence are shown on the y-axis. Chromatogram illustrates that protein was eluted in fractions 10-17 ml but the signal at fractions 10 ml and 11 ml suggests protein aggregation. Panel B represents the pre-concentrated fractions 8 to 17 on SD-PAGE gel scanned for GFP fluorescence. The Coomassie-stained SDS-PAGE gel is shown in the right panel, while the left panel displays the SDS-PAGE for GFP fluorescence. bbABCG1 at about ~93 kDa was seen.

A)

B)

C)
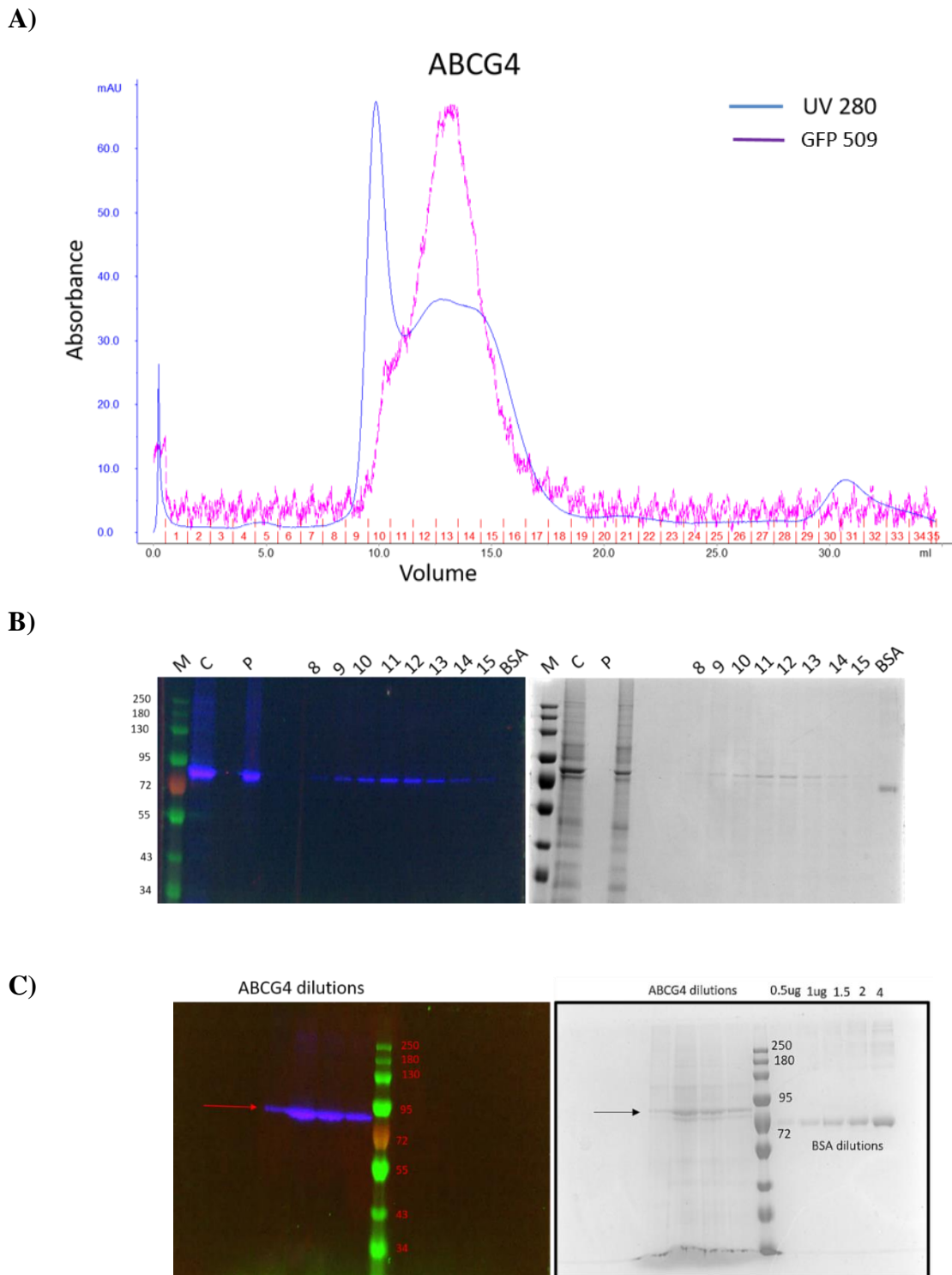
Figure 5.10: Mouse ABCG4 size exclusion chromatography (SEC). The chromatogram of mABCG4 is shown in **Panel A**. Volume is shown on the x-axis, while absorbance at 280 nm and GFP fluorescence are shown on the y-axis. Chromatogram illustrates that protein was eluted in fractions 10-18 ml but the signal at fractions 10 ml and 11 ml suggests protein aggregation. **Panel B** represents the pre-concentrated fractions 8 to 15 ml collected during SEC. Panel C represents the concentrated mABCBG4 loaded in different dilutions along with BSA dilutions. The Coomassie-stained SDS-PAGE gel is shown in the right panel, while the left panel displays the SDS-PAGE for GFP fluorescence.

5.4 Discussion and Conclusion

The aim of the work reported in this chapter was to enrich the proteins to a purity suitable for high resolution structural studies. In this context, microsomes were made from protein expressing yeast cells. In microsomes preparation, controlling the temperature during bead beating process was important for maintaining the protein quality. During the two-minute cycle of bead beating, the temperature can reach as high as 60–70 °C. So, the bead beating process carried out in cold room (4 °C) and dry ice was placed on the outer chamber of midi-bead beater throughout the process. In addition, protease inhibitors were added to the slurry to prevent yeast intrinsic proteases from destroying our recombinant protein (Lingam et al., 2017, O'Ryan et al., 2012a).

For microsomes solubilisation, DDM was chosen because of its mild and non-denaturing character and also ~80 % of the three dimensional (3D) structures of ATP binding cassette proteins (ABC family) on the protein data bank (PDB) was solved in the presence of DDM (Hofacker et al., 2007, Nandigama et al., 2019). In the pH range of 7.0 to 9.0, gpABCB6 solubilisation remained largely consistent, however there was a slight increase from roughly 10 % at pH 7.0 to 60 % at pH 9.0. (Figure 5.1). These findings are encouraging given that the His-tag affinity chromatography techniques demand a pH above 7.5 (in between 7.5-8.0) (Loughran and Walls, 2011).

We successfully expressed all of the four proteins, gpABCB6, mABCB5, bbABCG1 and mABCG4, in yeast expression systems and purified them *via* 2-step purification including nickel affinity chromatography and SEC. Following nickel affinity chromatography, SDS gels were analysed for GFP signals, which confirmed that all four of these proteins were binding to the His-trap column and eluting in the imidazole solution. However, after nickel affinity chromatography, SDS gels stained with Coomassie revealed protein impurities, with mABCG4 having the highest purity and bbABCG1 having the lowest purity. In mABCG4, an additional band of ~120 kDa band was also observed on the western blot but not visible under the GFP scan. This band could be mABCG4 that has been glycosylated or another intrinsic yeast protein that has a propensity to bind His-tag.

Aggregation in all of the four purified proteins (mABCB5, gpABCB6, bbABCG1 and mABCG4) was evident from the size exclusion chromatograms even after numerous size exclusion cycles and ultracentrifugation. These findings pointed to an interparticle equilibrium in the aqueous solutions between the aggregated and monomeric forms. It could

be because of the protein tendency to form a variety of oligomers and aggregates when concentrated, which is why in SEC chromatogram a strong peak at void volumes of 10 ml and 11 ml showed up. Another hypothesis is that all four of the proteins have GFP tags on them and the interparticle interactions in recombinant proteins have been reported to change when GFP is present (Montecinos-Franjola et al., 2020). Additionally, concentrating the protein also results in protein aggregation as well as protein loss because protein may link to the concentrator device filter.

bbABCG1 and mABCG4 showed better purity and homogeneity as judge by Coomassie stained SDS-PAGE gels after size exclusion chromatography (Figure 5.9B and Figure 5.10A). On the contrast, SDS-PAGE gels of mABCB5 and gpABCB6 showed polydispersity (see Figure 5.7B and Figure 5.8B). In order to improve protein purity and homogeneity of mABCB5 and gpABCG6 it may need to optimise its purification conditions or GFP resin purification might help to get pure protein. An additional negative purification step will be possible after tag cleavage which may increase the quality of the protein sample. Overall it could be argued that purified ABCG1 and ABCG4 was promising to use in high-resolution structural studies.

Conclusion

Taking everything into account, the UV280 readings and GFP fluorescent traces for each protein shows the aggregation in the void is mainly due to large impurities (mABCB5) or to aggregation of the target proteins (bbABCG1) or a combination (gpABCB6). This suggests that mABCG4 is the only one we could do structural work with or without cleaving the tag.

Chapter 06 Results and Discussion: Characterisation of purified proteins

Production of correctly folded, functional, and stable recombinant protein is challenging (Birch et al., 2020). To ensure that the purified protein is intact and fully functional, protein characterization is necessary before doing any structural research on it. In this chapter, the biochemical and biophysical characteristics of mABCB5, gpABCB6, bbABCG1, and mABCG4 were investigated under experimental conditions where the detergent, DDM, was present. Although the detergent mimics the membrane protein's natural environment, it may have some influence on the protein's characteristics to some extent (Thonghin, 2019).

6.1 Mass spectrometry (MS)

All constructs were newly synthesized and expressed in yeast, therefore, the purified mABCB5, gpABCB6, bbABCG1, and mABCG4 samples were examined for protein identification using mass spectrometry at the MS facility (Michael Smith building, the University of Manchester). Roughly 2 μg of purified protein was run on an SDS-PAGE gel which was then stained with Colloidal Coomassie stain. Protein bands of interest were excised and kept in 200 μl of MilliQ water (see section 2.21.2). Staff at the MS facility dissolved and processed gel fragments. The number of peptides that match the proteins in the Swiss-Prot and TrEMBL databases was used to display the results of the identification (Table 6.1-6.4). According to the MS facility's interpretation standards, an identification match with a threshold score of "3" is considered to be likely, whereas identification matches with higher levels are considered to be progressively assured.

The purified proteins were identified as the mouse ABCB5, giant panda ABCB6, little brown bat ABCG1, and mouse ABGC4 as shown by 40, 39, 30, and 34 peptide matches, respectively, in Tables 6.1- 6.4. Keratin, which is known as a common contaminant in MS (Keller et al., 2008), appeared to be the second-best match in purified gpABCB6, bbABCG1, and mABCG4. Keratin can be commonly found in skin and hair as well as in the form of dust in the laboratory. Keratin can sometimes interfere with analysis, however, in this study, its presence was not so great as to have an adverse impact. Interestingly, besides keratin, many other proteins of mouse origin were also found in purified mABCB5 protein as shown in table 6.1.

Table 6.1: MS identification scores for the mABCB5 sample using the mammal swissprot database. The table was re-written based on reports provided by MS facility, Michael Smith building, UOM. The first ten identified proteins are only shown. **OS**= OrganismName, **GN**= GeneName, **PE**= ProteinExcistence, **SV**= SequenceVersion

Identified proteins	No. of Peptides Identified
ATP-binding cassette sub-family B member 5 OS=Mus musculus GN=Abcb5 PE=2 SV=1	40
Fatty acid-binding protein, heart OS=Mus musculus GN=Fabp3 PE=1 SV=5	5
Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	4
MCG23377, isoform CRA_b OS=Mus musculus GN=Gm8797 PE=4 SV=1	4
Keratin, type II cytoskeletal 6A OS=Mus musculus GN=Krt6a PE=1 SV=3	2
Junction plakoglobin OS=Mus musculus GN=Jup PE=1 SV=3	9
ADP/ATP translocase 1 OS=Mus musculus GN=Slc25a4 PE=1 SV=4	6
Desmoplakin OS=Mus musculus GN=Dsp PE=1 SV=1	12
Actin, beta OS=Mus musculus GN=Actb PE=2 SV=1	9
ATP synthase subunit alpha OS=Mus musculus GN=Atp5a1 PE=1 SV=1	6

Table 6.2: MS identification scores for the gpABCB6 sample using the mammal swissprot database. The table was re-written based on reports provided by MS facility, Michael Smith building, UOM. The first ten identified proteins are only shown. **OS**= OrganismName, **GN**= GeneName, **PE**= ProteinExcistence, **SV**= SequenceVersion

Identified proteins	No. of Peptides Identified
ATP binding cassette subfamily B member 6 (Langereis blood group) OS=Ailuropoda melanoleuca GN=ABCB6 PE=4 SV=1	39
ABCG4 OS=Mus musculus GN=Abcg4 PE=2 SV=1	6
Keratin 1 OS=Homo sapiens GN=KRT1 PE=3 SV=1	27
Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2	25
Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3	21
Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6	23
ATP binding cassette subfamily B member 6 (Langereis blood group) OS=Callithrix jacchus GN=ABCB6 PE=4 SV=1	2

Identified proteins	No. of Peptides Identified
Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3	13
Keratin 14 OS=Macaca mulatta GN=KRT14 PE=3 SV=1	9
Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1	5

Table 6.3: MS identification scores for the bbABCG1 sample using the mammal swissprot database. The table was re-written based on reports provided by MS facility, Michael Smith building, UOM. The first ten identified proteins are only shown. **OS**= OrganismName, **GN**= GeneName, **PE**= ProteinExcistence, **SV**= SequenceVersion

Identified proteins	No. of Peptides Identified
ATP binding cassette subfamily G member 1 OS=Myotis lucifugus GN=ABCG1 PE=4 SV=1	30
Keratin 1 OS=Homo sapiens GN=KRT1 PE=3 SV=1	2
Keratin, type II cytoskeletal 5 OS=Pan troglodytes GN=KRT5 PE=3 SV=1	4
Uncharacterized protein OS=Pongo abelii PE=3 SV=1	2
Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	18
Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2	23
Keratin, type II cytoskeletal 6C OS=Homo sapiens GN=KRT6C PE=1 SV=3	16
Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6	21
Heat shock protein 70-2 OS=Mus musculus PE=3 SV=1	2
Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3	24

Table 6.4: MS identification scores for the mABCG4 sample using the mammal swissprot database. The table was re-written based on reports provided by MS facility, Michael Smith building, UOM. The first ten identified proteins are only shown. **OS**= OrganismName, **GN**= GeneName, **PE**= ProteinExcistence, **SV**= SequenceVersion

Identified proteins	No. of Peptides Identified
ABCG4 OS=Mus musculus GN=Abcg4 PE=2 SV=1	34
ATP-binding cassette sub-family G member 4 OS=Fukomys damarensis GN=H920_07488 PE=4 SV=1	2
ATP binding cassette subfamily G member 4 OS=Felis catus GN=ABCG4 PE=4 SV=1	2

ATP binding cassette subfamily G member 4 OS=Macaca mulatta GN=ABCG4 PE=4 SV=2	2
ATP-binding cassette sub-family G member 4 OS=Heterocephalus glaber GN=GW7_18683 PE=4 SV=1	4
Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2	20
Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3	10
Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1	10
Keratin, type II cytoskeletal 6C OS=Homo sapiens GN=KRT6C PE=1 SV=3	2
Heat shock 70kDa protein 2, isoform CRA_a OS=Homo sapiens GN=HSPA2 PE=3 SV=1	2

6.2 ATPase activity assays

It is important to look into a purified protein's function, especially when it comes to membrane proteins where detergent-induced denaturation may be a problem. Since the mABCB5, gpABCB6, bbABCG1, and mABCG4 were expressed and produced in a yeast expression system, it was always a concern that a recombinant higher eukaryotic protein produced in lower eukaryotic organisms could be incorrectly processed. The activity of the protein is a way to verify that functional protein can be produced in the expression system. This is also essential to check whether the protein is active under experimental settings. Additionally, this is essential for future structural studies to establish that any found structures have biological relevance.

The ATP binding cassette transporter family transports allocrites using a process powered by ATP hydrolysis (Locher, 2004). Based on their ATP hydrolysis reactions, ABC transporters are frequently monitored for their activity (Kuhnke et al., 2006, Simpson et al., 2016, Yamashita et al., 2014, Ward et al., 2013, Chavan et al., 2013) and this assay has been used to examine the functional activity of purified mABCB5, gpABCB6, bbABCG1, and mABCG4 proteins. The Chifflet assay was used to measure the ATPase activity of all of these four purified proteins (Chifflet et al., 1988). This assay measures how much inorganic phosphate (P_i) is released when ATP is hydrolyzed by ABC transporters which is directly proportional to protein activity. The basal, as well as, substrate-stimulated ATPase activity of all of the four DDM purified proteins were measured. Because we know from the purification (Chapter 05), that all of these four purified proteins contain contaminating proteins at varying levels, so to ensure that the target purified protein is the source of the

basal ATPase activity and not some other contaminating protein, substrate-stimulated ATPase activity was also performed in the presence of known substrates. For purified mABCB5 protein, verapamil, doxorubicin, vincristine, paclitaxel, camptothecin, and ivacaftor –six well-known substrates of mABCB5- (Tangella et al., 2021), were used as a stimulator, at a concentration of 10 µM. The three substrates used for purified gpABCB6 protein were piperlongumine, benzothium chloride, and glutathione (Polireddy et al., 2012). Piperlongumine, benzothium chloride was used at 10 µM concentration while glutathione was used at 5 mM concentration. For bbABCG1 protein, five substrates *i.e*. cyclosporinA, L-thyroxine, estradiol, estrone, and mestranol at 10 µM concentration were used (Cserepes et al., 2004a). For substrate-stimulated ATPase activity of mABCG4, substrates were classified into 2 categories: One class is known substrates of ABCG4 (Cserepes et al., 2004a), while, in the 2nd class, mABCG4 ATPase activity was determined in the presence of tyrosine kinase inhibitors (see detail in section 6.5)

Kinetics parameters were determined using different ATP concentrations ranging from 0-2.5 mM. Whereas, substrate-stimulated ATPase activity was calculated at a fixed concentration of 3 mM at 25 °C for 25 minutes (see Chapter 02, Section 2.21.2). Kinetics data were fitted using GraphPad prism 7 software according to the Michaelis-Menten equation (Rebbeor and Senior, 1998, Raaijmakers, 1987):

$$V = \frac{Vmax[S]}{Km+[S]}$$

The ATPase activity of mABCB5 was tested in just two independent experiments (due to a decrease in the amount of purified mABCB5 protein); therefore, to reduce the likelihood of error, the mABCB5 mean data from the two studies as well as the individual data points from each experiment are shown. For the other three target proteins (gpABCB6, bbABCG1, and mABCG4) as 3-4 independent experiments were conducted, therefore data are shown as Mean $\pm$ SD. A one-way ANOVA test followed by Dunnett's test was applied to examine the statistical significance in GraphPad Prism 7. Maximal activity (V_{max}) and a Michaelis-Menten constant (K_m) were determined by Michaelis-Menten kinetics fittings.

 DDM Purified mABCB5 protein in apo-state exhibited ATPase activity with maximal activity (V_{max}) of 17.97 $\pm$ 0.95 nmol/min/mg of protein and a Michaelis-Menton constant (K_m) of 0.77 $\pm$ 0.11 mM (Figure 6.1A). This shows that the mABCB5 possesses basal ATPase activity under experimental conditions. Substrate-stimulation results showed that

the maximal activity of mABCB5 increased by 3-fold (47.010 ± 20 nmol/min/mg) and 5-fold (63 ± 8 nmol/min/mg) in the presence of doxorubicin and paclitaxel respectively (Figure 6.1B). On the other hand, verapamil and ivacaftor didn't stimulate the activity of detergent-purified mABCB5. Similarly, vincristine and camptothecin did not show any significant effect on mABCB5 basal ATPase activity. Kawanobe et al., reported a basal ATPase rate of 65 nmol/min/mg and K_m 1.8 mM for the ABCB5 protein purified from an insect expression system this is somewhat higher than the results we found, additionally, Kawanobe and co-workers found 1.5-fold and 3-fold increase in mABCB5 basal ATPase activity in the presence of doxorubicin and docetaxel respectively (Kawanobe et al., 2012). Possible explanations for this variation might include the ATPase activity may be influenced by the expression system utilised to produce the protein, removal of detergents, detergents to protein ratio, ortholog variation, and incubation time. The amino acid identity and similarity of ABCB5 and ABCB1(P-gp) sequences are 55.0% and 89.8%, respectively, indicating significant homology (Kawanobe et al., 2012). In previous studies, it has been shown that verapamil, ivacaftor, and vincristine can stimulate the ABCB1 ATPase activity (Thonghin, 2019, Sáfár et al., 2016, Gottesman et al., 1995); while in our study these three substrates showed no significant effect on mABCB5 ATPase rate. These variances could be the result of how these transporters recognise substrates differently.

DDM Purified gpABCB6 in apo-state exhibited ATPase activity with maximal activity (V_{max}) of 24.83 ± 0.85 nmol/min/mg and a Michaelis-Menten constant (K_m) of 0.26 ± 0.04 mM (Figure 6.2 A). Substrate-stimulation results showed that glutathione increased the gpABCB6 ATPase rate 2-fold $i.e.$ 48.7 ± 1.28 nmol/min/mg (Figure 6.2 B). This is consistent with an earlier report in which they reported the basal ABCB6 as V_{max} 27 nmol/min/mg and 3-fold increase in the presence of glutathione (Song et al., 2021). Piperlongumine and benzothium chloride appeared to slightly reduce the baseline gpABCB6 ATPase activity, which is also consistent with the previous research (Polireddy et al., 2012)

DDM Purified bbABCG1 exhibited ATPase activity in apo-state with maximal activity (V_{max}) of 54.78 ± 1.63 nmol/min/mg and a Michaelis-Menton constant (K_m) of 0.35 ± 0.04 mM (Figure 6.3 A). Estradiol and mestranol stimulated bbABCG1 ATPase activity by 1.3-fold (71.4 ± 0.625 nmol/min/mg) and 1.5-fold (90 ± 5.50 nmol/min/mg) respectively (Figure 6.3 B). This agrees with Kalpana et al., 2019 who reported V_{max} of detergent-purified ABCG1 18.8 ± 0.4 nmol/min/mg and K_m 1.8 ± 0.1 mM (Shanmugarajah et al., 2019), but interestingly, this contradicts Hiroshi et al., 2013, who reported that the detergent-purified

ABCG1 does not show any ATPase activity (Hirayama et al., 2013). CyclosporinA, L-thyroxine, and estrone seemed to have an inhibitory effect on detergent-purified bbABCG1 ATPase activity (Figure 6.5 B).

Detergent-purified mABCG4 exhibited ATPase activity in the apo-state with maximal activity (V_{max}) of 21.69 ± 1.23 nmol/min/mg and a Michaelis-Menten constant (K_m) of 0.39 ± 0.07 mM respectively (Figure 6.4 A). This is consistent with the previous finding which reported 18 nmol/min/mg basal ATPase rate of ABCG4 (Cserepes et al., 2004a). Whilst, cyclosporinA and estrone increased the mABCG4 ATPase activity these two substrates had an inhibitory effect on bbABCG1 ATPase activity; while L-thyroxine, hexestrol, erythritol, and calcifediol do not show any significant effect on mABCG4 ATPase activity (Figure 6.4 B). Since the data for erythritol, hexetrol, and L-tyroxine are not very significant, it is possible that there was an experimental error in these three substrates. Several tyrosine kinase inhibitors (TKIs) have been shown to interact with ABC transporters in recent years, including imatinib, nilotinib, dasatinib, ponatinib, gefitinib, erlotinib, lapatinib, vandetanib, sunitinib, and sorafenib (*e.g.,* ABCB1, ABCC1, ABCG2, ABCC10). TKIs can act as substrates or inhibitors depending on the expression of particular pumps and their affinity for transporters (Beretta et al., 2017). For mABCG4, seven different TKI (cediranib, dovitinib, gefitinib, lenvatinib, tivozanib, masitinib, and linfanib) were tested for ATPase activity. Results revealed that, aside from cediranib, the other six TKIs enhanced mABCG4's ATPase activity. This can be justified by TKIs exhibiting substrate-like qualities at low concentrations but inhibiting the function of the transporter at higher concentrations (Anreddy et al., 2014). Dovitinib, Gefitinib, and Masitinib are more significant hits. A dose-response curve for these three medications, as well as one that includes additional pharmaceuticals, can be produced by follow-up studies in the future.

Generally, it appears that the ABC transporter family exhibits significant variation in terms of ATPase activity depending on the expression system, detergent type, detergent-to-protein ratio, lipid environment, and reconstitution efficiency (Meng, 2017, Chavan et al., 2013). Overall, the collective results imply that all of these four purified proteins possess a basal ATPase activity under conditions where the detergent is still present. This indicates that mABCB5, gpABCB6, bbABCG1, and mABCG4 prepared in this study retained their proper functionality in ATP hydrolysis. However, because of time limitation, we determined the basal as well as substrate-stimulated ATPase rate of only detergent-purified proteins. It is interesting to see the ATPase activity of these protein targets after reconstitution into

liposomes or nanodisc as in previous studies after reconstitution ATPase activity elevated significantly (Thonghin, 2019, Chavan et al., 2013, Hirayama et al., 2013, Klappe et al., 2009).
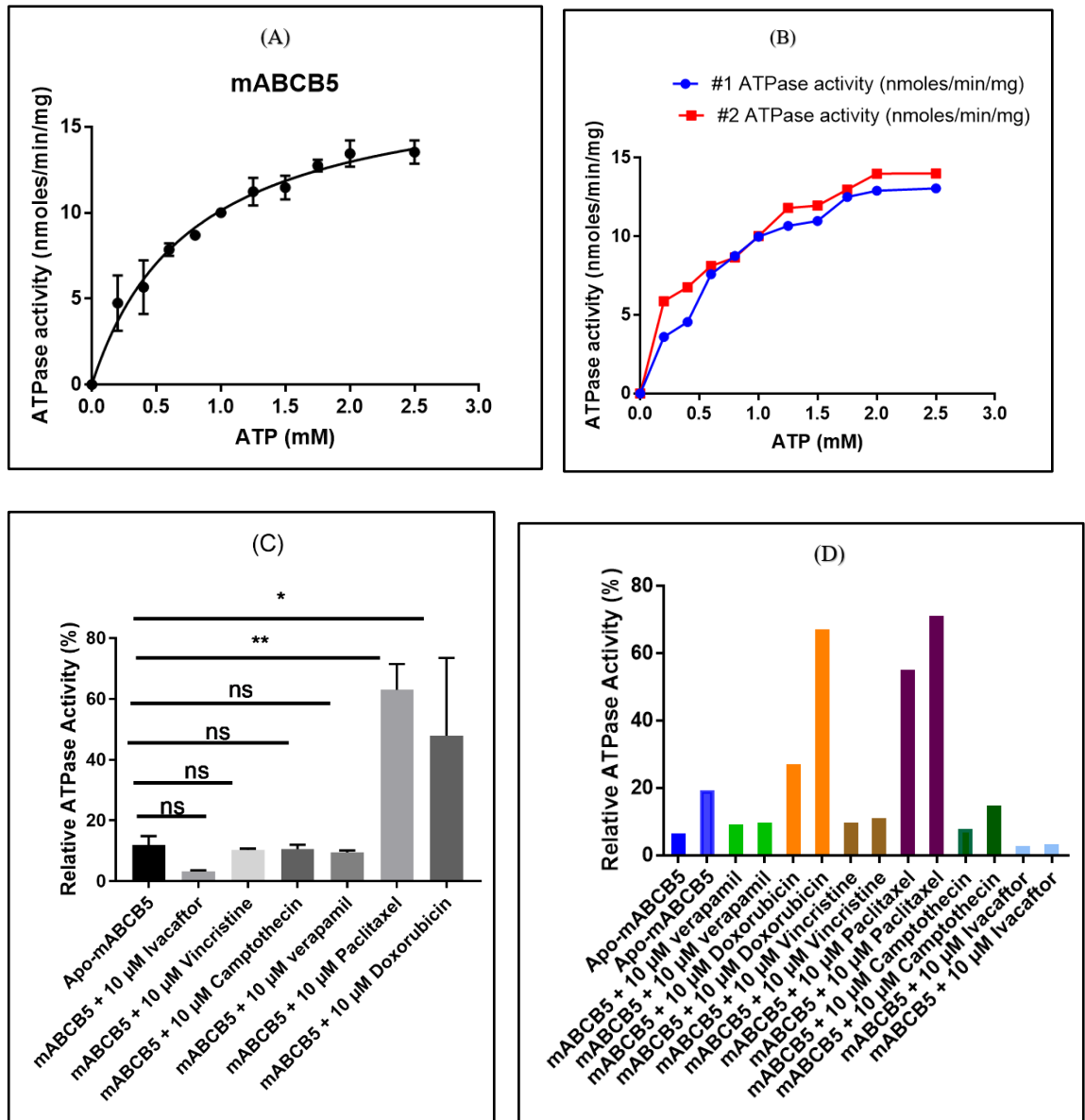
Figure 6.1: ATPase activity of mABCB5 measured *via* ATPase chifflet assay. (A) Apo-mABCB5 in detergent (DDM) shows a maximal ATPase activity of 17.97 ± 0.9 nmol/min/mg and a K_m estimated at 0.77 ± 0.10 mM. **(B)** Apo-ATPase activity of each experiment **(C)** Relative maximal ATPase activities of mABCB5 showed that paclitaxel stimulated ATPase activity increased by 5-fold compared to that of the basal activity in a detergent environment. Each data point represents the mean $\pm$ SD of the duplicate experiment. When there is no visible error bar, the SD is less than the corresponding symbol. **(D)** Relative maximal ATPase activities of mABCB5 of each experiment. **ns:** non-significant.

Table 6.5: Kinetics parameters of mABCB5, gpABCB6, bbABCG1 and mABCG4 in apo-state.

	mABCB5	gpABCB6	bbABCG1	mABCG4
Vmax (nmoles/min/mg)	17.97 ± 0.95	24.82 ± 0.85	54.78 ± 1.63	21.69 ± 1.23
Km (mM)	0.77 ± 0.11	0.26 ± 0.04	0.35 ± 0.04	0.39 ± 0.07

A)

B)
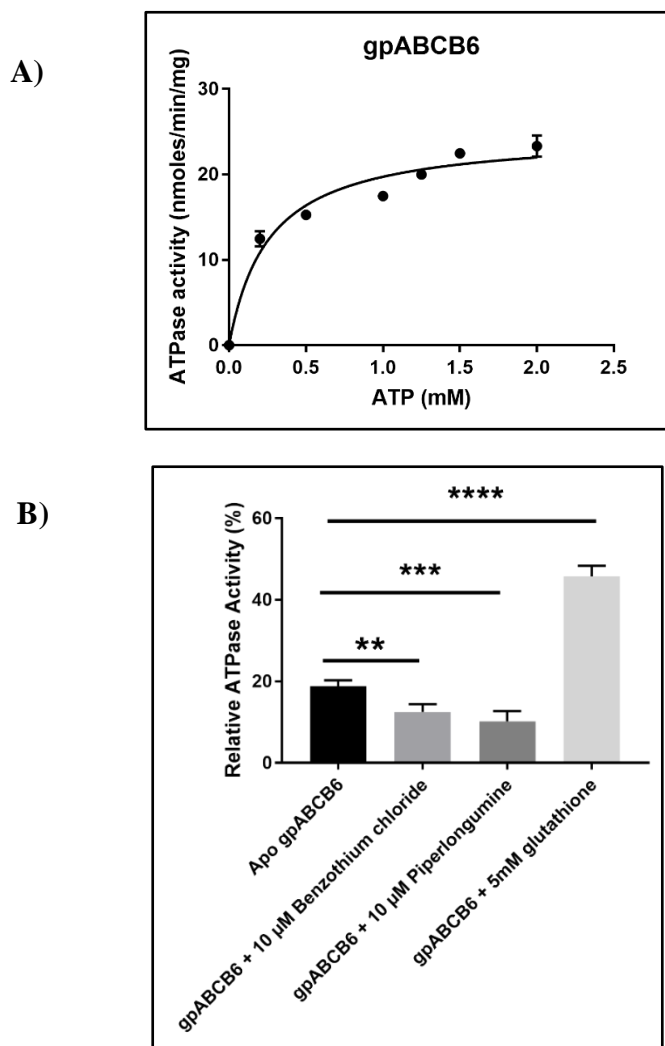
Figure 6.2: ATPase activity of gpABCB6 measured *via* ATPase chifflet assay. (A) Apo-gpABC6 in detergent (DDM) shows a maximal ATPase activity of 24.83 ± 0.85 nmol/min/mg and a K_m estimated at 0.2586 ± 0.03 mM. (B) Relative maximal ATPase activities of gpABCB6 showed that glutathione stimulated ATPase activity increased by 2-fold compared to that of the basal activity in a detergent environment. Each data point represents the mean $\pm$ SD calculated from four independent experiments. When there is no visible error bar, the SD is less than the corresponding symbol. A one-way ANOVA test is applied for comparison of statistical significance. The p-values <0.05, 0.01, 0.001 and 0.0001 are indicated with *, **, ***, and **** compared to apo-gpABCB6.
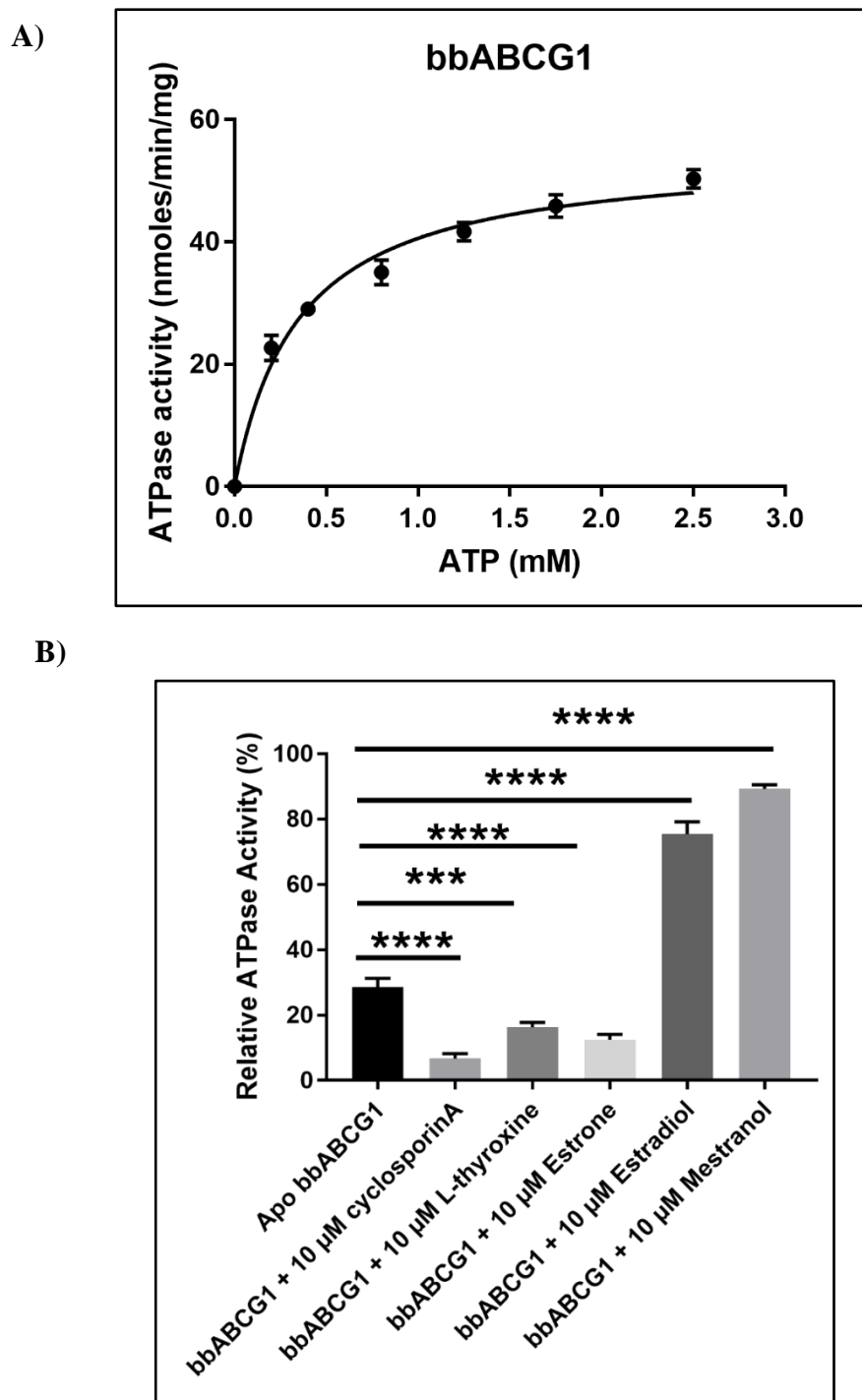
Figure 6.3: ATPase activity of bbABCG1 measured *via* ATPase chifflet assay. (A) Apo-bbABCG1 in detergent (DDM) shows a maximal ATPase activity of 54.78 ± 1.63 nmol/min/mg and a K_m estimated at 0.35 ± 0.03 mM. **(B)** Relative maximal ATPase activities of bbABCG1 showed that mestranol stimulated ATPase activity increased by 6-fold compared to that of the basal activity in a detergent environment. Each data point represents the mean $\pm$ SD calculated from three independent experiments. When there is no visible error bar, the SD is less than the corresponding symbol. A one-way ANOVA test is applied for comparison of statistical significance. The p-values <0.05, 0.01, 0.001 and 0.0001 are indicated with *, **, ***, and **** compared to apo-bbABCG1.

A)

B)

C)
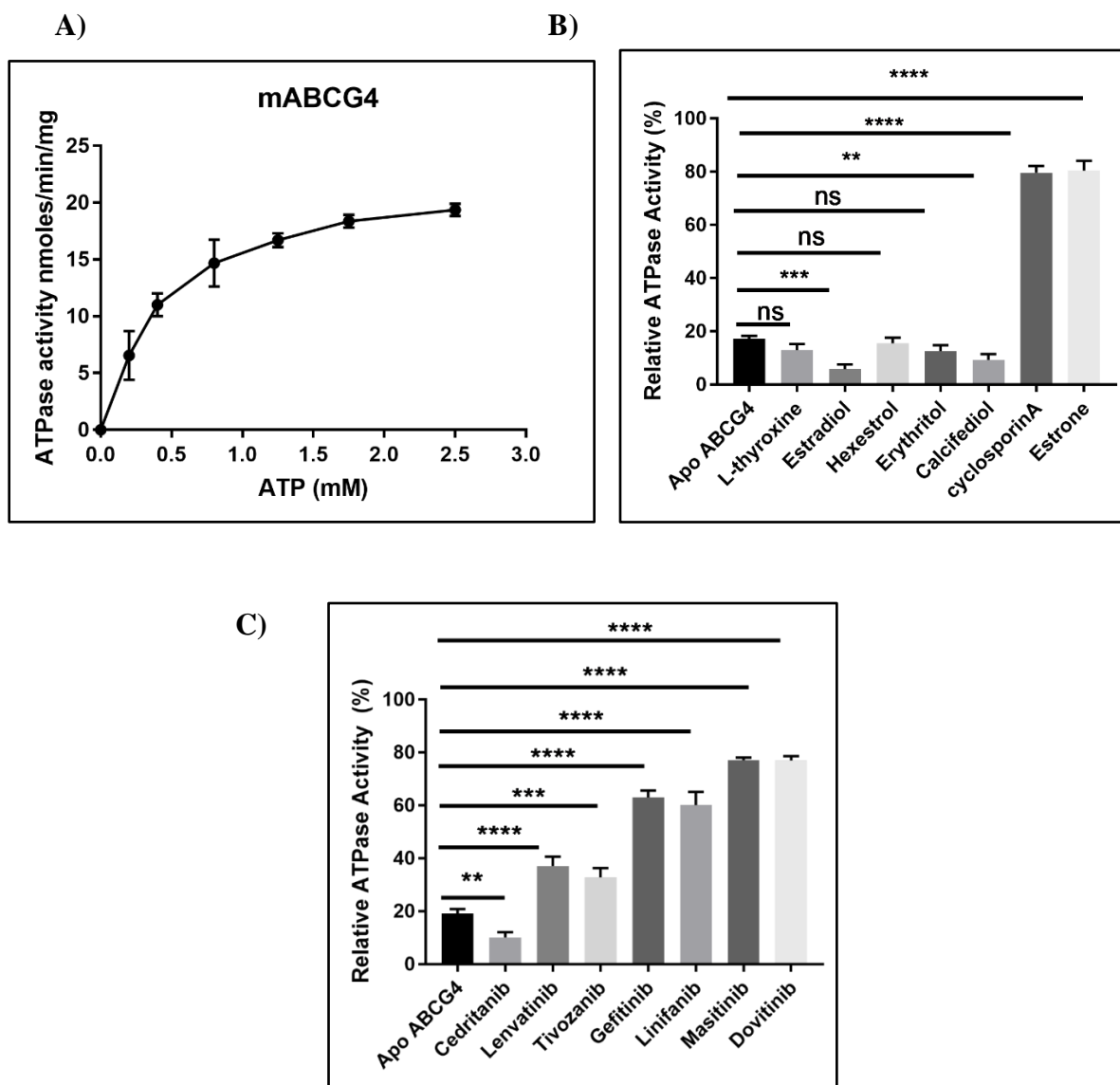
Figure 6.4: ATPase activity of mABCG4 measured *via* ATPase chifflet assay. (A) Apo-mABCG4 in detergent (DDM) shows a maximal ATPase activity of 21.69 ± 1.228 nmol/min/mg and a K_m estimated at 0.4 ± 0.07 mM. **(B)** Relative maximal ATPase activities of mABCG4 showed that cyclosporinA and estrone stimulated ATPase activity increased by 4-fold compared to that of the basal activity in a detergent environment. **(C)** Relative maximal ATPase activities of mABCG4 in the presence of tyrosine kinase inhibitors. Each data point represents the mean $\pm$ SD calculated from three independent experiments. When there is no visible error bar, the SD is less than the corresponding symbol. A one-way ANOVA test is applied for comparison of statistical significance. The p-values $<0.05, 0.01, 0.001$ and 0.0001 are indicated with *, **, ***, and **** compared to apo-mABCG4. **ns:** non-significant.

6.3 Nucleotide binding assay

Nucleotide binding is one of the key characteristics of ATP-binding proteins including mABCB5, gpABCB6, bbABCG1, and mABCG4. A crucial trigger for the ABC protein's transport mechanism to begin appears to be the binding of ATP (Frank et al., 2016, Oldham

et al., 2008). mABCB5 as a full-length transporter possesses two nucleotide-binding sites located in separated NBDs (Kawanobe et al., 2012), whereas the half-transporters gpABCB6, bbABCG1, and mABCG4 only encode one NBD; a functional homodimer, however, always has two NBDs (Wang et al., 2020b, Skarda et al., 2021, Yoshikawa et al., 2002). To analyse the ATP binding, A fluorescent ATP analog (TNP-ATP) was used (Sharom et al., 2003). It is easier for researchers to analyse ATP-Protein interactions when 2,4,6-trinitrophenol (TNP) is conjugated onto an ATP molecule. The ribose sugar of the ATP molecule is coupled to the TNP fluorophore, which Azegami and Iwai created in 1964 (AZEGAMI and IWAI, 1964). In 1975, Hiratsuka used infrared spectroscopy and NMR to confirm the structure of TNP-ATP (Hiratsuka, 1975). TNP-ATP exhibits a broad excitation spectrum ranging from 350 to 520nm with excitation maxima at 410 nm and 470 nm and an emission spectrum between 500 nm-600 nm.

In this research, an excitation peak of 410 nm was selected to prevent the overlapping of the GFP signal because all target proteins were GFP-tagged. Periodic measurements of the emission spectra between 510 nm and 560 nm were taken. To prevent cross-signalling between the wells, the binding assay was carried out in 96-well plates with clear bottom wells (see methods section 2.21.7). Figure 6.5 displays the fluorescence spectra of buffer baseline, mABCB5/gpABCB6/bbABCG1/mABCG4 alone, TNP-ATP alone, and protein combinations with varying TNP-ATP concentrations. BSA was used as a negative control in ATP binding assay. Fluorescence of TNT-ATP of these proteins is shown as mABCB5 (Figure 6.5 A), gpABCB6 (Figure 6.5 B) bbABCG1 (Figure 6.5 C), and mABCG4 (Figure 6.5 D). The emission peak of TNP-ATP in water lies between 550 nm and 560 nm. The emission spectra of this substance resembled to BSA. The fluorescence signal was found to increase when mABCB5/gpABCB6/bbABCG1/mABCG4 were added to TNP-ATP. Additionally, the emission peak was moved from 560 nm to 530-540 nm. This finding is consistent with the phenomenon previously discussed in the study of bacterial histidine kinases, which suggests that when TNP-ATP binds to the ATP-binding protein, the emission maxima shift towards shorter wavelengths (Guarnieri et al., 2011).

In the presence of a fixed protein quantity, fluorescent signals were found to increase along with TNP-ATP concentrations. However, the signal appeared to be quenched at the high TNP-ATP concentrations (*e.g.* 500 µM; Figures 6.5A, 6.5B, and 6.5C). This can be explained by the fluorescence reabsorption phenomenon because the absorption spectrum of TNP-ATP spans between 350 nm and 540 nm, as a result, the longer wavelength section of

the absorption spectrum overlaps with its shorter wavelength emission region beginning at 500 nm. Therefore, the variation in fluorescence concentration on the fluorescence emission spectrum can be directly attributed to fluorescence reabsorption and can be modelled using Beers law *i.e.* $A = \epsilon/c$.
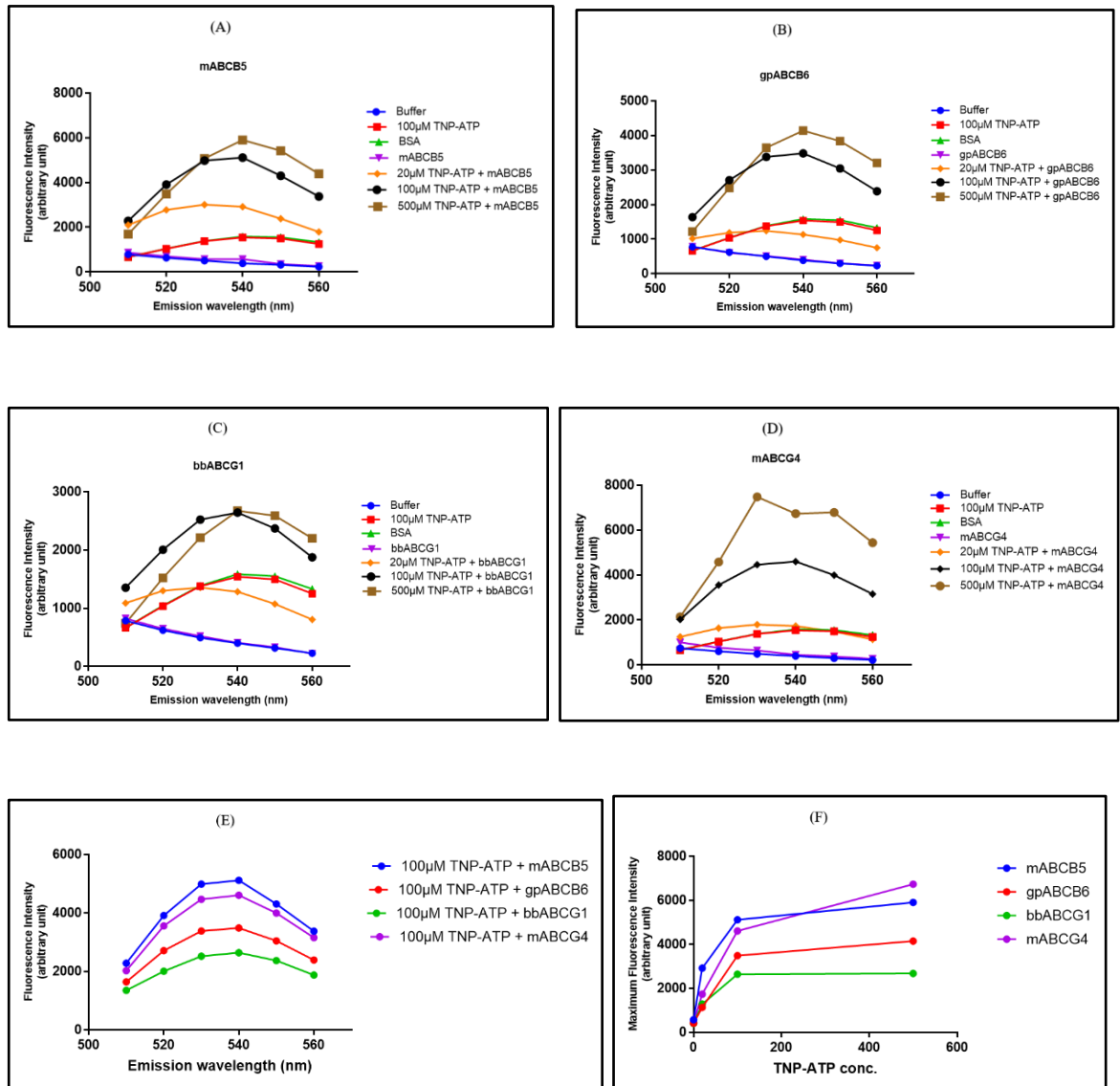
Figure 6.5: Graphical representation of TNP-ATP binding assay. A) Shows the binding of TNP-ATP with mABCB5. **B)** Shows the binding of TNP-ATP with gpABCB6. **C)** Shows the binding of TNP-ATP with bbABCG1. **D)** Shows the binding of TNP-ATP with mABCG4. **E)** Shows comparison of all four target proteins for TNP-ATP binding affinity at 100 μM concentration. **F)** Shows comparison of all four target proteins at different TNP-ATP concentration i.e. 0 μM, 20 μM, 100 μM and 500 μM.

TNP-ATP interacts with each of these proteins in the following order: mABCB5> mABCG4> gpABCB6> bbABCG1 as seen in Figure 6.5E and Figure 6.5F. Our conclusion, which is consistent with the previous section 6.4, is that all of the target proteins are

functional and have a affinity for ATP. TNP-ATP binding affinity, however, follows the reverse order of V_{max} which is as follows: bbABCG1 > gpABCB6 > mABCG4 > mABCB5. One possible explaniation for this might be amino acids variations at ATP-binding pocket of these target proteins because TNP-ATP fluorescence is extremely environment-sensitive, changing its spectrum features as a result of polarity and viscosity changes that occur when it binds to a protein's binding pocket (Woodbury et al., 2021). The early Hiratsuka and Uchida investigation clearly demonstrated and accepted an important point which is the kinetics of TNP-ATP binding to a protein differ from the kinetics of ATP binding to a protein (Hiratsuka and Uchida, 1973). Determining the binding parameters of ATP must therefore be done in a manner that does not solely rely on the information obtained through TNP-ATP binding. The addition of a second technique, such as P^{32}-labeled ATP in addition to TNP-ATP fluorescence, can be used in a follow-up study in the future to confirm the binding of these target proteins to ATP.

6.4 Dynamic light scattering (DLS)

The existence of proteins in a monodisperse state is a fundamental condition for high-resolution structural studies *i.e.* for cryo-EM, small angle X-ray scattering (SAXS), and also for the nucleation and development of protein crystals (Aivaliotis et al., 2003). Additionally, it was proposed that monodisperse macromolecule solutions crystallise easily, whereas polydisperse systems do so infrequently (Baritaki et al., 2002). Therefore, it is vital to undertake an experiment to ascertain whether a purified protein is monodisperse or not before conducting any high-resolution structural research on target protein. In this study, purified proteins *i.e.* mABCB5, gpABCB6, bbABCG1, and mABCG4 were investigated for homogeneity, aggregation, and dynamic size distribution *via* DLS. We chose DLS because it is a non-destructive process that requires a small amount of a pure protein sample, can be completed in a few minutes, and is quick to perform. The samples were prepared as described in section 2.21.6 and the UNcle instrument (Unchained Labs Inc.) was used to conduct a duplicate examination (Unchained Labs Inc.). The University of Manchester's Biomolecular Analysis Facility houses this instrument. The experiment was performed in four different states including apo-, substrate-bound, ATP-bound, and AMP-PNP-bound *i.e.* nucleotide-bound state. Figure 6.7 shows DLS profiles of mABCB5 in each state. The left panel's intensity distributions depict a bimodal distribution that represents two populations with different hydrodynamic diameters in all four states. The right panel depicts the mass

distribution of this data, which revealed one main population of particles. The particle hydrodynamic diameter of the apo-mABCB5 particle, which is revealed by the first peak of the intensity distribution, is 10 ± 2.1 nm, which is about the theoretical length of the mABCB5 *i.e.* 9.22 nm as determined in the Chimera using the homology model (Eswar et al., 2008). Particles with a diameter of ~100 ± 18 nm made up another population. This population could be a mixture of oligomers or small protein aggregates. The sixth power of the particle radius (=squared of volume) directly correlates with the scattering intensity of biological molecules (Goldburg, 1999). This refers to the fact that a small number of aggregates or high-mass particles can produce a significant DLS amplitude and, in some situations, dominate the distribution. This demonstrates that DLS is a sensitive tool for analysing sample aggregation and homogeneity (Walian et al., 2004). mABCB5 mass distribution plot suggested that aggregation was present in a negligible amount of the sample. Therefore, under the conditions described, the purified mABCB5 sample can be said to be monodisperse which is not in agreement with Coomassie stained SDS-gel (see Chapter 05). Additionally, mABCB5 displays comparable DLS profile patterns across all experimental setups, indicating that nucleotides hardly have any impact on protein aggregation or mass distribution.

The DLS profiles of gpABCB6 are shown in Figure 6.6. The intensity distribution plot of apo-gpABCB6 shows three populations; the first, extremely small peak, with a hydrodynamic diameter of only ~10 ± 0.6 nm, corresponds to the gpABCB6 monomer which is slightly bigger than the theoretically determined diameter of gpABCB6 monomer (8.44 nm). In contrast, the second and third peaks with hydrodynamic diameters of ≥ 100 nm greater correspond to large aggregates. It's interesting to note that the third peak (>100nm hydrodynamic diameter) disappeared in substrate-bound gpABCB6. The mass distribution plots revealed that aggregation appeared to be very small. This can be explained by if large particles are present in the purified protein sample, the intensity and volume will tend to represent a bigger diameter, because those large particles will scatter more light and also occupy more space, even if they are minority in the sample. In a number distribution, however, they will represent less weight in the mass distribution because they are minority (Goldburg, 1999). The analysis of post-hydrolytic trapped (ADP-bound vanadate-trapped, ADP-V_i) and AMP-PNP-bound data for gpABCB6 was not performed.

The DLS profiles of bbABCG1 and mABCG4 are shown in Figures 6.8 and 6.9, respectively. Intensity distribution graphs demonstrate a bimodal distribution of particles in all four states with hydrodynamic diameters of ~10 ± 0.2 nm which corresponds to

bbABCG1/mABCG4 monomer and ~100 ± 0.4 nm which corresponds to small aggregates. These findings are consistent with mABCB5 findings. Theoretically, bbABCG1 and mABCG4 have 7.86 nm and 7.76 nm diameters, respectively determined in Chimera using homology model. bbABCG1 mass distribution plots reveal that there is only a slight aggregation. In contrast, the mABCG4 mass distribution plots only display one peak suggesting mABCG4 close to homogeneity.

As demonstrated by (Papish et al., 2002) who analysed calmodulin in the presence and absence of Ca^{+2} ions and predicted a detectable hydrodynamic diameter change as tiny as 0.5 nm or 5 Å, DLS has also been proposed as a useful method for detecting large-scale conformational changes in the form of altered hydrodynamic diameter. It is interesting to note that in all four proteins, this study found no evidence of a substantial change in hydrodynamic diameter under the various experimental settings, suggesting that DLS is unable to resolve any modifications brought on by nucleotide-induced conformational changes. It is possible that only a small fraction of the population (if none) underwent conformational changes (Thonghin, 2019). A tiny proportion of the population would not significantly contribute to a detectable separated population (peak) in the intensity distribution. The polydispersity index (PDI) of the apo states of all four of these proteins in apo-state which also suggests that mABCG4 (PDI: 0.20) is monodisperse and homogeneous. Upon examination of all of the data collected from this single experiment (**Table 1**), prioritizing the target proteins was done as: mABCG4> gpABCB6> bbABCG1> mABCB5 for future structural studies.
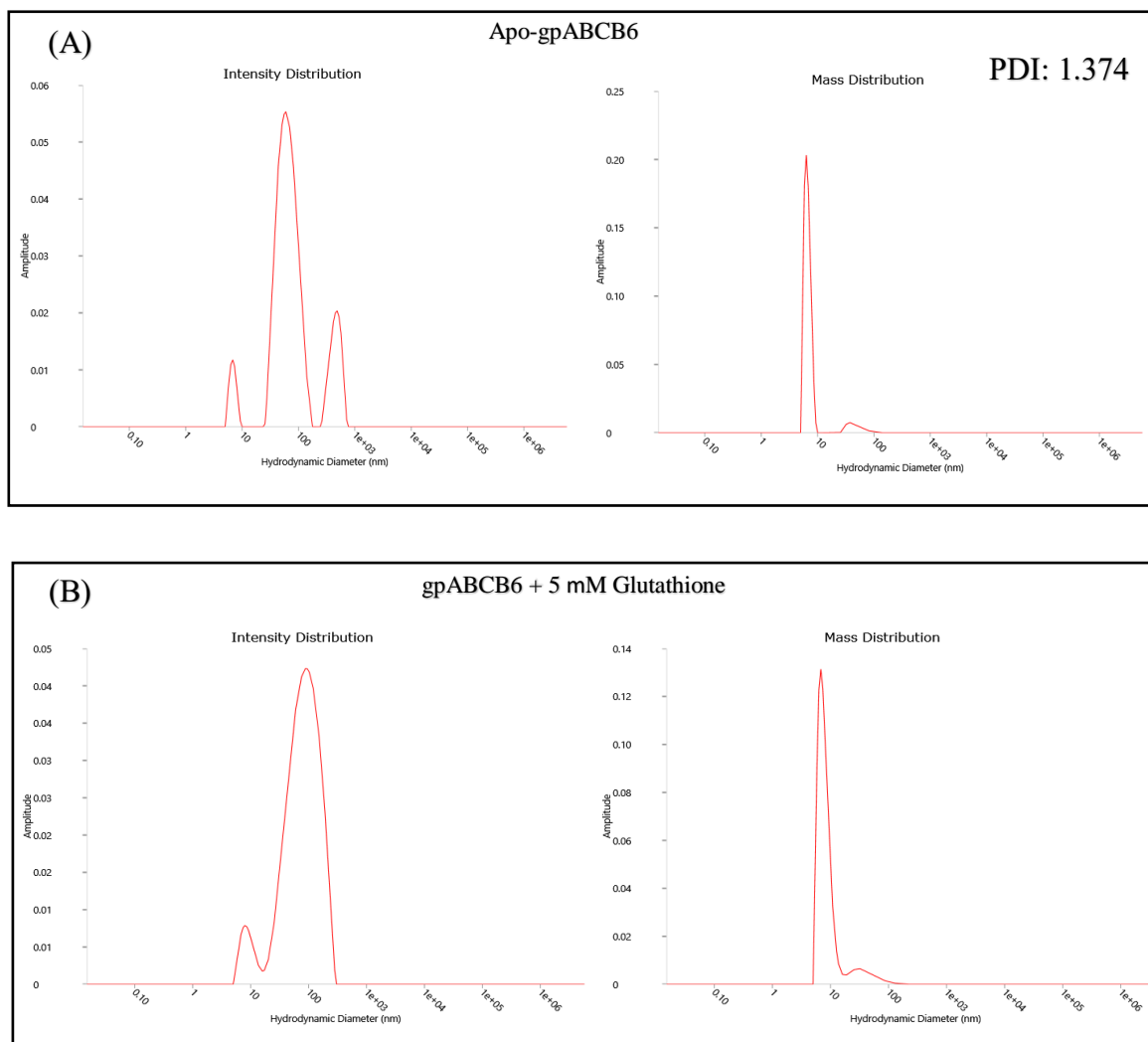
Figure 6.6: Dynamic light scattering profiles of gpABCB6 in different states (A) apo-gpABCB6, **(B)** Substrate-bound gpABCB6. The intensity distribution plot is displayed in the left panel while the mass distribution is displayed in the right panel. In apo-states three populations with hydrodynamic diameters of ~6 nm-10 nm, ~100 nm and >100nm were observed; while in the presence of substrate two populations were observed. The monomer appears to be predominant in the sample as judge by the mass distribution plots. **PDI:** Polydispersity Index

(A) Apo-mABCB5   PDI: 0.8655

(B) mABCB5 + 10 μM tariquidar
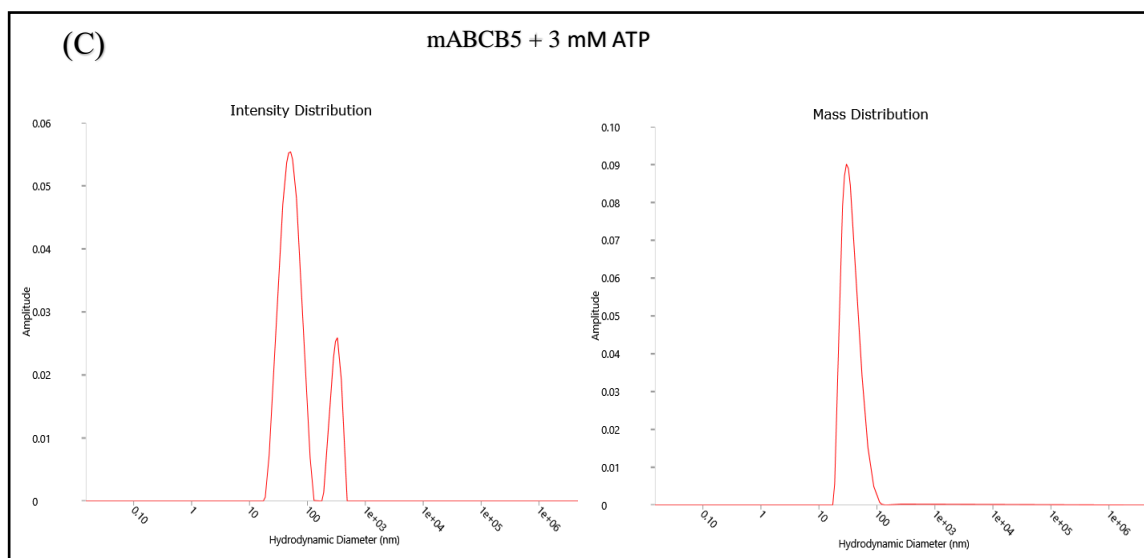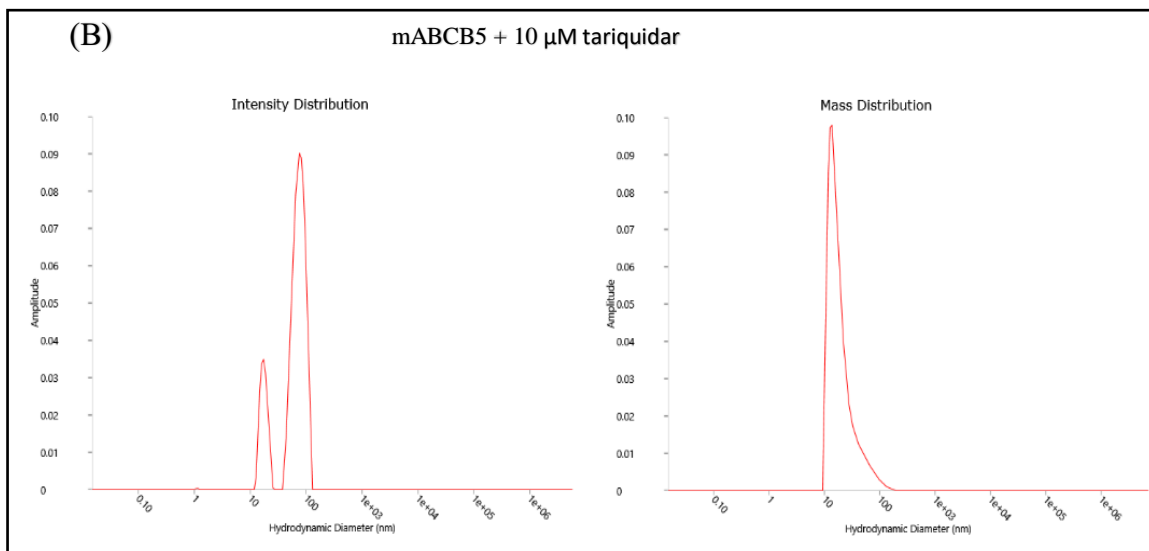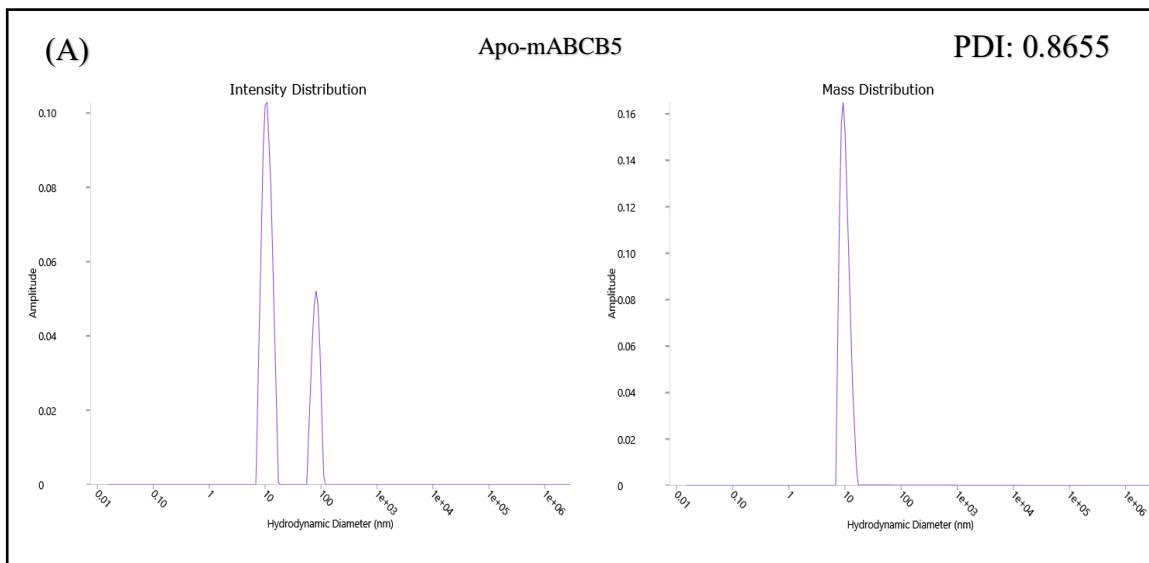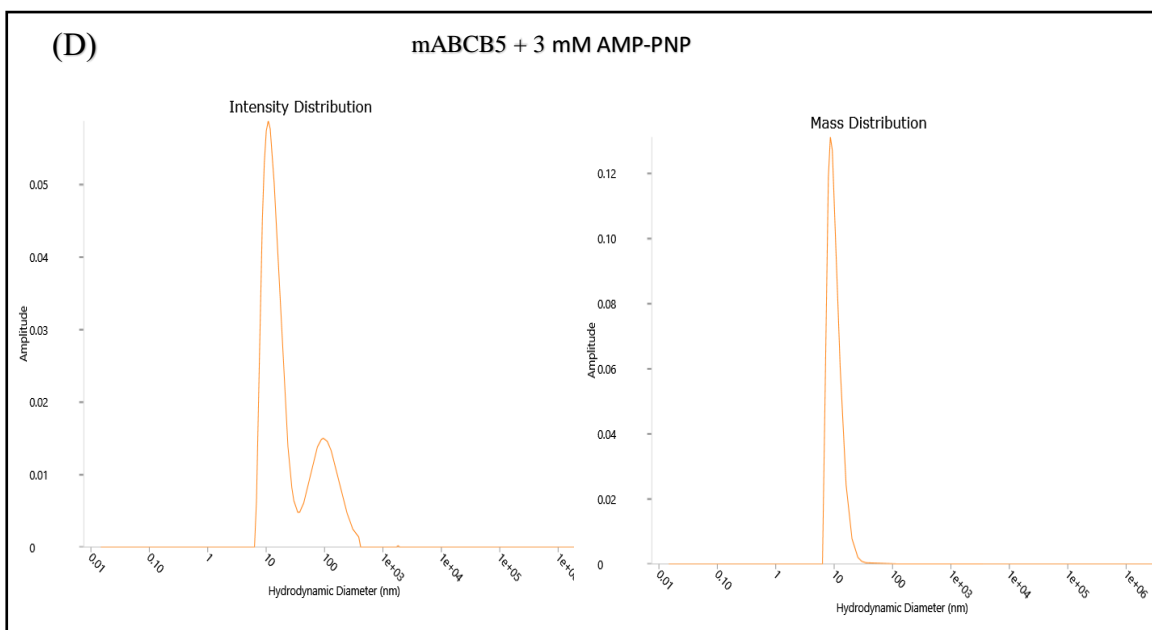
(C) mABCB5 + 3 mM ATP

Figure 6.7: Dynamic light scattering (DLS) profiles of mABCB5 in different states. A) apo-mABCB5 **B)** mABCB5 in the presence of 10 μM tariquidar **C)** mABCB5 in the presence of 3 mM ATP **D)** mABCB5 in the presence of 3 mM AMP-PNP. The intensity distribution plot is displayed in the left panel while the mass distribution is displayed in the right panel. A bimodal intensity distribution shows two distinct populations with hydrodynamic diameters of ~ 13 nm and > 100 nm representing monomers and oligomers/ aggregates, respectively. The monomer appears to be predominant in the sample as judge by the mass distribution plot. **PDI:** Polydispersity Index.
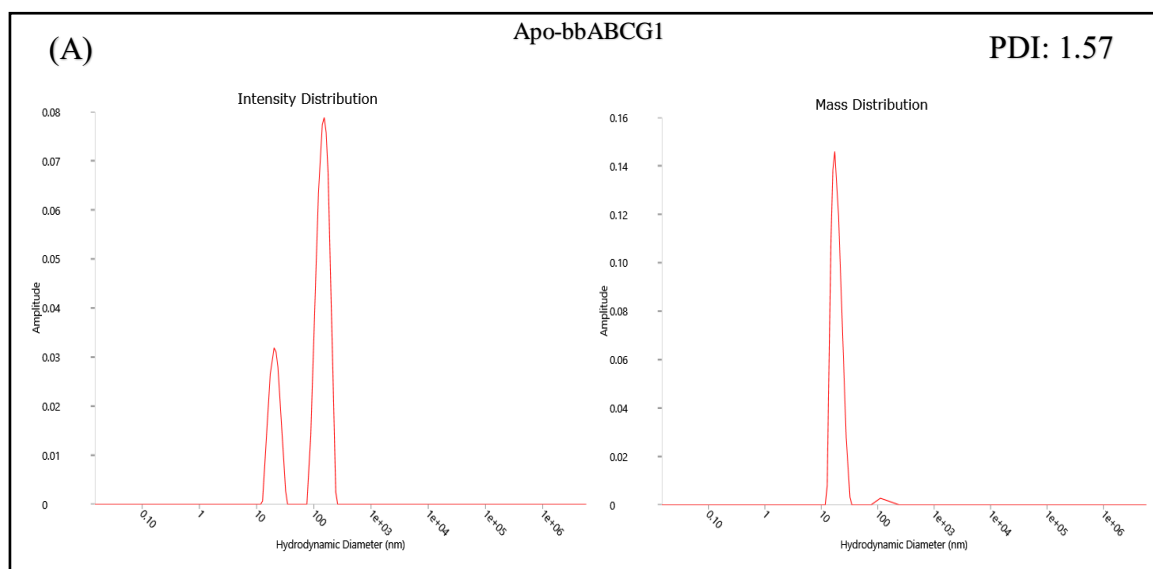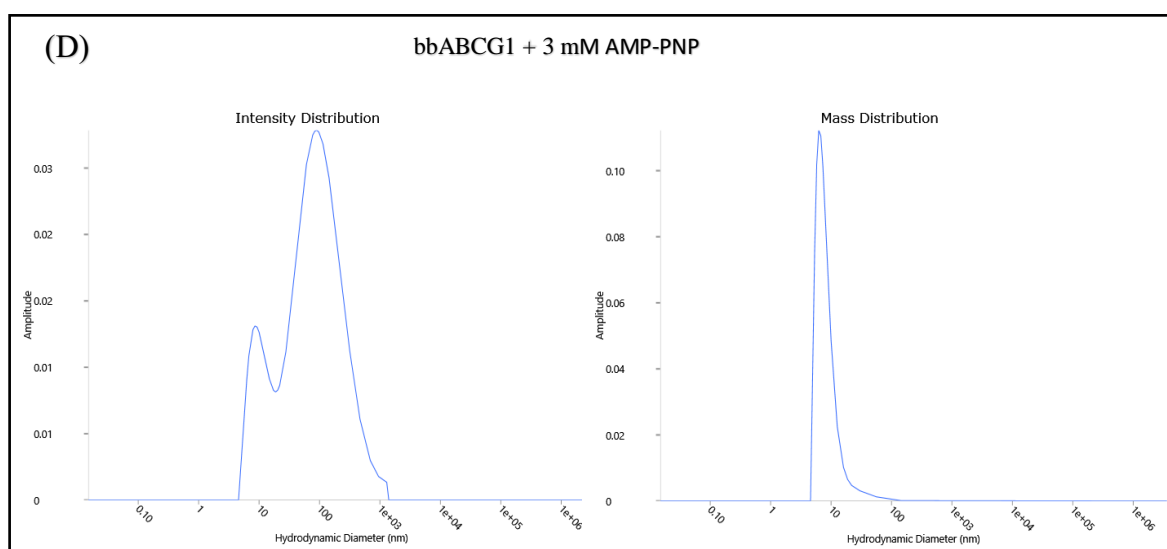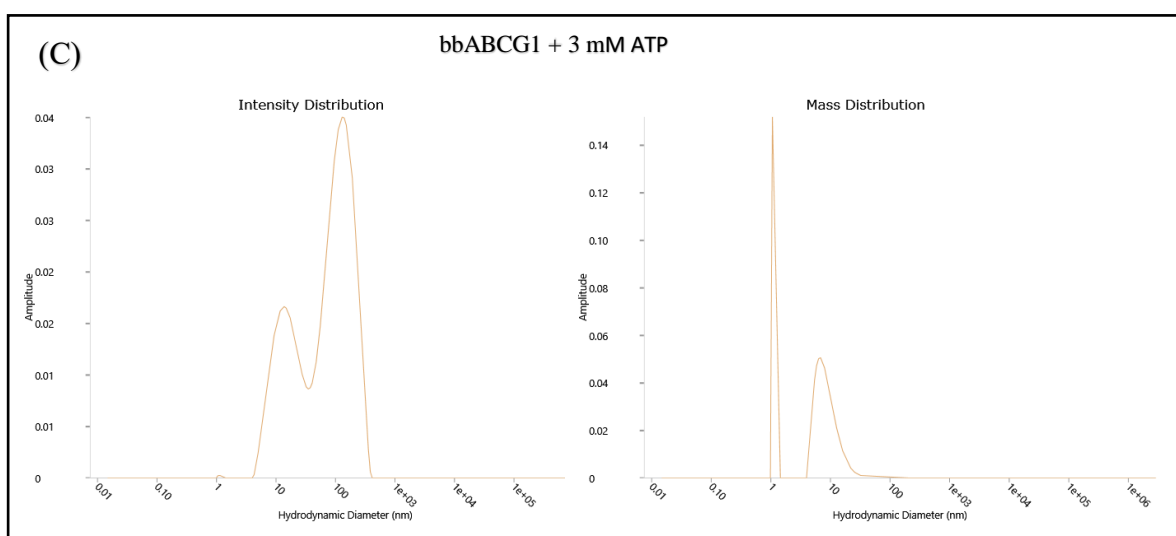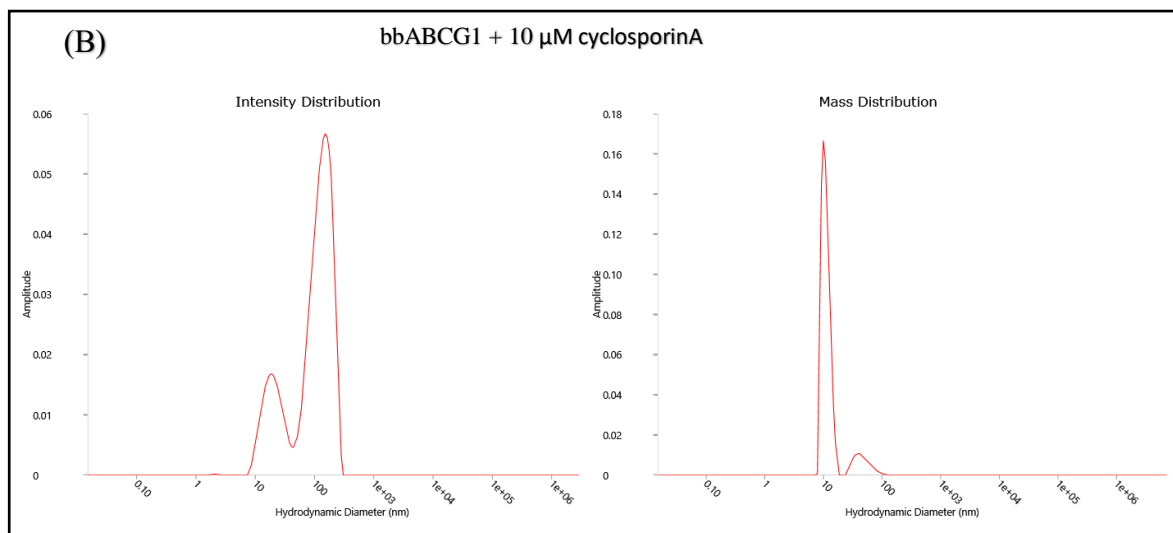
Figure 6.8: Dynamic light scattering (DLS) profiles of bbABCG1 in different states. A) apo-mABCB5 **B)** mABCB5 in the presence of substrate **C)** mABCB5 in the presence of 3 mM ATP **D)** mABCB5 in the presence of 3 mM AMP-PNP. **PDI:** Polydispersity Index

(A) Apo-mABCG4   PDI: 0.20

Intensity Distribution   Mass Distribution

(B) mABCG4 + 10 µM cyclosporin A

Intensity Distribution   Mass Distribution

(C) mABCG4 + 3 mM ATP

Intensity Distribution   Mass Distribution
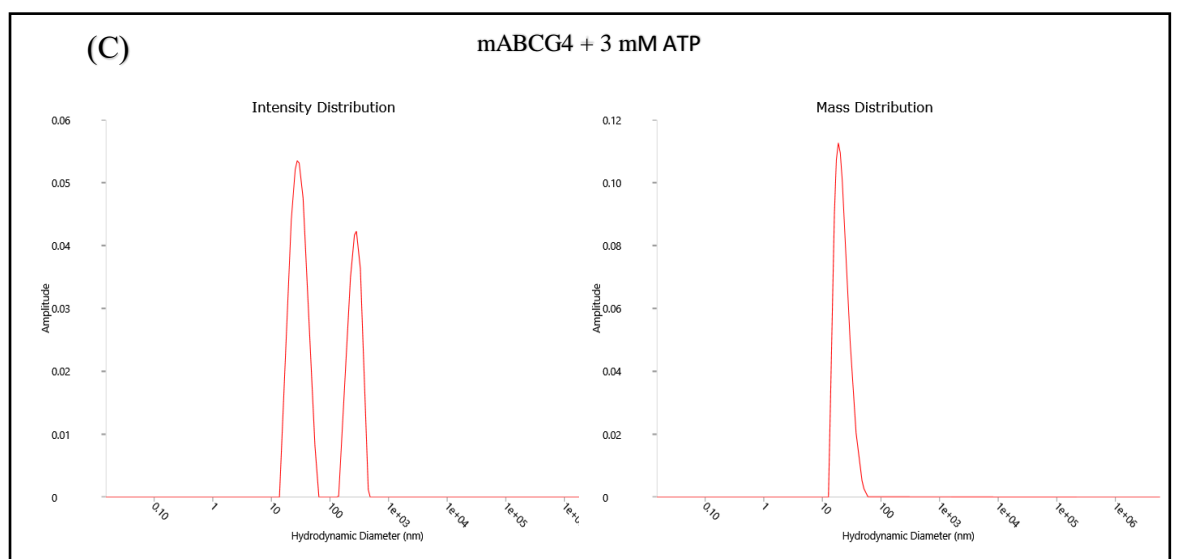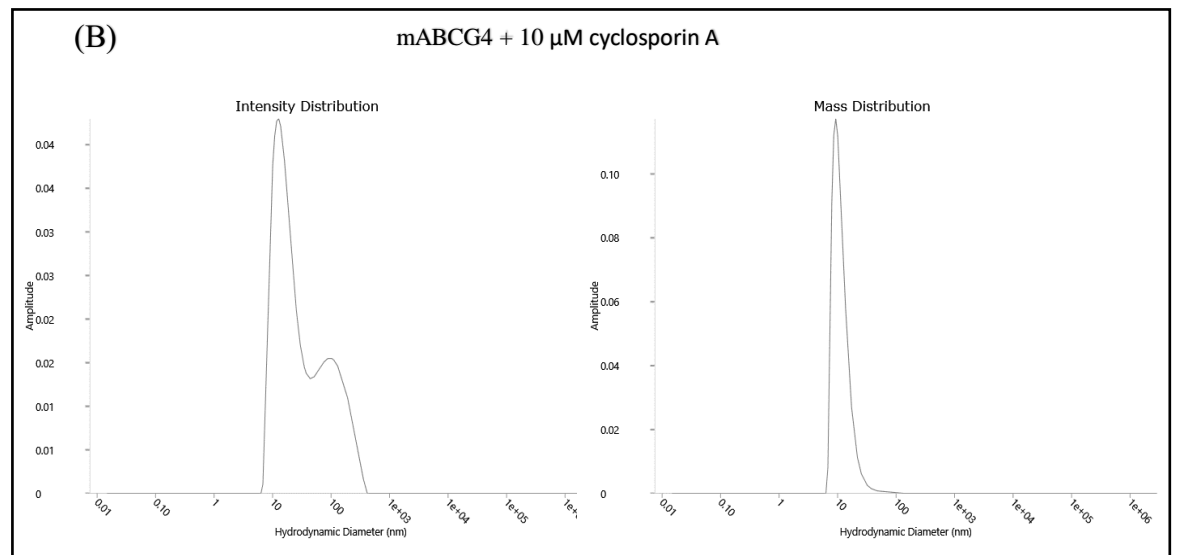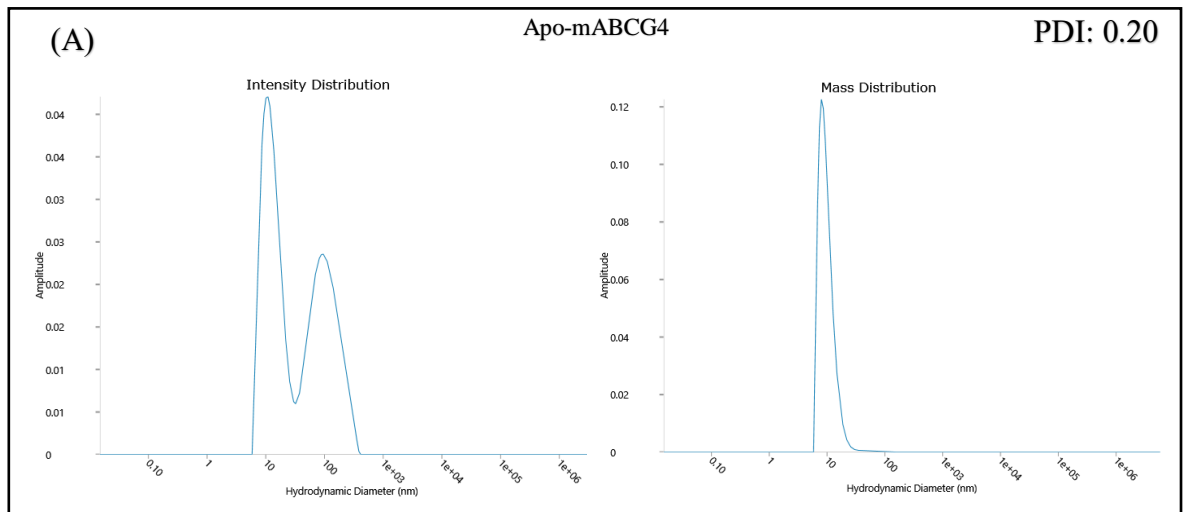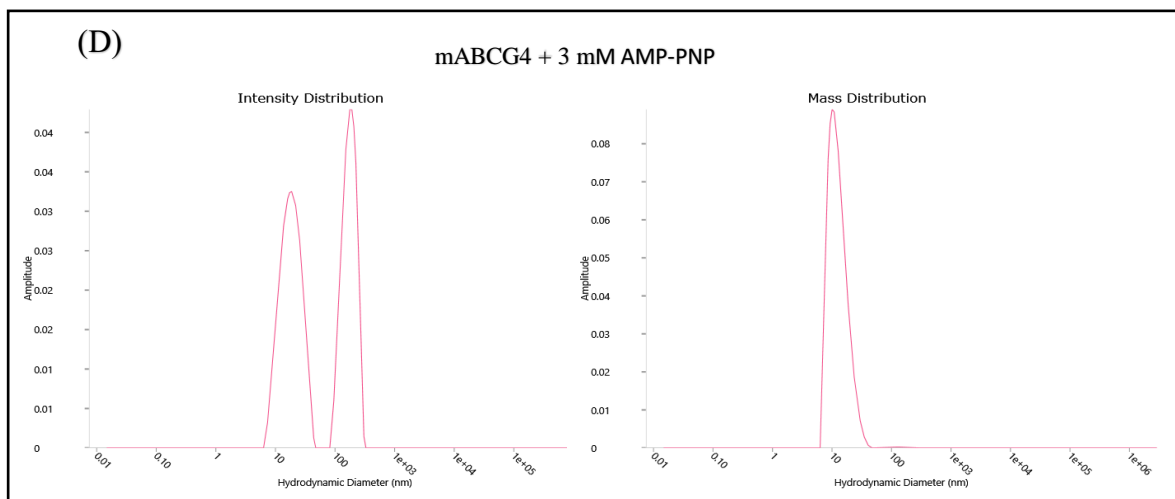
Figure 6.9: Dynamic light scattering (DLS) profiles of mABCG4 in different states. A) apo- mABCG4 **B)** mABCG4 in the presence of substrate **C)** mABCG4 in the presence of 3 mM ATP **D)** mABCG4 in the presence of 3 mM AMP-PNP. The intensity distribution plot is displayed in the left panel while the mass distribution is displayed in the right panel. A bimodal intensity distribution shows two distinct populations with hydrodynamic diameters of ~ 13 nm and > 100nm representing monomers and oligomers/ aggregates, respectively. The monomer appears to be predominant in the sample as judge by the mass distribution plots. **PDI:** Polydispersity Index

6.5 Thermal stability

Thermal stability is one of the protein characteristics that may have an impact on the quality of protein preparations and is considered a key factor when doing 3D crystallization trials (Carpenter et al., 2008, Malawski et al., 2006, Dupeux et al., 2011). The majority of the known structures of integral membrane proteins were found to be proteins with high stability (Sonoda et al., 2011). Overall, investigating the thermal stability can help in assessing the quality of purified proteins and shed additional light on how they behave. Differential scanning calorimetry (DSC), circular dichroism, and other assays that are frequently used to examine protein thermal stability exist, however not all of them can be adapted to membrane proteins. (Hong et al., 2009). Most of the assays require a relatively large amount of protein usually on a milligram scale which is a challenge for poor-yielding membrane proteins. For most of the thermal stability assays, detergent is also a problem. It is impractical to use techniques like circular dichroism that rely on the secondary structure because the detergent can mask the thermal unfolding of the protein and retain the secondary structure level of membrane proteins. (Kean et al., 2008). The thermal stability of mABCB5, gpABCB6,

bbABCG1, and mABCG4 were thus determined using three techniques that were considered to be compatible with membrane proteins and detergents as follows:

1- Using intrinsic protein fluorescence (IPF), tryptophan fluorescence is primarily monitored (Garidel et al., 2008, Hannemann et al., 2002, Ghisaidoobe and Chung, 2014, Rabbani et al., 2014, Cai et al., 2009, Kotov et al., 2019). The unfolding transition temperature T_m and the aggregation temperature Tagg of the four purified proteins, mABCB5, gpABCB6, bbABCG1, and mABCG4, were investigated and compared to determine the thermal stability of each protein.

2- By CPM assay (Alexandrov et al., 2008). This assay needs less protein and can be performed with 1-10 μg pure protein This assay has been successfully used to determine the thermal stability of membrane proteins including ABCA4, CFTR, and also P-gp (Pollock et al., 2014b, Meng et al., 2017). This CPM test was used in two independent experiments. In the first experiment, the thermal stability of all isolated proteins—aside from gpABCB6—was examined in two states *i.e.* the apo state and in the presence of nucleotides. The second experiment was carried out on pure mABCG4 to investigate ligands that stabilize the mABCG4.

3- Membrane thermal shift assay (Ashok et al., 2015). This assay does not require purified protein and can be done on membranes/vesicles. This assay has been successfully used to identify ligands for ABCG2 stability (Gose et al., 2020). In this study, this assay was employed to explore the effect of tyrosine kinase inhibitors (TKIs) on the stability of mABCG4.

6.5.1 Experiment 1: Tm and Tagg

The thermal stabilities of these four target proteins were determined by measuring melting temperature, T_m monitored by Intrinsic Protein Fluorescence (IPF), and the onset Tagg monitored by SLS (266 nm and 473 nm) using the UNcle instrument (Unchained Labs Inc.) (see Methods, section 2.21.4). Tm denotes the mid-point temperature at which the native protein conformation changes into a fully unfolded configuration; while the Tagg temperature corresponds to the point at which the presence of insoluble aggregates causes a noticeable change in the protein solution's light scattering to be detected. The thermal stability of membrane proteins has been assessed by intrinsic tryptophan fluorescence recorded simultaneously with SLS (Mastrotto et al., 2018, Cai et al., 2009, Ghisaidoobe and Chung, 2014).

9 µL of each protein sample of 10 µg concentration were loaded in duplicate in the Uncle capillary cassette and run with a thermal ramp from 20 °C to 95 °C at a ramp rate of 1 °C per minute (see section 2.21.4). To calculate the melting temperature, T_m, which is the midpoint of the unfolding process, the fluorescence from intrinsic aromatic amino acid residues (mainly tryptophan amino acid) excited by the excitation wavelength of 266 nm was determined, and T_m was calculated for each protein sample using the barycentric mean (BCM) of fluorescence intensity by the Uncle program. As shown in Table 6.6 gpABCB6, bbABCG1 and mABCG4 showed similar melting temperatures (T_m) 37.3 ± 0.8 °C, 36.5 ± 0.5 °C, and 36.8 ± 1.8 °C respectively, while the T_m of mABCB5 is 41.3 ± 6.3 °C suggesting that the mABCB5 is more stable than the other three proteins that had almost similar thermal stability. Moreover, as the Tagg 266 of mABCB5, gpABCB6, bbABCG1 and mABCG4 is 51.4 ± 0.8 °C, 41 ± 0.6 °C, 52.8 ± 3.6 °C, 36.5 ± 0.0 °C respectively, whereas, Tagg 473 of mABCB5, gpABCB6, bbABCG1 and mABCG4 is 50.6 ± 13 °C, 53.6 ± 6.0 °C, 47.1 ± 10.6 °C and 41.6 ± 2.4 °C (Table 6.6). The Tagg temperature is determined by measuring the point at which the presence of insoluble aggregates causes a discernible change in the protein solution's light scattering (Mastrotto et al., 2018). These numbers also suggest that mABCB5 is more stable.

At 266 nm, the mABCB5 and bbABCG1 exhibited a low level of scattered intensity and showed almost no aggregation at the start of a thermal ramp with a slight and gradual increase starting at around 60 °C. The SLS 266 nm profile of gpABCB6 showed a sudden and sharp increase at 40 °C, reached a maximum scattering intensity, and then had a decrease in signal intensity between 80–95 °C (Figure 6.10 A). Since SLS is sensitive to apparent molecular weight, the sharp increase in SLS at the start of the transition suggests that aggregation starts to happen at that temperature. Afterward at 80°C the decline, suggests that precipitation and gravitational clearing of aggregated material occurred *via* settlement. mABCG4 had a higher initial SLS266 intensity as compared to all other proteins tested (Figure 6.10 A). SLS measurements at 473 nm of mABCG4 painted a similar picture. Aggregation was not observed for mABCG4, which is consistent with the 266 nm measurements, and the lower sensitivity of the 473 nm read. The SLS 473 nm profiles were consistent with SLS 266 nm profiles *i.e.* mABCB5 and bbABCG1 had aggregation onsets at 60 °C; while gpABCB6 at around 40 °C (Figure 6.10 B). It can be difficult to make strong conclusions about the absolute amount and nature of aggregates based on the melting temperature when comparing proteins to one another but SLS measurements in isolation, suggest that mABCG4 is the more stable protein as they seem to have low levels of

aggregation or no aggregates at all and thermal stability ranking of these four proteins based on SLS measurements is as **mABCG4 > mABCB5/bbABCG1 > gpABCB6**.
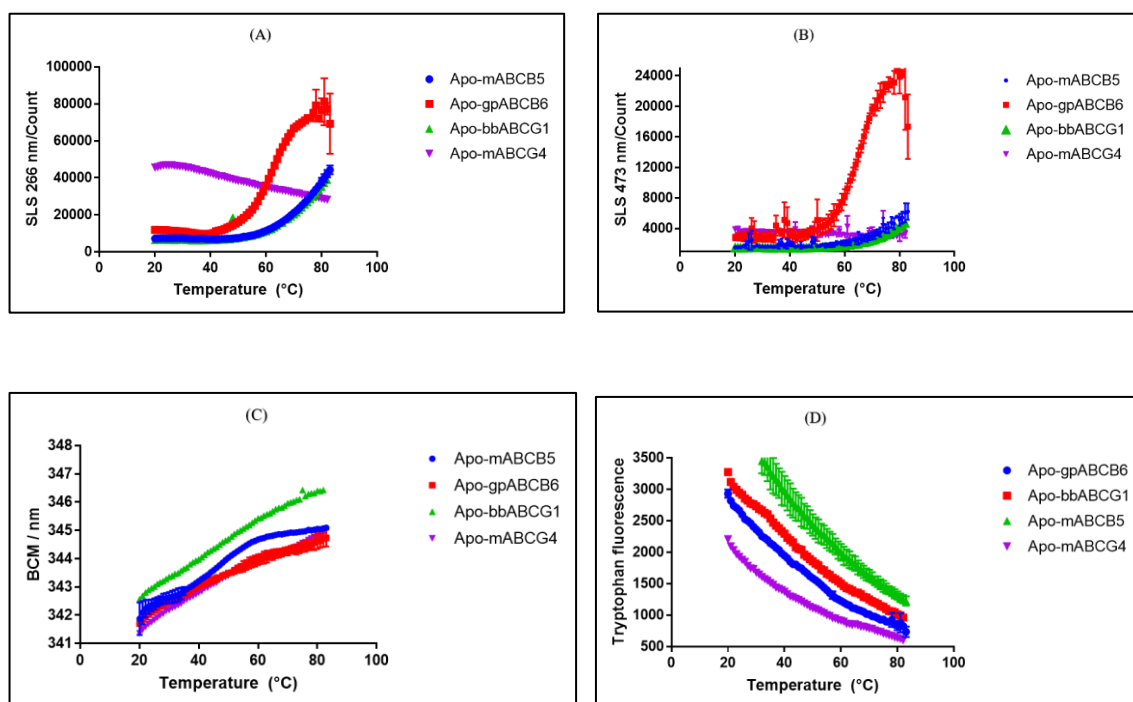
Figure 6.10: Tₘ and Tagg characterisation of mABCB5, gpABCB6, bbABCG1 and mABCG4 A) Aggregation curves overlaid for all of these proteins as measured by scattering intensity at 266 nm over a thermal ramp. **B)** Aggregation curves overlaid for all of these proteins as measured by scattering intensity at 473 nm over a thermal ramp. **C)** BCM melting curves. **D)** Tryptophan fluorescence intensity at 325 nm is plotted against temperature. The fluorescence was quenched as temperature increased.

Table 6.6: Summary of stability, aggregation and sizing data of mABCB5, gpABCB6, bbABCG1 and mABCG4 proteins in apo-state of Tₘ and Tagg experiment. Tm: Melting temperature; **Tagg:** Onset of aggregation temperature

	mABCB5	gpABCB6	bbABCG1	mABCG4
Tm (°C)	41.3 ± 6.3	37.3 ± 0.8	36.5 ± 0.5	36.8 ± 1.5
Tagg (°C) at 266 nm	51.4 ± 0.8	41 ± 0.6	52.8 ± 3.6	36.5 ± 0.0
Tagg (°C) at 473 nm	50.6 ± 13	53.6 ± 6	47.1 ± 10.6	41.6 ± 2.4
Initial SLS at 266 nm (counts x104)	0.6	1.19	0.6	4.5
Final SLS at 266 nm (counts x104)	4.45	6.9	3.7	2.8
Initial SLS at 473 nm (counts)	1.5	2.8	1.5	3.9
Final SLS at 473 nm (counts)	6.06	1.7	4.7	3.2

6.5.2 Experiment 02: CPM assay

ABC transporter conformational changes are triggered by nucleotides (Campbell et al., 2004, Wen and Tajkhorshid, 2008), and their binding may have significant influences on protein thermal stability (Celej et al., 2003). Thus, the thermal stability of mABCB5, bbABCG1, and mABCG4 was examined, comparing the apo-state and states in the presence of nucleotides. CPM assay was not conducted on gpABCB6 because of a lack of purified protein. The assay adopted in this study was previously optimised to use less protein, which is more suitable for membrane protein studies (Alexandrov et al., 2008).

In this experiment, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM), a dye specific to thiols, was used to monitor the unfolding of proteins as the temperature increases (see section 2.21.4). When CPM dye bounds to exposed cysteine residues in proteins, the dye fluoresces intensely, but in an aqueous solution, it hardly emits a signal (Alexandrov et al., 2008). Hence, the CPM fluorescent signal is specific to cysteine binding during the thermal unfolding process. Because the dye is so sensitive, only a minimal amount of protein is needed for this assay. Other advantages include a good signal-to-noise ratio and compatibility with a variety of detergents. Therefore, it offers many benefits that make it a suitable fluorescent dye that can be utilised in studies of membrane protein thermal stability (Alexandrov et al., 2008, Lee et al., 2013). The CPM has an excitation peak at ~384 nm, but for these experiments, it was excited using the instrument's 267 nm UV laser. Light output was continually measured between 200 nm and 600 nm, enabling the detection of both CPM and tryptophan fluorescence as well as a light scattering to be recorded. This is crucial because any significant change caused by protein aggregation or air bubble production in the capillaries could have an impact on its interpretation. The CPM fluorescence was detected between 475 nm and 490 nm and this region of the emission spectrum was integrated as a measure of CPM fluorescence intensity. Tryptophan fluorescence was detected between 323 nm and 350 nm and integrated before being ratioed with the CPM fluorescence signal. This allows deconvolution and adjustment for thermal quenching of the CPM fluorescence as well as any variation between duplicate experiments caused by sample loading, according to Xin Meng's thesis (The University of Manchester), where this approach was created and optimised. The protein was heated to 95 °C with a 2 °C increase per step and a heating rate of 1 °C/min. Raw data were firstly corrected by subtraction of data for buffer controls for each protein. This was also important to eliminate background effects. Due to the slight variations in each sample's capillary optics, each

dataset in a triplicate experiment was first normalised to the greatest value of each dataset, with 0% being designated as the signal of the buffer control with no protein present. The data is, therefore, presented here as a plot of relative thermal unfolding percentage against temperature increase, with the triplicate, normalised experiments averaged and then mean and standard deviation (SD) of fluorescence at each temperature plotted. The data was analysed for melting temperature (T_m) using non-linear regression of sigmoidal curve fit in GraphPad Prism 7. The difference was considered to be significant when P-value < 0.05 *via* unpaired, two-tailed student's T-test.

Figure 6.11 presents unfolding profiles of mABCB5 in the presence of different nucleotides in comparison with the apo-state. ADP+P_i and AMP-PNP bound states data were too noisy because the nucleotides strongly absorb the 267 nm laser light (Thonghin, 2019). Apo-mABCB5 thermal unfolding transition pattern illustrating that the protein underwent unfolding between 40 °C and 70 °C. Apo-mABCB5 exhibited a T_m 52.45 ± 0.38 °C, which is higher than the previously reported value of 41.3 °C (determined by intrinsic tryptophan fluorescence measurements. Interestingly, the presence of nucleotide significantly increased the T_m of mABCB5 up to 2 degrees from 52.45°C to 54 ± 1.21 °C for ADP-V_i bound-state, while in AMP-PNP-bound state T_m increased by 6 degrees *i.e.* 58 ± 0.87 °C. There seemed significant difference between AMP-PNP and ADP-V_i conditions in terms of T_m. This implies that nucleotide plays a significant role in mABCB5 stabilisation. This is interesting because thermal stability has been recognised to have an inverse relationship with conformational flexibility (Celej et al., 2003, Tsai et al., 2001) meaning that the protein may gain structural rigidity upon nucleotide binding. Apo-bbABCG1 exhibited a T_m of 51± 0.36°C, which is significantly higher than the previously reported value of 36.5 °C (determined by intrinsic tryptophan fluorescence measurements) (Figure 6.12). The presence of nucleotide significantly increased the T_m of apo-bbABCG1 up to 16 degrees in ADP-V_i *i.e.* from 51 °C to 67.08 ± 2.03 °C, while in AMP-PNP-bound state T_m increased by 10 degrees which is 62 ± 3.14 °C. Apo-mABCG4 exhibited a T_m of 60.5 ± 0.34 °C which is higher than the previously reported value of 36.8 °C (determined by intrinsic tryptophan fluorescence measurements). The presence of nucleotide significantly increased the T_m of apo-mABCG4 up to 2 degrees from 60.5 °C to 62 ± 0.5 °C in ADP-Vi and 5 °C in AMP-PNP bound state from 60.5 °C to 65.3 ± 0.2 °C (Figure 6.13). There seems a significant difference in T_m of apo-states and AMP-PNP /ADP-V_i bound in all proteins (mABCB5, gpABCB6, mABCG4) and results were consistent in terms that the presence of nucleotides is increasing the T_m indicating that nucleotides have a stabilising impact. Although, there is

no published data on the thermal stabilities of these proteins yet, however, our data is consistent with the previously published data on P-gp and CFTR which demonstrated the stabilising effects of nucleotides (Meng, 2017, Thonghin et al., 2018b).

Figure 6.14 displays the comparison of mABCB5, bbABCG1, and mABCG4 in apo-state. CPM fluorescence plots depicts initial 40 % initial CPM labelling of mABCB5, which demonstrates that 40 % surface-exposed cysteine residues in mABCB5 at the 20 °C which increased upto 80 % as the temperature increased and then quenching occurred (Figure 6.14B). 18-20 % CPM initial labelling of mABCG4 and bbABCG1 respectively which demonstrates that 18-20 % surface-exposed cysteine residues in mABCG4 and bbABCG1at 20 °C which increased to 40 % in mABCG4 while 35 % in bbABCG1 (Figure 6.14B). The relative thermal unfolding patterns of mABCG4 and bbABCG1, which begin at roughly 40 °C, are identical. This is understandable given the 72 % sequence identity between mABCG4 and bbABCG1.

Table 6.7: Summary of Tm determined by CPM assay of mABCB5, bbABCG1 and mABCG4 in apo, ADP+V$_i$ and with AMP-PNP states.

	Apo-state	ADP+V$_i$	AMP-PNP
mABCB5	52.45 ± 0.38	54 ± 1.21	58 ± 0.87
bbABCG1	51± 0.36	67.08 ± 2.03	62 ± 3.14
mABCG4	60.5 ± 0.34	62 ± 0.5	65.3 ± 0.2

A)
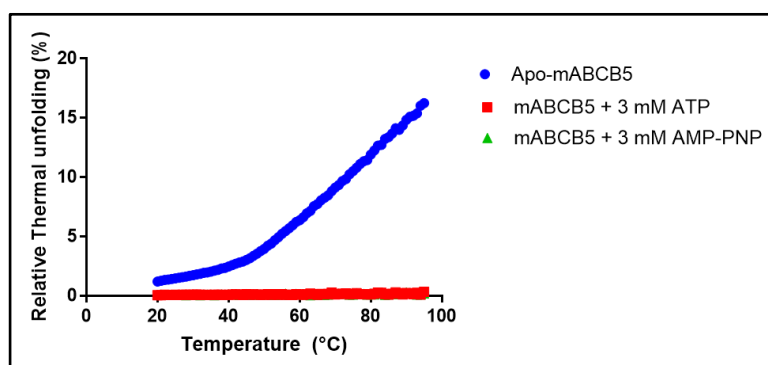
B)
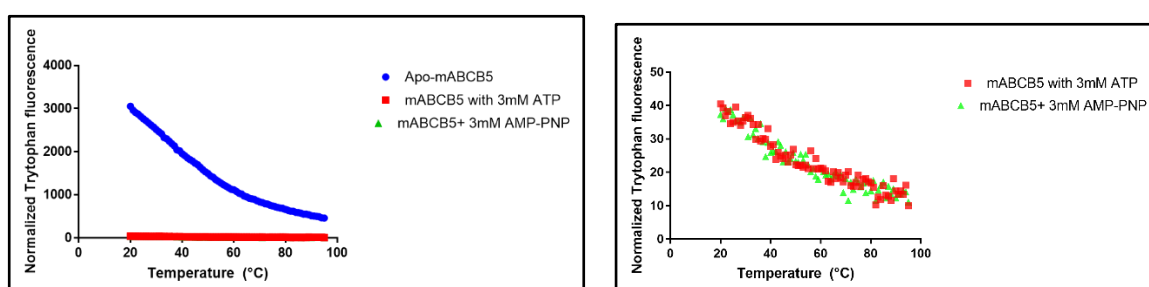
C)
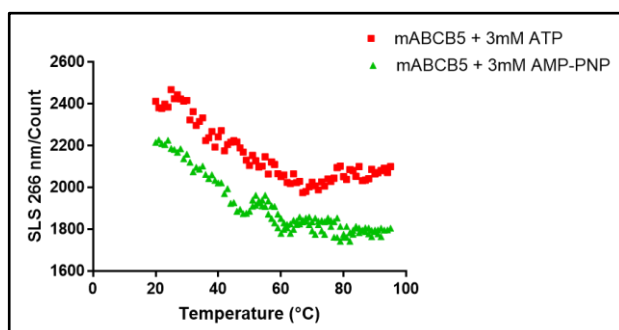
D)
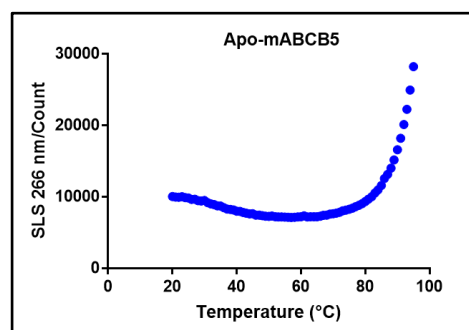
Figure 6.11: CPM Thermal stability of mABCB5 in different states. The thiol-specific dye (CPM), which binds to exposed cysteine residues on surfaces, was used in the experiment. **A)** Relative thermal unfolding data were obtained from ratioing CPM fluorescence data to intrinsic Tryptophan data. Melting temperature (T_m) given by the mid-point of the protein unfolding transition was calculated using GraphPad Prism 7 *via* sigmoidal nonlinear regression fitting. **B)** Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. Each data point represents mean ± SD of duplicate experiments. **C)** Aggregation curves of mABCB5 with ADP+ V_i and with AMP-PNP at static light scattering 266nm **D)** Aggregation curves of apo-mABCB5 at static light scattering 266 nm.

A)

B)

C)

D)

Figure 6.12: CPM Thermal stability of bbABCG1 in different states. The thiol-specific dye (CPM), which binds to exposed cysteine residues on surfaces, was used in the experiment. **A)** Relative thermal unfolding data were obtained from ratioing CPM fluorescence data to intrinsic Tryptophan data. Melting temperature (T_m) given by the mid-point of the protein unfolding transition was calculated using GraphPad Prism 7 *via* sigmoidal nonlinear regression fitting. **B)** Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. Each data point represents mean ± SD of duplicate experiments. **C)** Aggregation curves of bbABCG1 with ADP+ V_i and with AMP-PNP at static light scattering 266 nm. **D)** Aggregation curves of apo-bbABCG1 at static light scattering 266 nm.

A)
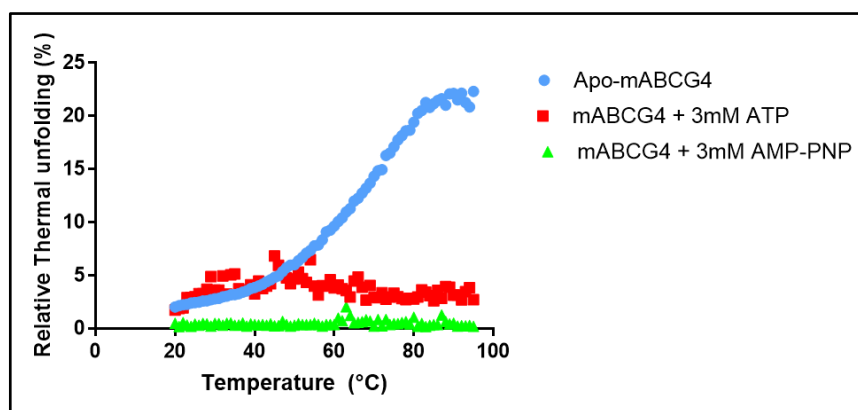
B)
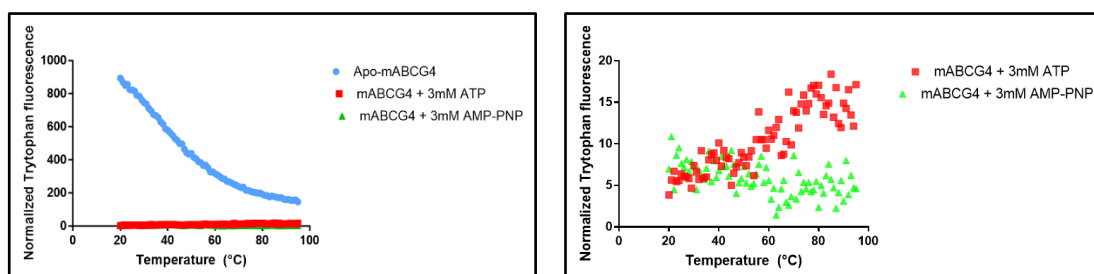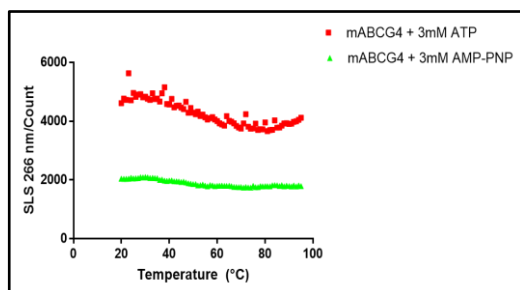
C) **D)**
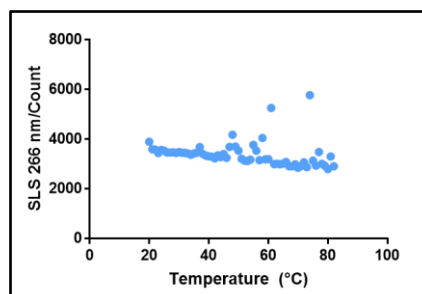
Figure 6.13: CPM Thermal stability of mABCG4 in different states. The thiol-specific dye (CPM), which binds to exposed cysteine residues on surfaces, was used in the experiment. **A)** Relative thermal unfolding data were obtained from ratioing CPM fluorescence data to intrinsic Tryptophan data. Melting temperature (T_m) given by the mid-point of the protein unfolding transition was calculated using GraphPad Prism 7 *via* sigmoidal nonlinear regression fitting. **B)** Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. Each data point represents mean ± SD of duplicate experiments. **C)** Aggregation curves of mABCG4 with ADP+ V_i at static light scattering 266 nm. D) Aggregation curves of apo-mABCG4 at static light scattering 266 nm.
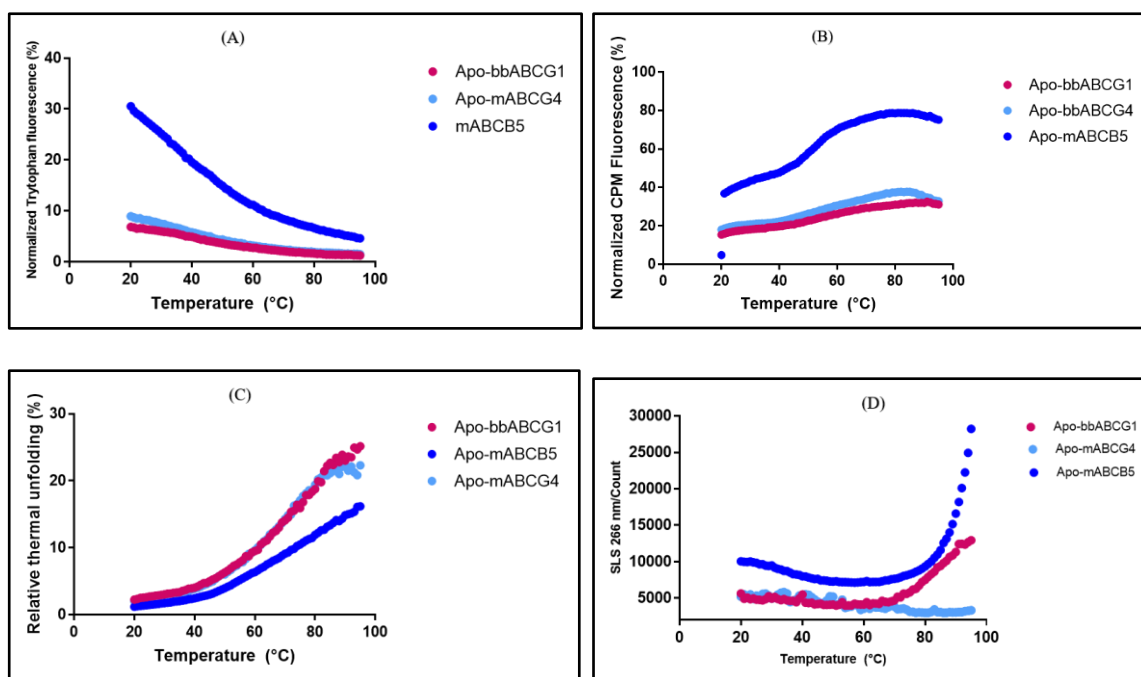
Figure 6.14: Comparison profiles of CPM Thermal stability of mABCB5, bbABCG1 and mABCG4 in apo-state. A) Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. **B)** Normalized CPM fluorescence data showing signal transition and distinctive thermal quenching. **C)** Relative thermal unfolding. **D)** Aggregation curves measured at static light scattering 266 nm.

6.5.3 Experiment 03: CPM thermal stability of mABCG4 in the prescence of subtrates

Purified mABCG4 was assessed further for thermal stability in the presence of different substrates. The assay was conducted using the UNcle machine *via* the CPM thermal stability assay protocol described earlier in section 6.5.2. The experiment was conducted three times. Raw data containing emission spectra were exported and then subtracted for buffer controls before analysis *via* GraphPad Prism 7. Unfolding data derived from CPM emission data were plotted against temperature and fitted with non-linear regression analysis assuming a sigmoidal curve fit (Figure 6.15). The midpoint of each curve was determined as the melting temperature (T_m). Tm difference was considered to be significant when P-value < 0.05 *via* unpaired, two-tailed student's T-test. The T_m of apo-mABCG4 was estimated at 63.01 ± 0.29 °C which is 3 degrees higher as compared to previous experiment 02. The purified protein utilised at higher concentrations in this experiment is one explanation that might be given for this. The T_m value increased by 14-15 °C in the presence of sphingolipids and cyclosporine A, and 10 °C in the presence of estriol (Figure 6.15). Interestingly, T_m decreased

significantly in the presence of cholesterol. One possible explanation for this could be partial solubility of cholesterol.
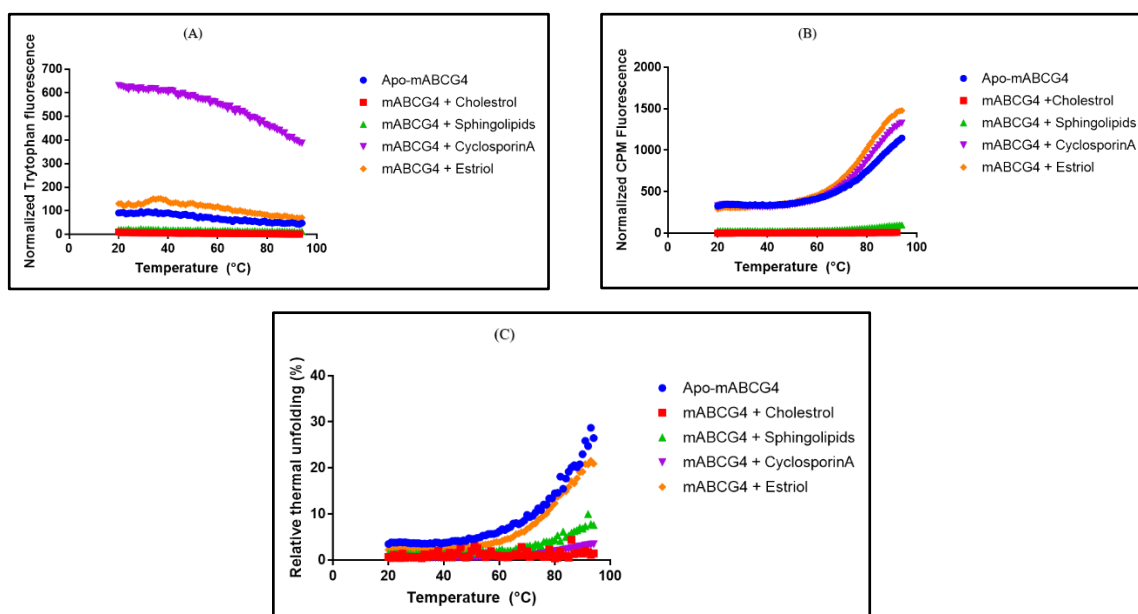
Figure 6.15: CPM Thermal stability of mABCG4 in the presence of known substrates. A) Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. **B)** Normalized CPM fluorescence data showing signal transition **C)** Relative thermal unfolding data were obtained from ratioing CPM fluorescence data to intrinsic Tryptophan data. Melting temperature (T_m) given by the mid-point of the protein unfolding transition was calculated using GraphPad Prism 7 *via* sigmoidal nonlinear regression fitting. B) Normalized Tryptophan fluorescence showing reduction of the signal along with an increasing temperature. Each data point represents mean ± SD of duplicate experiment.

Table 6.8: Summary of T_m of mABCG4 in the presence of substrates.

	$T_m(°C)$
Apo-mABCG4	63.01 ± 0.2939
mABCG4 + cholesterol	49.55 ± 2.333
mABCG4 + Sphingolipids	79.69 ± 0.3781
mABCG4 + cyclosporin A	78.09 ± 0.2534
mABCG4 + Estriol	71.25 ± 0.2109

6.5.3 Membrane thermal shift assay

Although CPM assay has high a throughput but it requires a pure membrane protein which is challenging to get in enough quantities. Because of this reason, in order to identify ligands that stabilise the ABCG4 protein, the membrane thermal shift assay was adapted, according to Ashok et al (Ashok et al., 2015). This assay can be done using microsomes/vesicles. The guiding principle is that, following heating, a target protein will unfold and precipitate,

whereas a protein that is coupled with a ligand will often need a higher temperature to unfold and precipitate (Reinhard et al., 2015, Molina et al., 2013, Martinez Molina and Nordlund, 2016, Jafari et al., 2014). There hasn't been any research into the stability of isolated full length mABCG4 to date. It could be argued that the improvement of thermal stability of full length protein will be crucial factor before doing any high-resolution structural studies. Additionally, it is possible that any substance that can improve the thermal stability of full-length mABCG4 could go into the crystallization trials.

mABCG4 membrane vesicles (5 mg/ mL, 80 µl final reaction volume) were heated in a PCR machine for 3 minutes at various temperatures (20-90 °C) to establish a thermal denaturation curve (see section 2.21.18) that revealed a T_m of 70 °C (Figure 6.16A). Based on the extrapolation from the thermal denaturation curve that was predicted to produce 99 % loss of protein in the supernatant, 70 °C was selected to assess the ability of ligands to thermally stabilize mABCG4. Membranes were incubated with forty-two different tyrosine kinase inhibitors at 10 µM concentration for 60 minutes at 37 °C and then heated at 70 °C for 3 minutes (see section 2.21.8). Immunoblot was imported to ImageJ for analysis where the bands were analysed for quantification of western blot signals. The signal intensity of the heated sample was normalized to the signal intensity of the control sample (mABCG4 heated at 70 °C without any drug) and the graph is plotted as normalized protein recovery in Graphpad Prism 7. Because of time limitation we performed only duplicate experiments with TKIs and were unable to calculate standard deviation. Results are displayed in Figure 6.16 Panel B, C and D. Results showed that Dovitinib, Tivozanib and Masitinib were most effective in stabilizing ABCG2 against thermal denaturation (Figure 6.16 B). On the other hand, linifanib, foretinib, cabozantinib, pazopanin, afatanib, ACL1717, alectinib had least stabilizing effect. Taken together, this study shows the ability of these tyrosine kinase inhibitors to thermally stabilize mABCG4 by the membrane thermal shift assay. Future research can be done to find out how many degrees of thermal stability can be achieved by running an experiment at temperatures between 70 and 80 degrees.
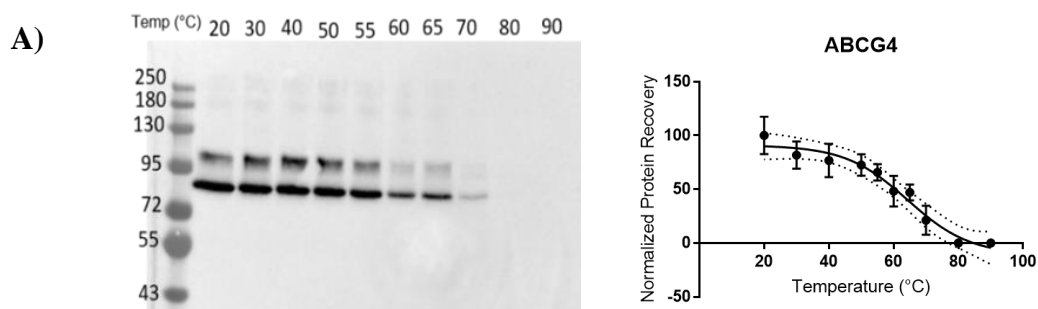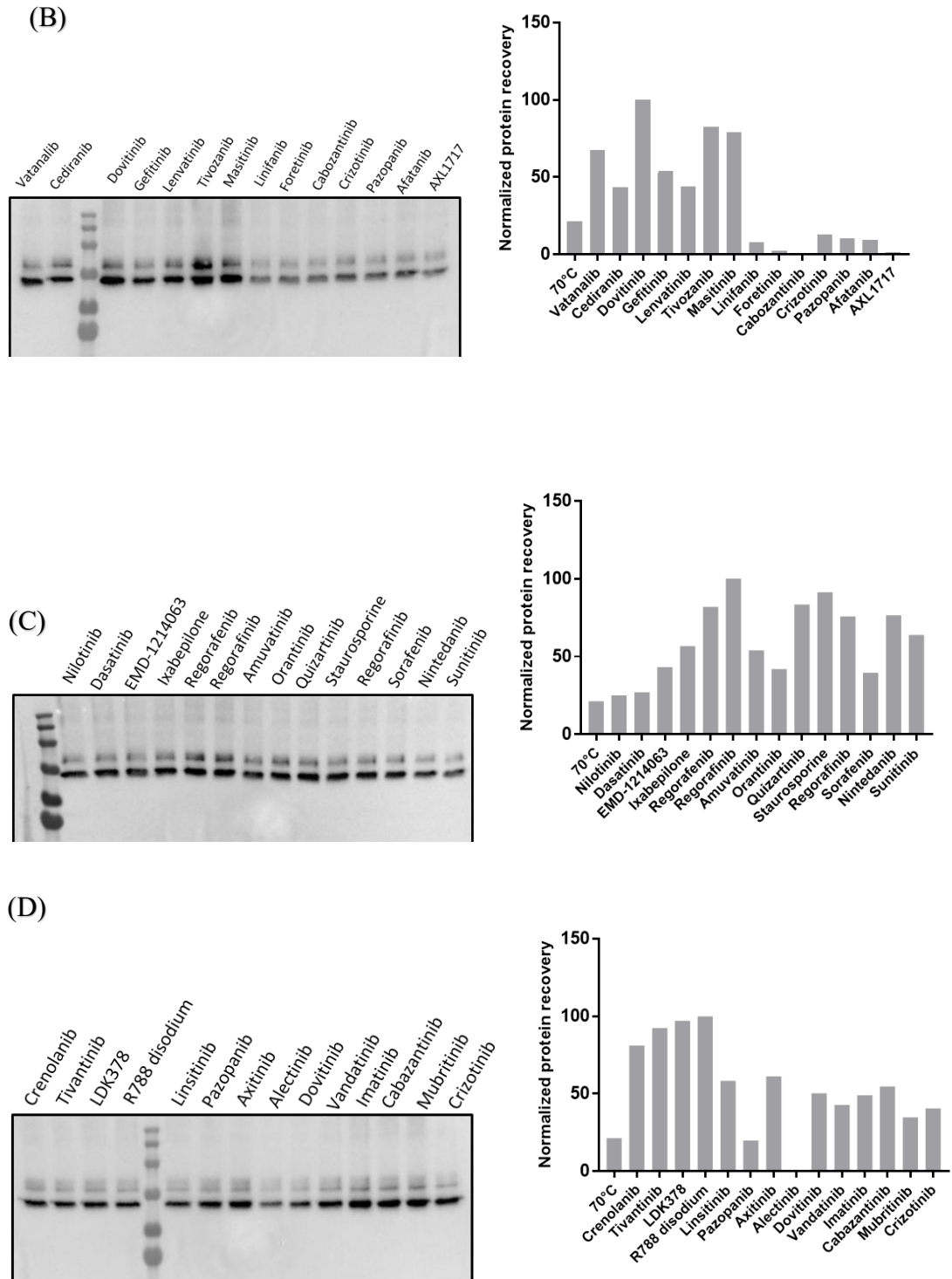
Figure 6.16: Membrane thermal shift assay of mABCG4. DDM-solubilized mABCG4 microsomes were incubated at each temperature for 3 minutes. **(A)** Left panel shows the western blot of mABCG4 which disappeared at 70 °C. Right Panel shows the graphical representation of quantification of protein recovery analyzed by imageJ after thermal unfolding. **B, C, and D**) Left panel shows the Western blot of binding affinity of Tyrosine kinase inhibitors (TKIs) to mABCG4 at 70 °C; while right panel shows the graphical representation of binding affinity of Tyrosine kinase inhibitors (TKIs) to mABCG4.

6.6 Negative stain electron microscopy

EM is a useful technique to study the structure of biological macromolecules, including membrane proteins (Frank, 2006, Rubinstein, 2007, Ohi et al., 2004, Oliver, 1973). Electron micrographs can provide qualitative information of the homogeneity and monodispersity of protein (Sjollema et al., 2012). For instance, EM was successfully employed to examine the CFTR aggregation state following purification from mammalian cells before further structural characterisation (Zhang et al., 2011, Zhang et al., 2009a, Zhou and Cross, 2013). The sample preparation for this approach is rather simple and quick. The minimum required protein concentration for negative stain EM is 10 μg/ml. This technique involves embedding a specimen in a layer of dried heavy metal solution that greatly improves the specimen contrast. For typical negative staining, a number of heavy metal stains are available such as ammonium molybdate, tungstate stains, and uranyl stains (Ohi et al., 2004). In this experiment we have used uranyl acetate because research has shown that uranyl acetate solution fixes protein structure on a millisecond time frame and having a finer grain size results in improved staining of the specimen (Zhao and Craig, 2003).

After a two-step purification process, we used negative staining to examine the mABCG4, gpABCB5, bbABCG1, and mABCG4 proteins to assess homogeneity and particle size. It is a best practice to first do a negative stain study when working with a structurally uncharacterized molecule before making any attempt to work with crystallography/ vitrified specimens. Each target protein was applied on a 300-mesh carbon-coated copper grid at a concentration of 25 μg/ml, and then rinsed with water to remove extra detergent and other solutes before being stained with uranyl acetate stain. A transmission electron microscope was used to view the grids. Protein is seen in micrographs as white particles on a grey background.

In figures 6.17–6.20, two micrographs for each target protein are displayed. Results for each target protein almost similar except mABCB5. mABCB5 micrographs reveal heterogeneity which is consistent with SDS gels (section 5.3). Cumulative results of gpABCB6, bbABCG1, and mABCG4 are written in this section. Micrographs showed nearly homogeneous and evenly distributed purified protein particles. The majority of particles are ~10 nm in size, which is consistent with the findings of dynamic light scattering (section 6.4). These positive findings indicated that purified proteins are appropriate for cryo-EM or crystallography. However, these techniques demand high protein concentrations, and as of yet, we are unsure how these protein particles would function at high concentration. Previous

studies has shown concentrating a protein sometimes is problematic because of degradation propensity of proteins (Ma et al., 2017). Besides, few particles of 20-30 nm and aggregates ~ 100 nm are also seen in micrographs which might represent protein oligomers or protein-detergent complexes. Also, staining buildup, which is typical of negative staining, can be seen on micrographs (Ohi et al., 2004).

(A)

(B)
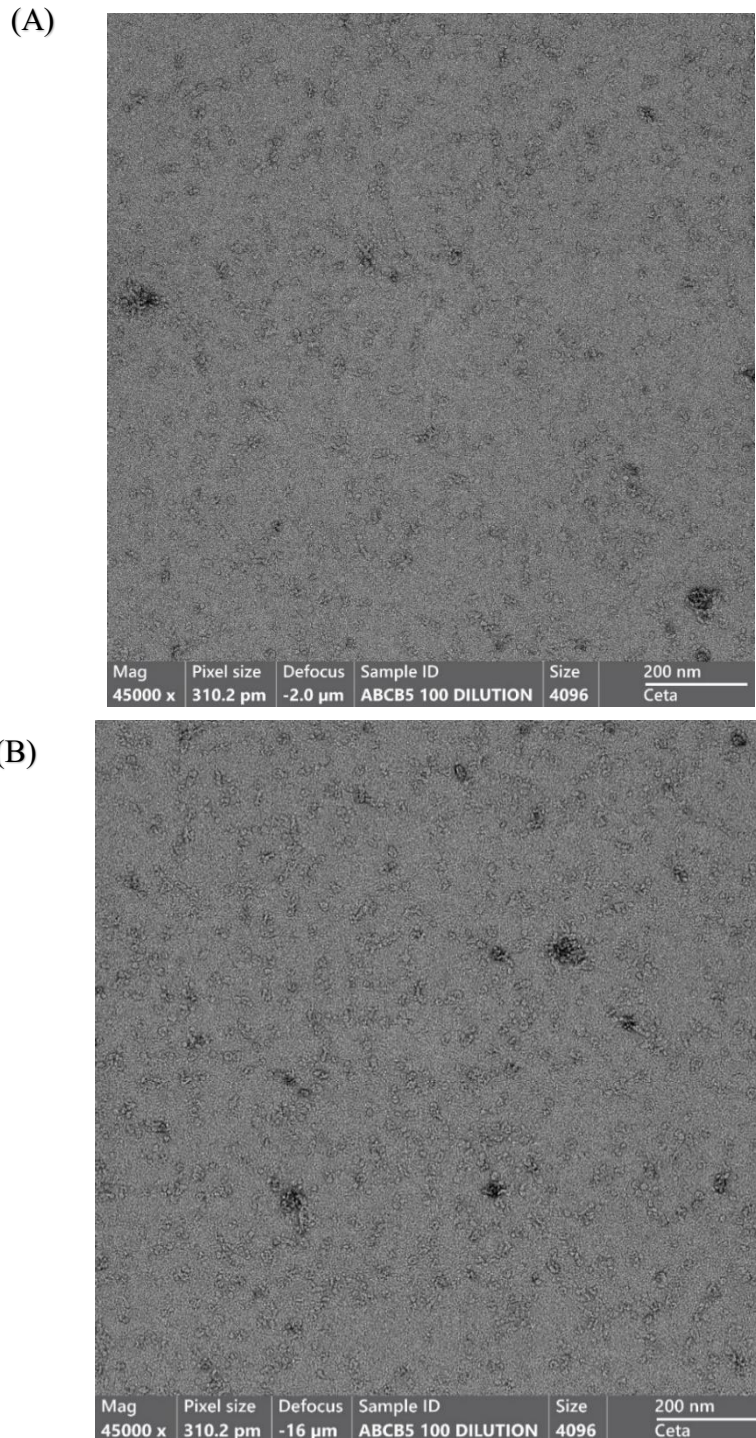
Figure 6.17: Negatively stained mABCB5 protein analysis using transmission electron microscopy. Panels A and B shows 10 nm single particles, as well as, 20 nm, 30 nm, and 100 nm aggregates**.**

(A)

200 nm

(B)

200 nm
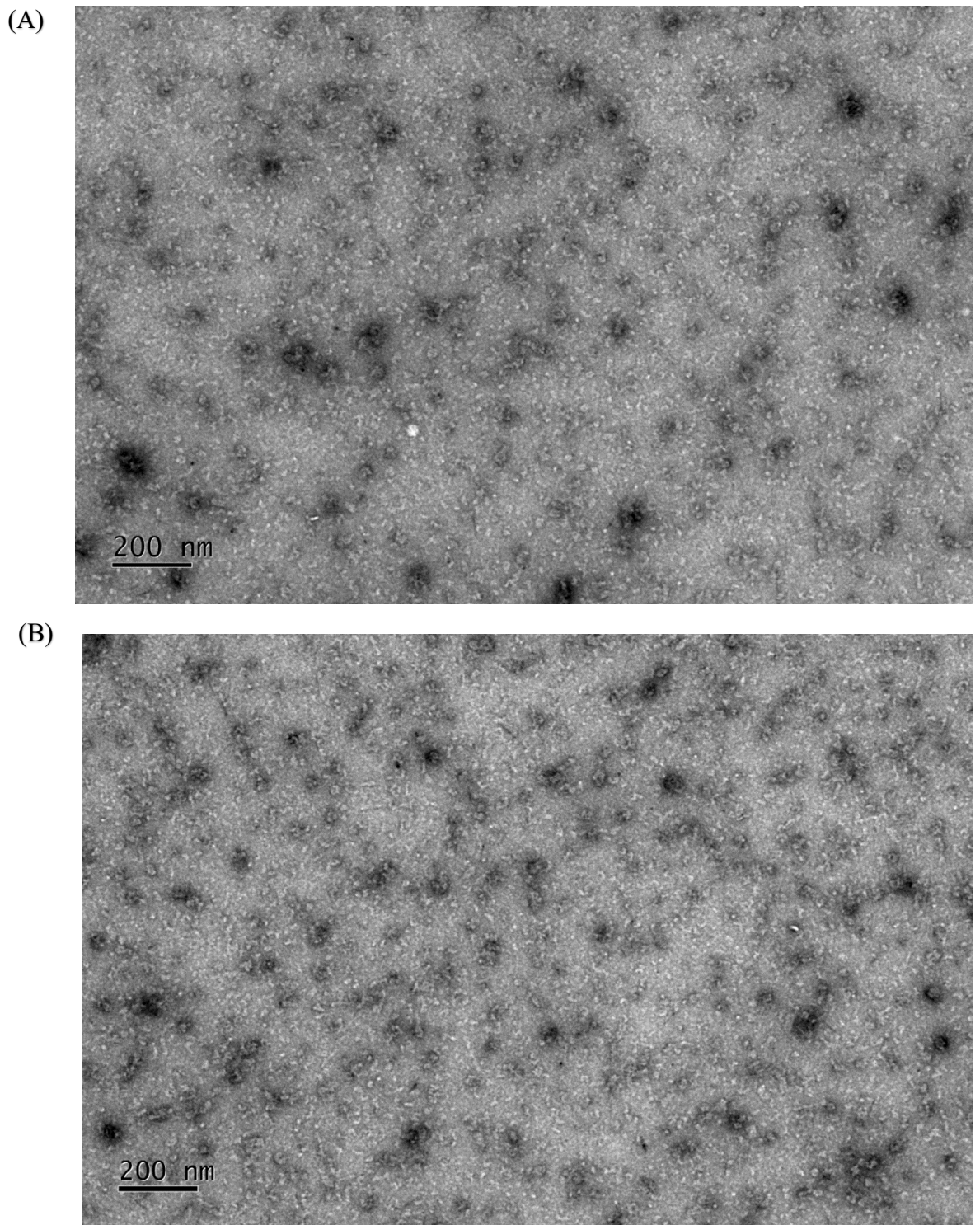
Figure 6.18: Negatively stained gpABCB6 protein analysis using transmission electron microscopy Panels A and B shows 10 nm single particles, as well as, 20 nm, 30 nm, and 100 nm aggregates.

(A)

Mag	Pixel size	Defocus	Sample ID	User	Size	100 nm
57000 x	244.9 pm	-13 μm	abcg1	iqra younus	4096	Ceta

(B)

Mag	Pixel size	Defocus	Sample ID	User	Size	100 nm
57000 x	244.9 pm	-11 μm	abcg1	iqra younus	4096	Ceta
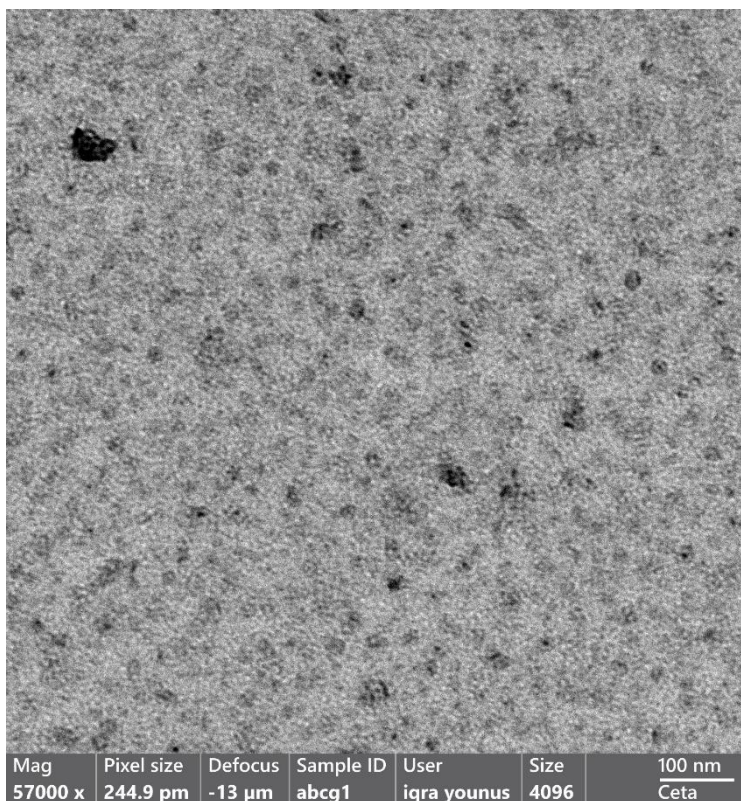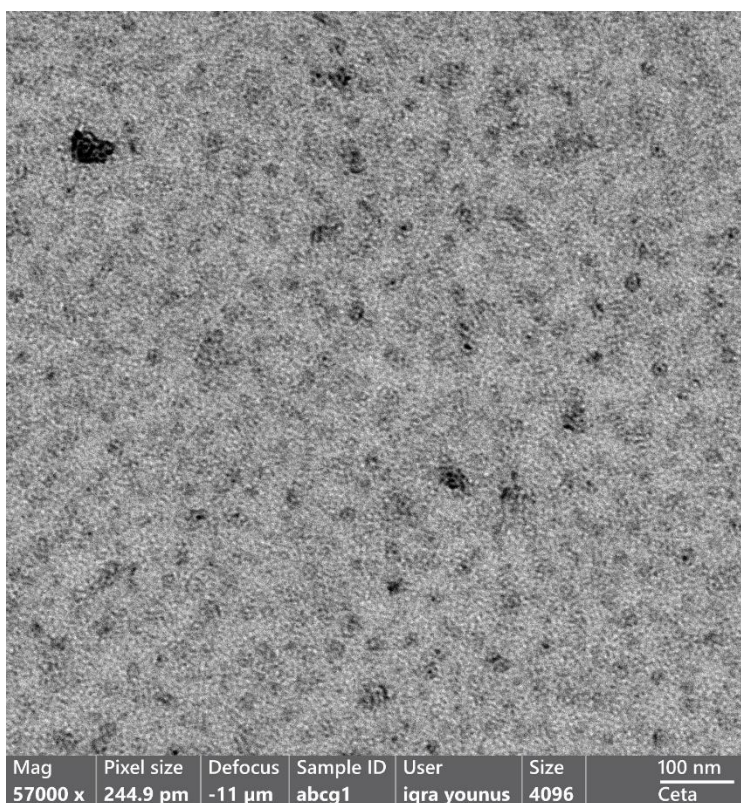
Figure 6.19: Negatively stained bbABCG1 protein analysis using transmission electron microscopy. Panels A and B shows 10 nm single particles, as well as, 20 nm, 30 nm, and 100 nm aggregates.

(A)
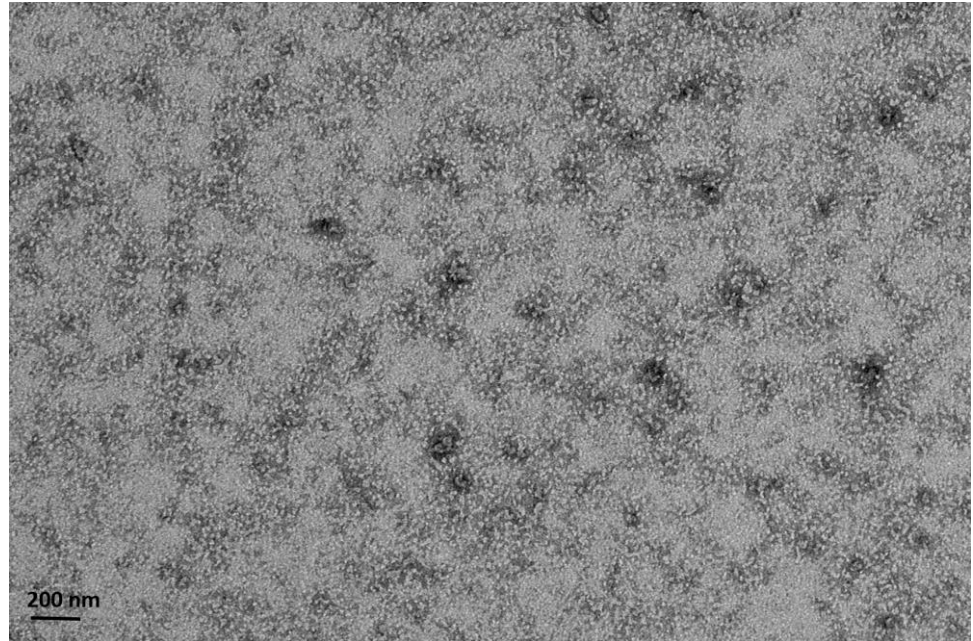
(B)
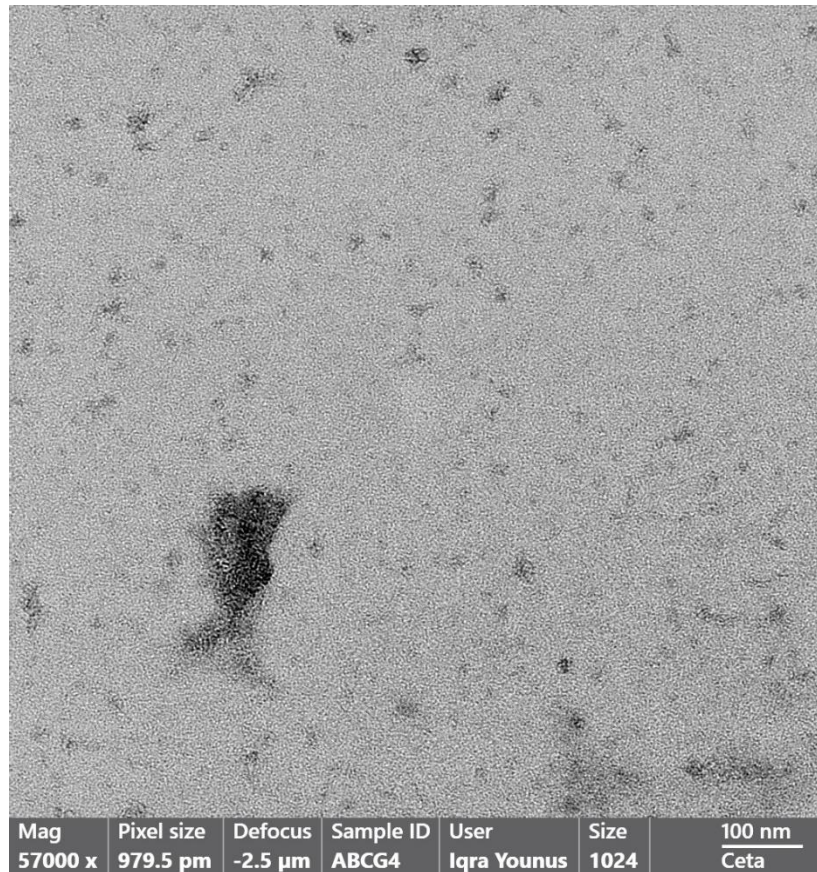
Figure 6.20: Negatively stained mABCG4 protein analysis using transmission electron microscopy. Panels A and B shows 10 nm single particles, as well as, 20 nm, 30 nm, and 100 nm aggregates.

6.7 Discussion and Conclusion

The monodispersity, thermal stability, and ATPase activity of all four of these proteins (mABCB5, gpABCB6, bbABCG1, and mABCG4) were assessed in this research work, and these characteristics are expected to be important factors before conducting high-resolution structural research. For ABC transporter research, ATPase activity assays are necessary to verify whether the protein is still active under the experimental conditions. There are many commonly used assays to measure ATPase activity, including the enzyme-coupled assay (Adam, 1965), the malachite green/molybdate assay (Baykov et al., 1988), and the Chifflet assay (Chifflet et al., 1988). These assays detect released products of ATP hydrolysis either ADP or inorganic phosphate (Pi). The Chifflet assay, which was developed to measure released P_i calorimetrically, was used in this study (discussed in section 6.2). All of these four proteins were shown to exhibit a significant basal ATPase activity in the absence of any substrate, in the detergent environment. The basal ATPase activity appeared to lower but low basal ATPase activity has been detected in some ABC transporters such as MRP1 and MRP3 (Li et al., 1996, Gorbulev et al., 2001). Additionally, in the presence of the well-known substrates, ATPase activity of all of these four purified proteins was determined in detergent environment (see section 6.2). This was important because purity of the protein may play a significant role in a total ATPase activity of the sample since only a small amount of highly active ATPase contaminants may contribute to the overestimation of an actual ATPase activity of the sample (Ban and Yang, 1998). According to Chapter 5, two-step purification produced proteins that were not of a high purity as demonstrated by Commassie stained gels. However, substrate stimulation or inhibition data of ATPase activity is a sign of purified target protein. Therefore, the likelihood of signals originating from ATPase contaminations is considered to be low. However, this study did not quantify transport activity. A protein that showed high ATPase activity might yet have low transport activity if it was somehow dysfunctional and uncoupled. Hence, the information seems insufficient to compare mABCB5, gpABCB6, bbABCG1 and mABCG4 protein samples in terms of how intact they were relatively to each other.

ATP hydrolysis is recognised as a vital activity of ABC transporters (including mABCB5, gpABCB6, bbABCG1, and mABCG4) because it takes part in the protein's conformational changes in the transport mechanism. ATP binding ability was examined using a series of TNP-ATP concentrations. Since it has been reported that nucleotide binding ability of ABC transporters is independent from substrate binding (Sauna and Ambudkar, 2001), no

substrate was included in this experiment as well as in the cryo-EM experiments performed later in this research.

Thermal stability assays were conducted using three methods. Monitoring the intrinsic tryptophan fluorescence with temperature increase is an efficient, cost-effective method for assessing the thermal protein structural stability (Garidel et al., 2008). The CPM approach makes use of a fluorescent dye that strongly fluoresces when it binds to cysteine residues that are exposed on the surface. The assay was designed to address the shortcomings of existing assays, which frequently have low signal-to-noise ratios because of the detergent background and necessitate many components (Yeh et al., 2006, Alexandrov et al., 2008) The APJ receptor, an essential membrane protein that is a member of the GPCR superfamily, had its thermal stability successfully determined by this assay (Alexandrov et al., 2008). Additionally, it was used to examine the thermal stability of the CFTR in the presence of different potentiators (Meng et al., 2017) the analysis of cytochrome c oxidase's thermal stability, helped to determine the melting points of the complex's component subunits (Kohlstaedt et al., 2015). When compared to the Tm established by intrinsic tryptophan fluorescence, the melting temperature determined by the CPM assay is often higher for each target protein (see section 6.5.2). This could be explained by the fact that each protein has a higher ratio of cysteine to tryptophan. mABCB5 have 18 tryptophan and 20 cysteine residues. On the other hand, gpABCB6 possesses 22 tryptophan residues and 16 cysteine residues, bbABCG1 possesses 14 tryptophan residues and 26 cysteine residues and mABCG4 possesses 8 tryptophan residues and 36 cysteine residues in protein dimer (Figure 6.16) (Table 6.8).

This study shows an increase in T_m of the mABCB5, bbABCG1, and mABCG4 proteins in the presence of nucleotides. This suggests that nucleotide binding may introduce some conformational changes to the protein and more kinetic energy is required for the protein to undergo unfolding to a similar degree. Interestingly, this finding is in agreement with the findings of (Yang et al., 2017) where they reported catalytically-inactivated E552A/ E1197A mouse P-gp in the presence of nucleotide showed an elevated T_m of up to 6.2 °C in the presence of nucleotide, indicating enhanced structural stability. All of these four proteins have almost similar melting temperatures (T_m), and at first glance, their aggregation temperature (Tagg) values would not distinguish them dramatically. However, supplementing the SLS intensity values determined by intrinsic tryptophan fluorescence with the DLS sizing and polydispersity information allows us to conclude that mABCG4 and mABCB5 are more "well-behaved" and do not form large heterogeneous aggregates

during thermal stress. Therefore, they may be better and more robust candidates for moving forward with high-resolution structural experiments. However, from purification chapter 05, Coomassie stained gels suggest mABCB5 is less pure than mABCG4, hence mABCG4 was prioritised for high-resolution structural research.

mABCB5 homodimer
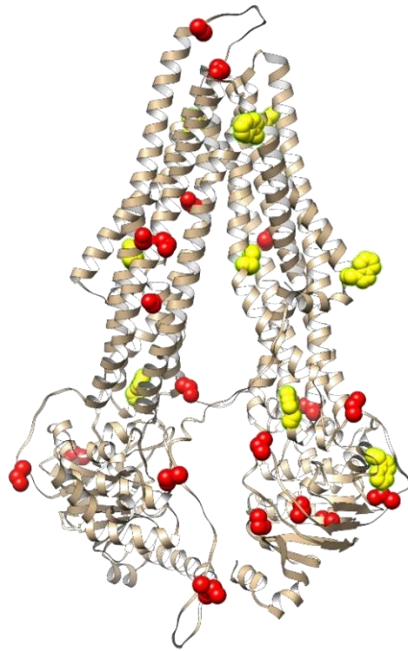
gpABCB6 homodimer
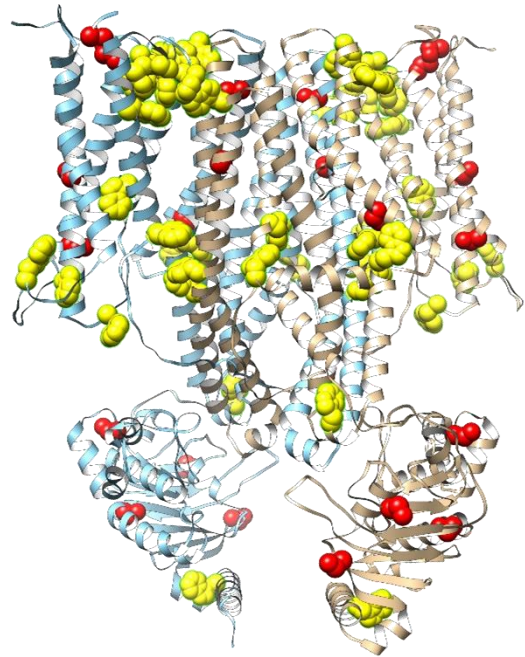
bbABCG1 homodimer
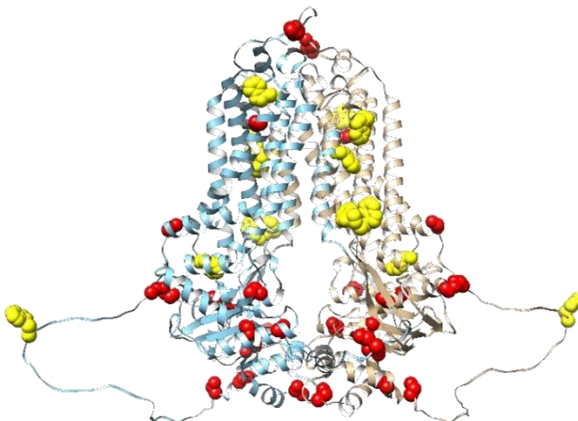
mABCG4 homodimer
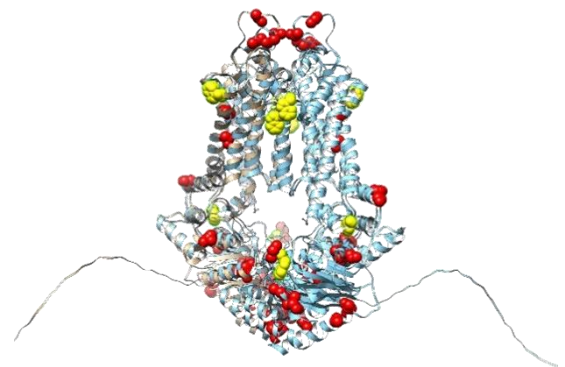
Figure 6.17: An illustration of tryptophan and cysteine residues as yellow and reds colour respectively.

Table 6.8: Total number of cysteine residues, as well as, surface-exposed and buried cysteine residues in each protein homodimer calculated in Chimera.

Protein	Surface-exposed cysteine residue in homodimer	Buried cysteine residue in homodimer	Total cysteine residue in homodimer
mABCB5	09	11	20
gpABCB6	08	14	22
bbABCG1	22	04	32
mABCG4	30	06	36

mABCG4 protein was further assessed by CPM assay and also by the membrane thermal shift assay. By CPM assay mABCG4 was investigated whether some well-known substrates have any effects on mABCG4 thermal stability. Except for cholesterol, sphingolipids, cyclosporin A and estriol increased the T_m of mABCG4 up to 10 degrees as compared to apo-state. Many tyrosine kinase inhibitors (TKIs) have been reported to interact with ABC transporters in recent years, including imatinib, nilotinib, dasatinib, ponatinib, gefitinib, erlotinib, lapatinib, vandetanib, sunitinib, and sorafenib (Beretta et al., 2017). TKIs interaction with mABCG4 was checked by membrane thermal shift assay. Our findings showed that all tested TKIs had a thermal stabilising impact on mABCG4 at a temperature of 70 degrees except alectinib, AXL1717, cabozantinib, and foretinib (see Figure 6.16). Some of the TKIs also tested for ATPase activity and results demonstrated that dovitinib, gefitinib, lenvatinib, tivozanib, masitinib at 10 μM concentration increased the ATPase rate of mABCG4 protein; while cedritanib had an inhibitory effect on mABCG4 (see figure 6.4C). This can be justified because the relationship between the molecular structure of TKIs and their capacity to engage with ABC transporters hasn't yet been established, though (Hegedűs et al., 2002). In fact, TKIs exhibit a large variability of transporter selectivity and affinity. TKIs typically have substrate-like qualities at low concentrations and can block the activity of transporters at high concentrations (Anreddy et al., 2014).

Negative stain data of gpABCB6, bbABCG1, and mABCG4 is optimistic at this step; while mABCB5 protein purification needs a bit more work. Even if a protein appears pure and only displays one band on an SDS PAGE gel, the particles are likely to appear heterogeneous when seen by negative stain EM (Ohi et al., 2004). The negative staining process itself has the potential to produce a small amount of heterogeneity because of protein distortions during adsorption and/or a varying degree of stain embedding. However, there may be more considerable heterogeneity in the purified protein population due to particles aggregating into various oligomeric forms, particles adopting various conformational states, and particles

adsorbing to the grid in various orientations (resulting in identical molecules having a distinct appearance). Even though heterogeneity is typically present in pure protein preparations, it is more obvious when macromolecular complexes that frequently disintegrate are researched (Ohi et al., 2004).

Chapter 7: Results and Discussion: Small Angle X-ray Scattering

SAXS is a powerful technique for analysing the three-dimensional (3D) structure of biomolecules in solution (Bernadó et al., 2007, Petoukhov and Svergun, 2013). SAXS is a simple method that can be used to analyse proteins ranging in size from a few kDa to GDa. It can give numerical evaluations of structural characteristics of proteins, such as size and shape. In addition, the intrinsic disorder, dynamics, and conformational polydispersity of proteins in their native state can also be revealed by it (Doniach, 2001, Pelikan et al., 2009). The goal of this experiment was to use SAXS to create low-resolution structural envelopes of ABCB5, ABCG1, and ABCG4. These findings might be useful for comprehending the protein's shape, structure, and molecular dynamics.

7.1 Data analysis

SEC-SAXS data was collected in apo-state for 2.5 mg/ml mABCB5 and 1.5 mg/ml bbABCG1 at Diamond Light Source Beamline 21. In contrast, SAXS data for mABCG4 was collected in batch mode at 2 different concentrations of 5 mg/ml and 10 mg/ml in the apo state (see section 2.21.2). The data were collected with 600 frames per sample in SEC-SAXS mode and 21 frames per sample in batch mode to maintain consistency and detect any radiation damage. The data analysis program ATSAS from EMBL was utilised for data processing (Manalastas-Cantos et al., 2021). Each frame was imported into the PRIMUS interface *i.e.* SEC-SAS data processing and for batch into SAS data analysis (Konarev et al., 2003). Figures 7.1 and 7.2, respectively, illustrate the elution profiles of mABCB5 and bbABCG1 produced by ATSAS (mABCG4 elution profile is not shown). Frames were chosen for the protein sample, as indicated by the green colour, and the buffer, as indicated by the red colour. The buffer was then subtracted from the mABCB5 and bbABCG1 apo samples after the frames had been averaged. Experimentally determined 1D scattering profiles of 5 mg/ml and 10 mg/ml mABCG4 is shown in Figure 7.3.
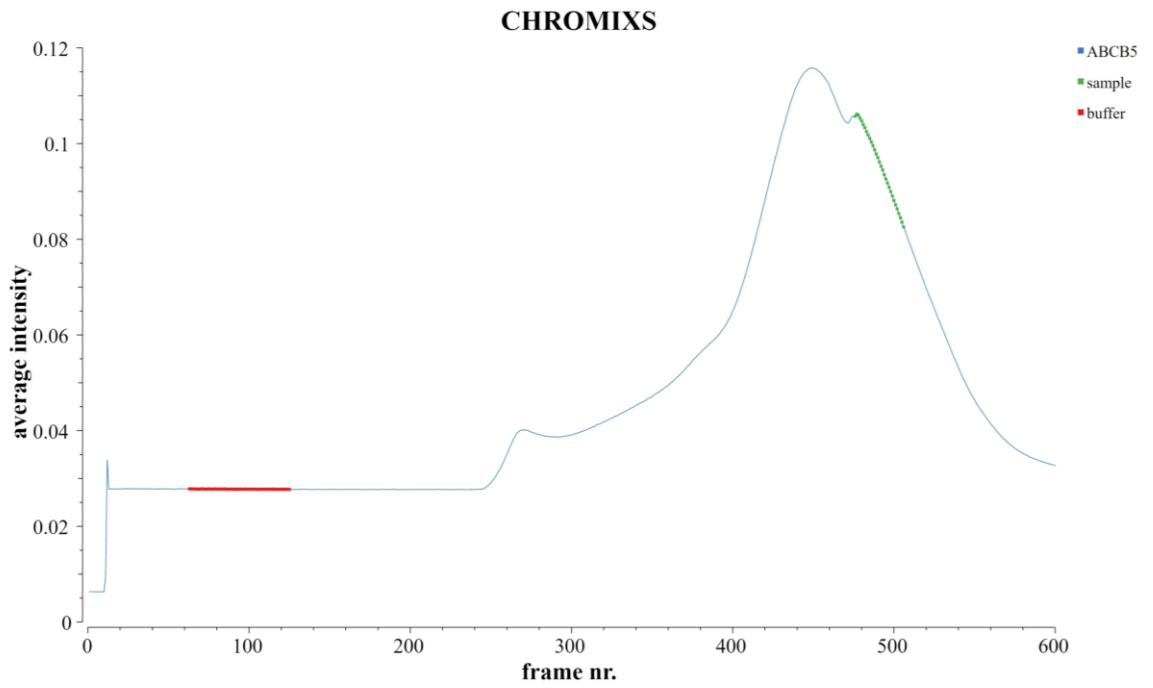
Figure 7.1: mABCB5 elution profile. Total frames: 600, **Buffer selection:** 33-97 frames, **sample selection:** 475-506 frames
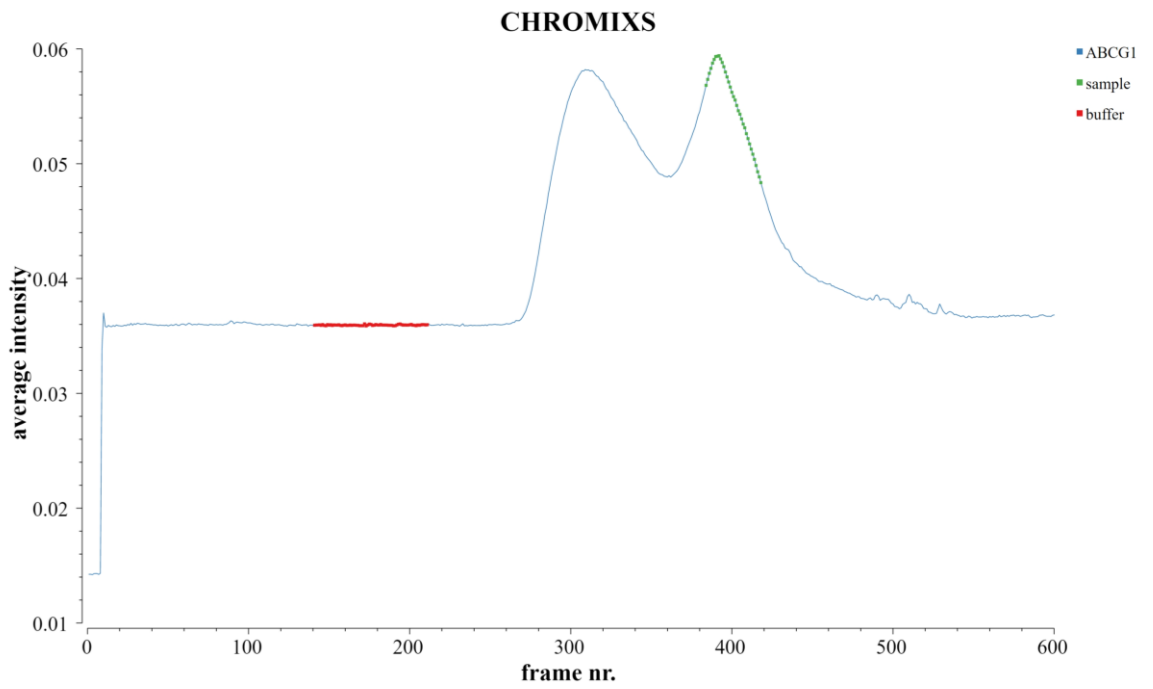
Figure 7.2: bbABCG1 elution profile. Total frames: 600, **Buffer selection:** 141-211 frames, **sample selection:** 384-418
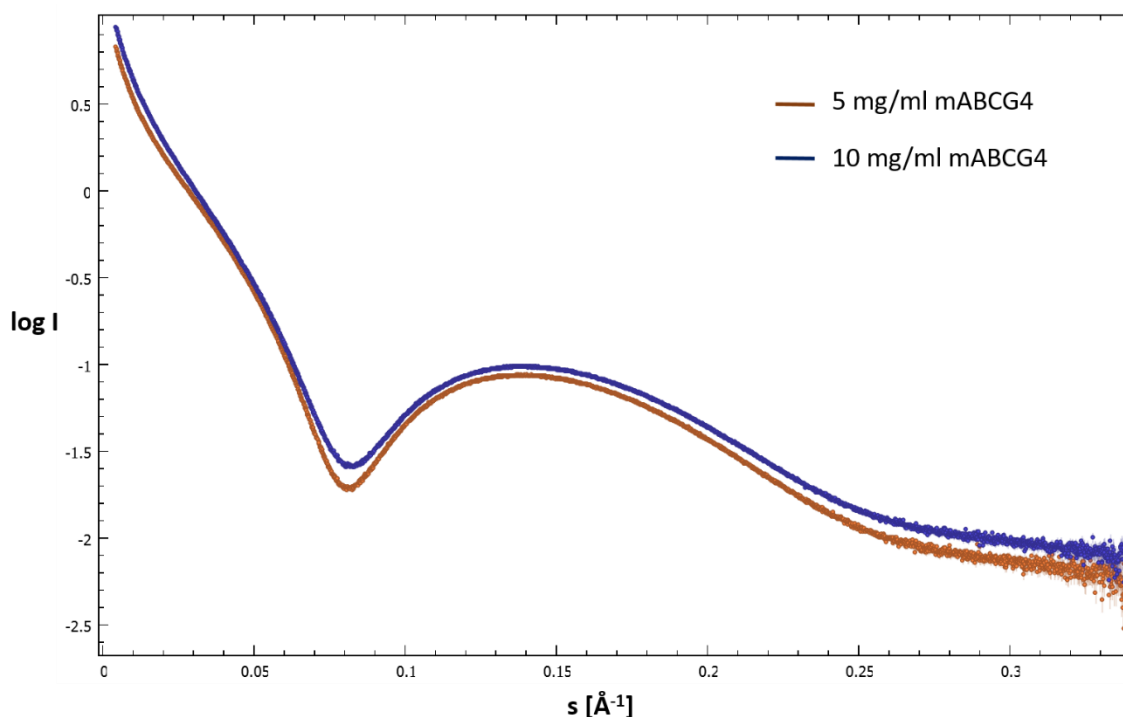
Figure 7.3: Experimentally determine scattering profiles of mABCG4 at 5 mg/ml and 10 mg/ml concentration.

A qualitative examination of each scattering profile shows that at low values of momentum transfer (S) the gradient of the log(I) versus S plot increases as S decreases: This is consistent with the presence of aggregates (Figure 7.5 A). The homology model of mABCG4, bbABCG1 and mABCB5 (see section 6.9.1) was used to generate the theoretical scattering profile server for comparison with experimentally determined scattering profiles (Figure 7.5B). The theoretical scattering calculations was done using the FoxS server (Schneidman-Duhovny et al., 2010). These calculated profiles have no contributions from oligomer or aggregate structures and do not show the same intensity variation at low S as the measured data.

Comparing the calculated profiles with the measured profiles at intermediate values of momentum transfer (0.07<s<0.25): It is interesting to note that the experimentally determined scattering profiles of mABCB5, bbABCG1, and mABCG4 all exhibited a hump which is not seen in theoretical scattering profiles.

Gross structural parameters can be extracted from the SAXS data using a Guinier plot and *via* the calculation of a pair distribution function. Figure 7.6 depicts the Guinier analysis used for calculating the radius of gyration (R_g) (Sorensen and Shi, 2000). R_g describes the overall distribution of the mass of the protein around its centre of gravity. The Guinier analysis of mABCB5, bbABCG1, and mABCG4 showed radii of gyration of 65.88 Å, 58.85

Å, and 70.83 Å respectively. This Guinier analysis has a drawback in that it is solely dependent on low "q" values, making it extremely susceptible to discrepancies at large scattering angles or in conditions where there is inter-particle interaction. Pairwise Distance Distribution Function (PDDF) is another reliable method of describing the SAXS data (Debye 1915). With the use of this PDDF, a Fourier transformation, the reciprocal space data from scattering is transformed into a real space distance distribution histogram, where each point represents the separation of two atoms in the given sample. This distance distribution graph provides information about the particle's size and general shape. The pair distance distribution analysis for mABCB5, bbABCG1, and mABCG4 showed D_{max} of ~300 Å, 353.26 Å, and 250 Å respectively.
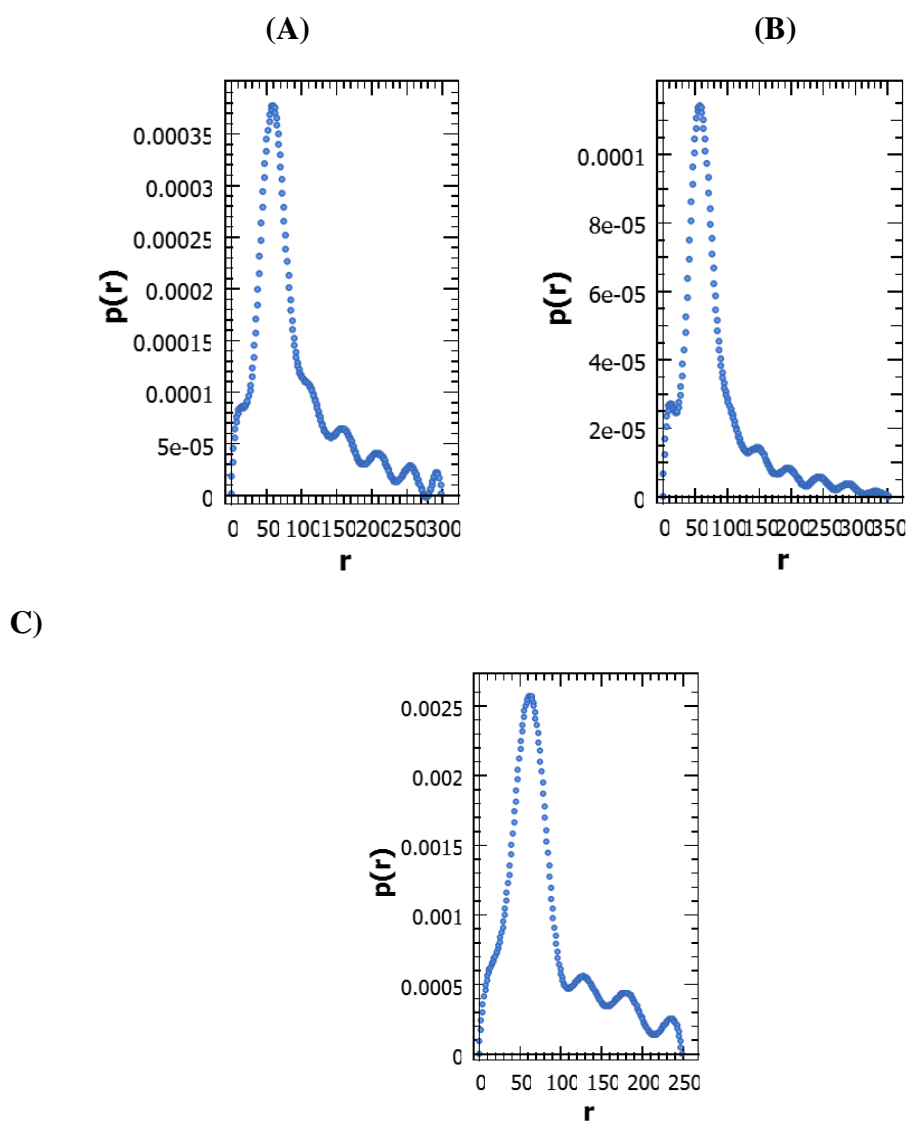
Figure 7.4: Experimentally determined pair-distance distribution plots. A) mABCB5 **B)** bbABCG1 **C)** mABCG4

A)

B)
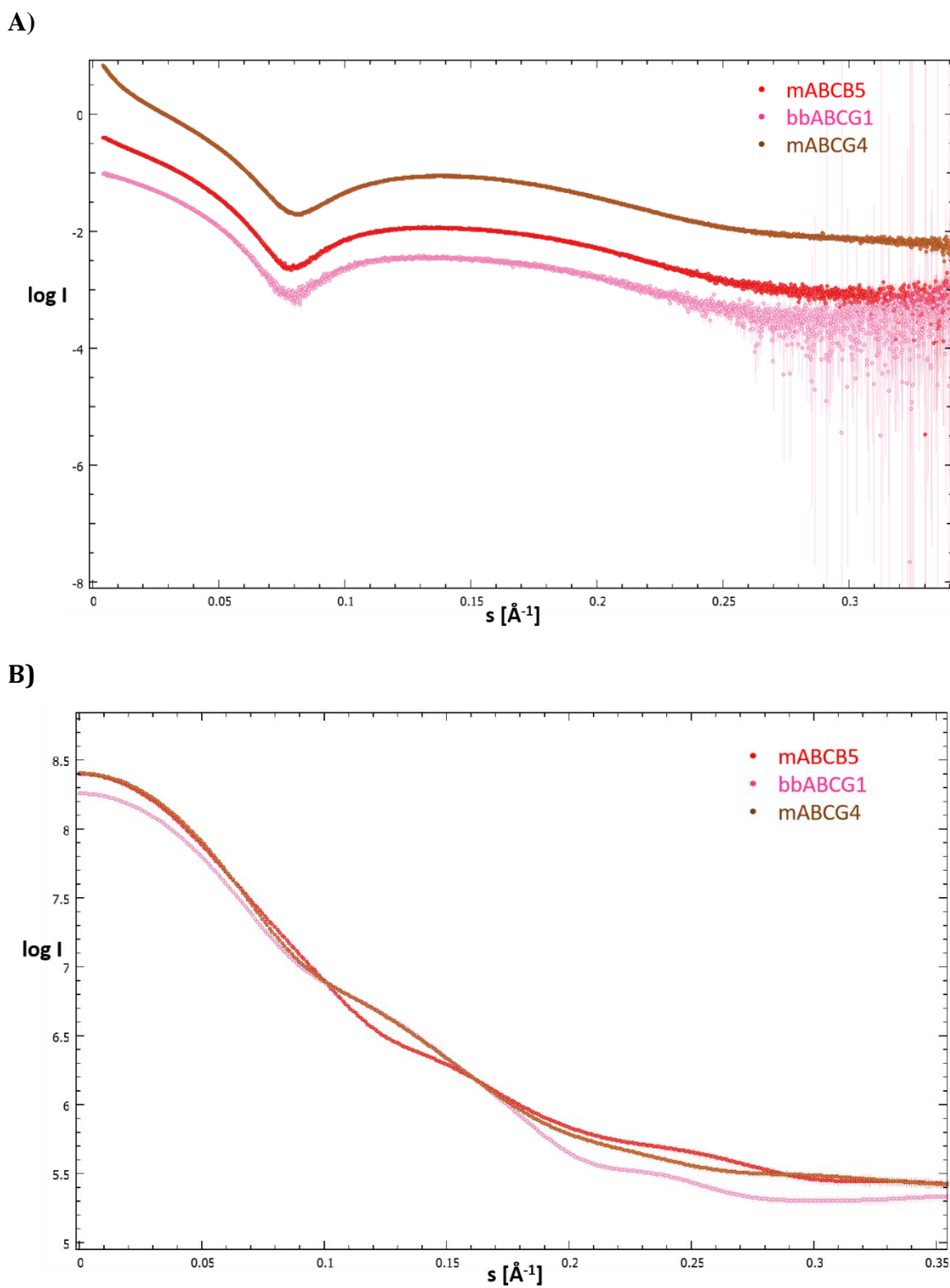
Figure 7.5: Scattering profiles of mABCG4 (5 mg/ml), bbABCG1, and mABCB5 in apo state, plotted as a function of intensities vs momentum transfer (s=4πsinΘ/λ). A) Theoretical scattering profile of mABCG4, bbABCG1 and mABCB5 generated using FoXS server. **B)** Experimentally determined scattering profile of mABCG4, bbABCG1, and mABCB5.

A)
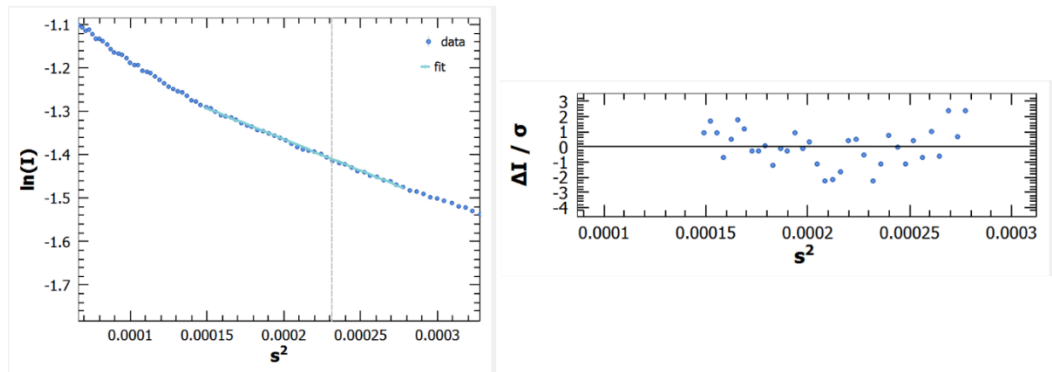
ABCB5

B)
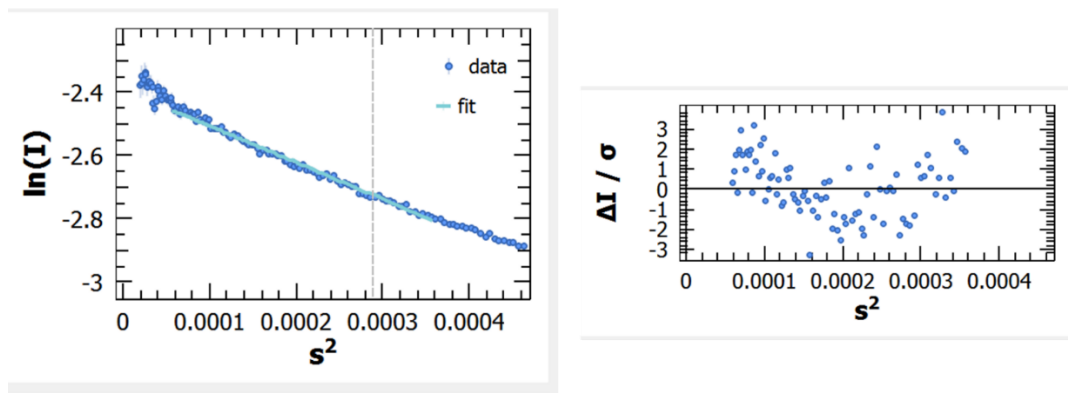
ABCG1

C)

ABCG4
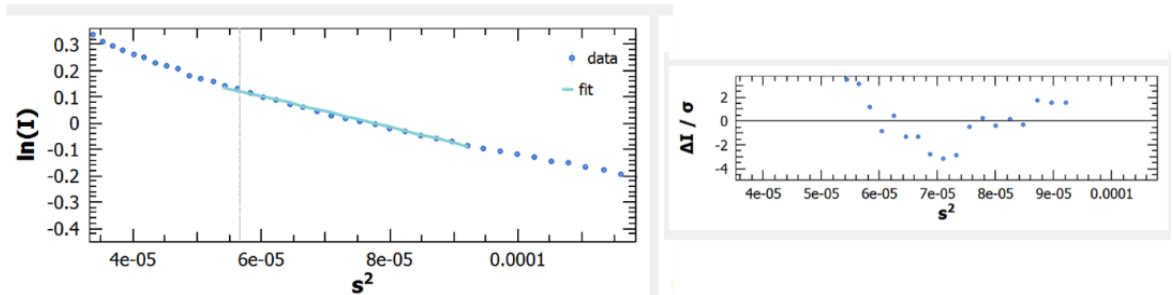
Figure 7.6: Guinier analysis of ABCB5, ABCG1, and ABCG4 in the apo state, panel A, B, and C respectively. In the Guinier plot (ln(I) versus s^2) blue dots represent the experimentally determined scattering intensity, Light blue line is the best fit of a straight line to the transformed scattering data. The corresponding residual plot (deltaI/sigma versus s^2) shows the deviation of each point from the fitted line. In each case the residual plot is curved suggests that the Guinier plot is not fitted well with a straight line

Table 7.1: Summary of structural features determined by SAXS.

	Rg (Guinier) (Å)	I (0)	Rg/ I_0 p(r) (Å)	Theoretical calculated Rg (Å)	Dmax (Å)
mABCB5	65.88 ± 0.17	0.34 ± 0.0005	75.26 ± 0.35	38.74	300
bbABCG1	58.85 ± 0.19	0.09 ± 0.0002	71.12 ± 0.10	34.74	353.26
mABCG4	70.83 ± 0.11	2.51 ± 0.003	71.47 ± 2.42	36.23	250

7.2 Discussion and Conclusion

It is difficult to study membrane proteins structurally using SAXS because of the detergent issues (Lipfert and Doniach, 2007). It may offer system-related information that can be used to inform high-resolution experiments. In this experiment, structural parameters from the scattering profile suggest large aggregations in mABCB5, bbABCG1, and mABCG4. One possible explanation might be presence of detergent. Membrane proteins must be solubilized during purification to maintain their structural integrity in aqueous solutions. The most straightforward and extensively used way of solubilizing membrane proteins is detergent molecules above the CMC (le Maire et al., 2000). These detergent molecules form a corona in the vicinity of the lipid-embedded area (Pérez and Koutsioubas, 2015). These areas have the potential to strongly scatter X-rays, which can change the ultimate scattering patterns of integral membrane proteins (Koutsioubas, 2017). This correlates with the common features seen in the scattering profiles of all three samples at intermediate ranges of S. The detergent micelle that surrounds the protein may be the cause of hump characteristic of scattering profiles. With aquaporin0 solubilized with DDM, a similar hump scattering characteristic was also seen (Molodenskiy et al., 2020).

Conclusion

All-together, in this study, we investigated the structural characteristics of mABCB5, bbABCG1, and mABCG4 in the apo state using a low-resolution small-angle scattering approach. Unfortunately, our findings indicated protein aggregation and since data from a homogeneous sample is required for ab initio modelling of envelopes from SAXS data (Kikhney and Svergun, 2015) we were unable to proceed with this analysis.

Chapter 08: Structural studies on mABCG4

To determine structure of protein at high resolution two methods are commonly employed X-ray crystallography and cryo-EM. For this research work, after bioinformatics analysis four proteins were selected for crystallization trials. but by the end of PhD stage, 3D crystallisation was taking a lesser role for determining the structures than cryo-EM. Also as seen in chapter 5 further purification of these proteins would be required to increase their quality and quantity because protein crystallisation requires substantially more concentrated and 99.95% homogeneous protein (Caffrey, 2003). Furthermore, it is necessary to consider each individual membrane protein, as well as its preferred detergent and occasionally lipids, when planning crystallisation experiments (Saidijam et al., 2003). Considering these limitations and time constraints on completion of thesis work, unfortunately, no target protein has been screened for crystallisation in order to determine its structure, in this study. Alternatively, cryo-EM was employed on most promising protein for structure determination. mABCG4 showed the best size exclusion profile in the last purification step (chapter 05). we performed four cryo-EM trials on the mABCG4 (Figure 8.1).

It has been reported that glycerol and detergent potentially have negative influences on both specimen preparation process and micrograph quality (Thonghin et al., 2018). Cryo-buffer (SEC buffer with reduced DDM and glycerol content) was used in cryo-EM. Grids were prepared as described earlier. Test micrographs were collected using an in-house Titan Krios electron microscope (EM facility, the University of Manchester). Each of the four attempts is depicted in a representative micrograph in Figures 8.6, 8.7, 8.8, and 8.9.

In the first attempt 2.14 mg/ml mABCG4 was applied on the Quantifoil 1.2/ 1.3 copper grids and images were recorded (see section 2.23.3.1). The test micrographs were of poor quality for processing and heterogenous particles including long string like structures were seen in micrographs Individual particles were featureless and consistent with the presence of excess detergent micelles (Figure 8.6). The lengthy strings could be explained in two ways as follows:

1. mABCG4 may be aggregated and stacked on top of each other. In previous research it has been shown that protein aggregation is one of the persistent problem in cryo-EM grid preparation (Kampjut et al., 2021).

2. Particles are predominantly detergent, which was also observed with the similar protein P-gp in our lab (Shafi, 2022). One reason for this could be that the detergent micelles were

concentrated in the samples by ultrafiltration instead being filtered away. As seen in Figure 7.2, the coverage of mABCG4 particles is also low and not uniform, concentrated mostly at the edges. This may be due to protein's propensity to preferentially adsorb at the air-water interface or more likely due to concave lens formation because of the presence of detergent(Schmidt-Krey and Rubinstein, 2011).
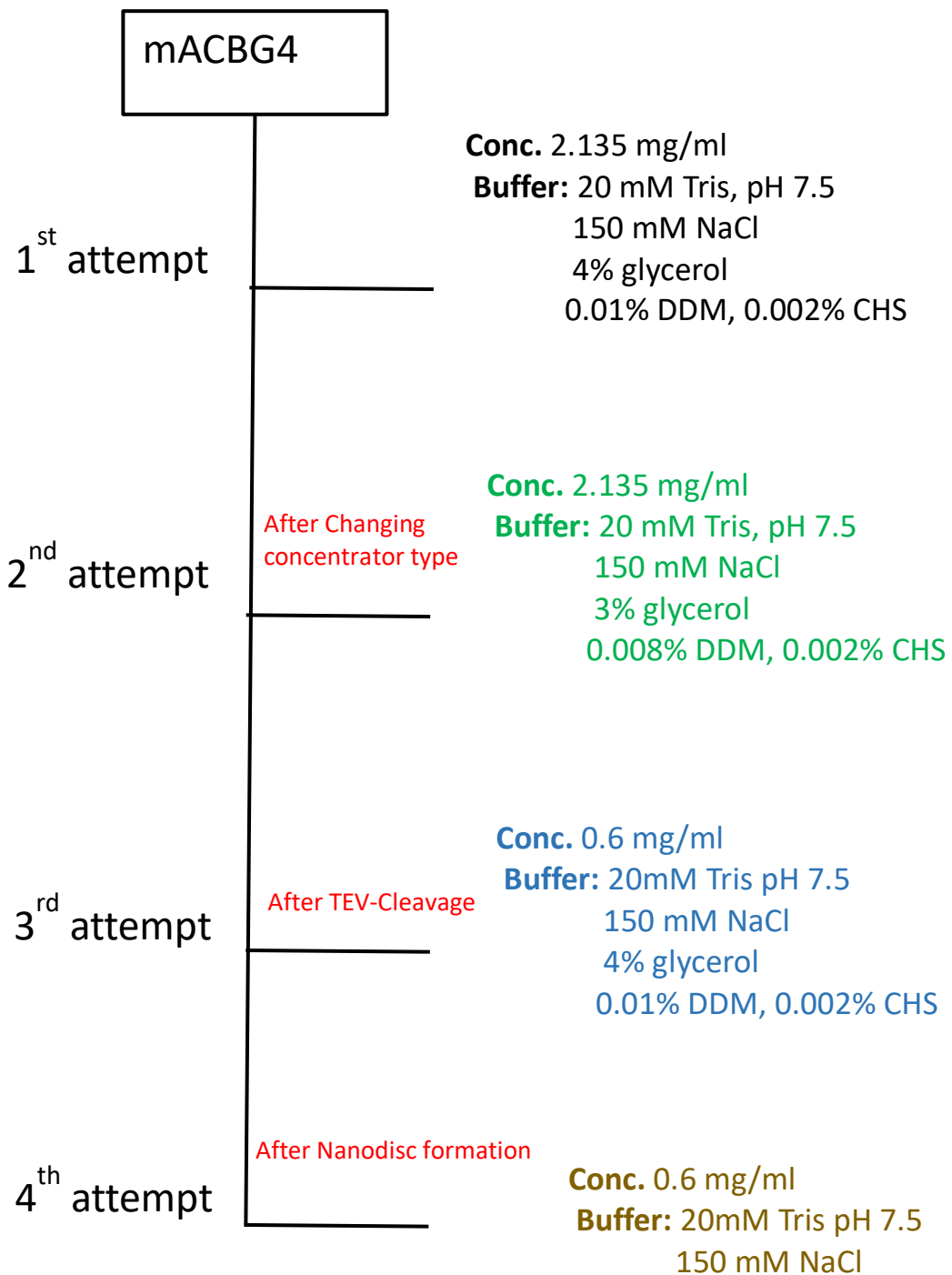
Figure 8.1: An overview of the mABCG4 cryoEM experiments
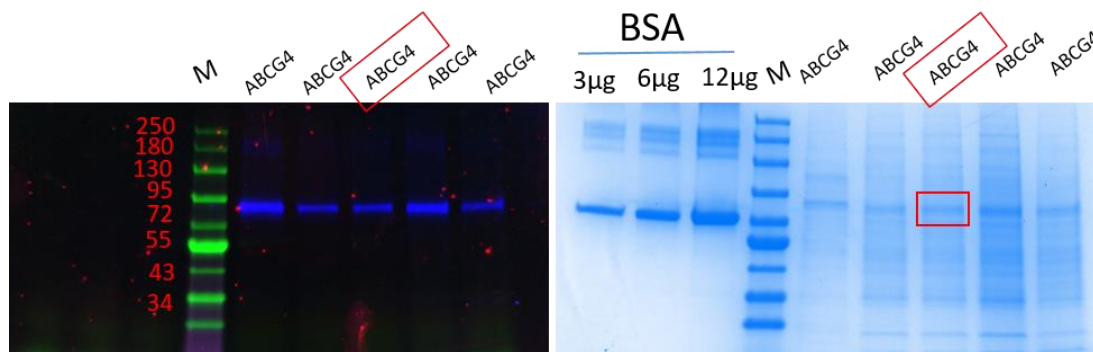
Figure 8.2: Five different batches of Purified ABCG4 on SDS-PAGE. On Quantifoil 1.2/1.3 copper grids, concentrated ABCG4 at a concentration of 2.13 mg/ml highlighted in red box was applied, and test micrographs were recorded in Titan Krios electron microscope.

To improve data quality in our second attempt, we switched from a polyethersulfone membrane (PES) to a regenerated cellulose membrane (RCM) to concentrate the protein, which decreased the ratio of these long strings and allowed us to see some, but very few, ABC type particles (Figure 8.7). This can be explained by the difference between cellulose and polyethersulfone (PES) membranes; the former is more hydrophobic while the latter has high hydrophilicity and, as a result, has an excellent porosity (Qu et al., 2010, Zhang et al., 2009b).

In our third method, mABCG4 was negatively purified by TEV protease treatment to enrich the pure mABCG4. First, optimization of the TEV enzyme was done on mABCG4, which revealed 100 times more TEV than the amount of TEV suggested by the manufacturer (Figure 8.3). We did TEV treatment using 3 times more TEV suggested by the manufacturer which was far less than the optimisation and leads to TEV treatment ineffective (Figure 8.4). However, 0.5 mg/ml ABCG4 protein with no tag was applied on the Quantifoil 1.2/ 1.3 copper grids. Grids were screened using Titan Krios electron microscope and we could hardly detect any ABC particles and mostly micelle was observed (Figure 8.8).

As TEV treatment efficiency was not good so, in order to reduce the detergent effects seen in previous trials, we made a mABCG4 nanodisc/bicelle using membrane scaffold protein 1D1 (MSP1D1) (Merck) (see section 2.20.1). Size exclusion chromatogram of reconstituted ABCG4 is represented in figure 8.5A. Fraction 18 was collected, concentrated using a 100 kDa cutoff filter, and run on 8% SDS-gel before applying on the Quantifoil 1.2/ 1.3 copper grids. Grids were screened using titan krios electron microscope. Micrographs showed few

ABC particles that were evenly distributed, although the large black dots observed which most likely indicated ethane contamination (Figure 8.9). These data are being processed.
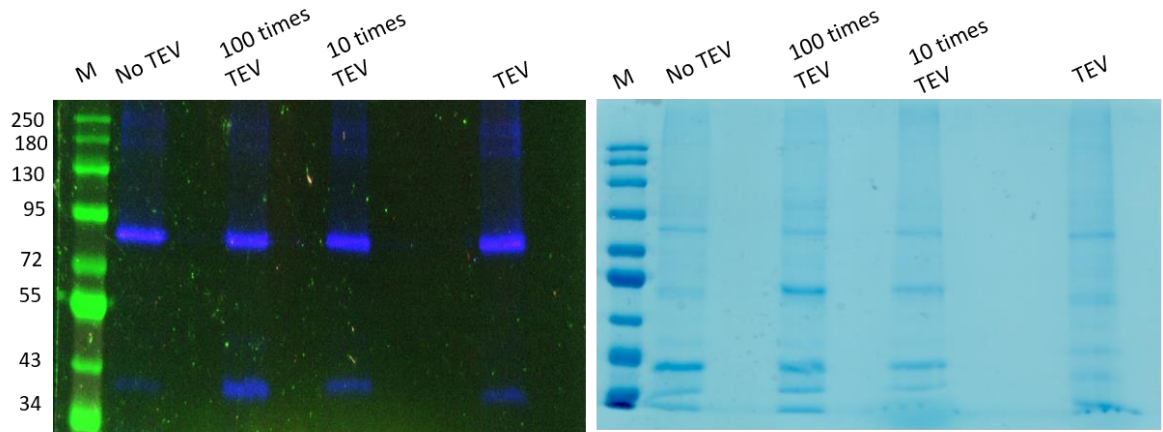
Figure 8.3: TEV optimisation of mABCG4. ABCG4 was incubated overnight with TEV protease at 4 with end-to-end rotation. The following day efficiency of TEV protease was tested by running ABCG4 treated with TEV on SDS gel. The right Panel shows the SDS gel after Coomassie staining. Left Panel shows the SDS gel scanned for GFP signal using Alexflour488.
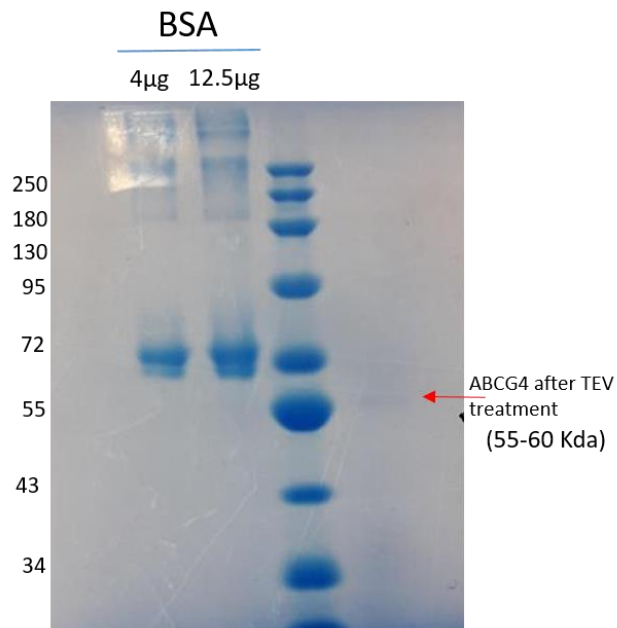
Figure 8.4: Confirmation of TEV treated mABCG4 on 8% SDS-gel before loading to quantifoil 1.2/1.3 grid.

A)

B)
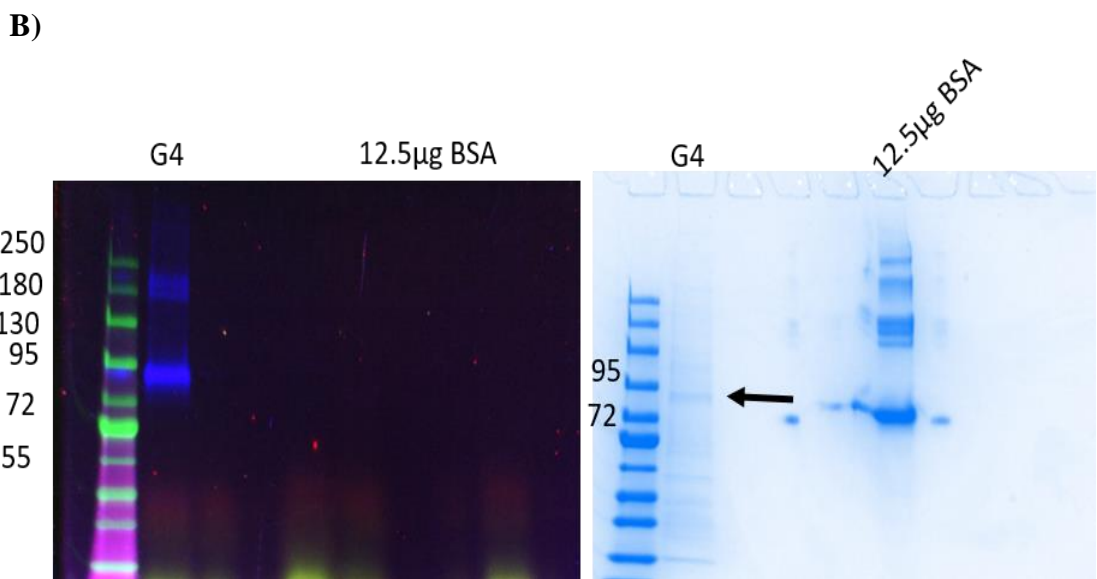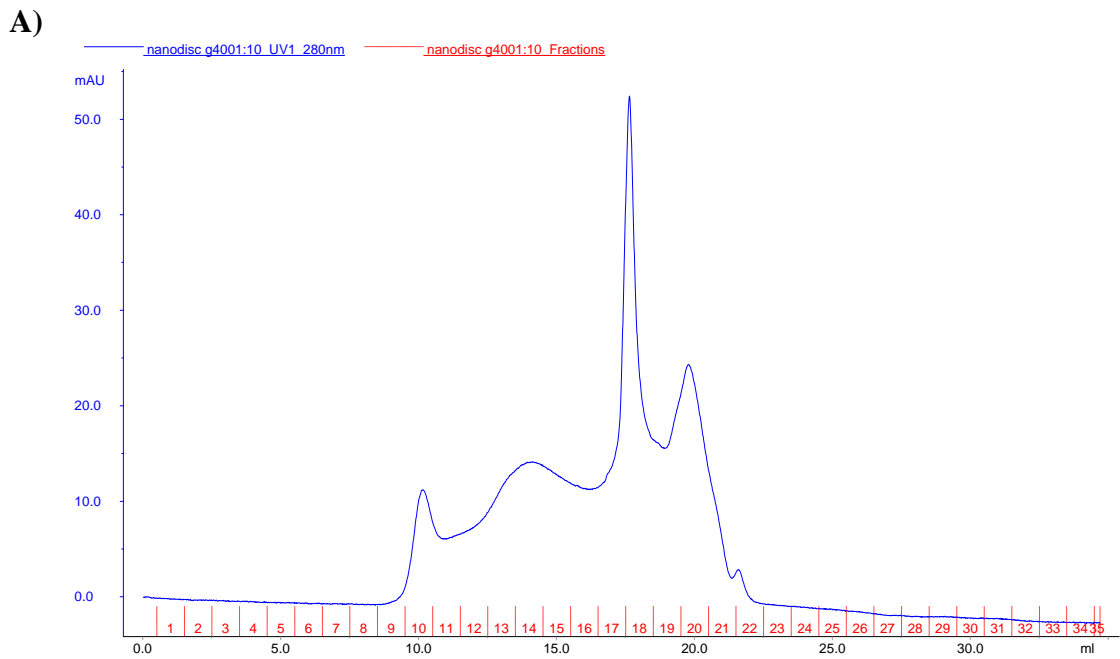
Figure 8.5: Nanodisc formation of mABCG4. Panel A represents the size exclusion profile after nanodisc formation. Fraction 18 was collected, concentrated and run on 8% SDS-gel. **Panel B** shows the SDS gel of concentrated fraction 18. The right Panel shows the SDS gel after Coomassie staining. Left Panel shows the SDS gel scanned for GFP signal using Alexflour 488.
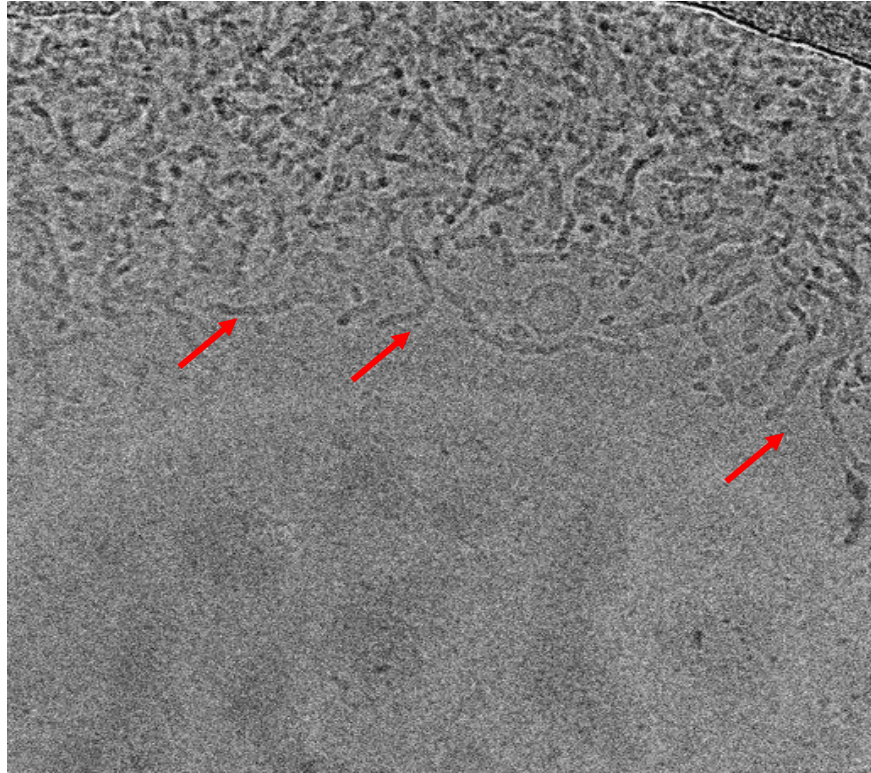
Figure 8.6: Representative micrograph of cryo-EM grids of 2.14 mg/ml mABCG4 on Quantifoil 1.2/ 1.3 grid (1st attempt). Micelles are indicated by red arrows.
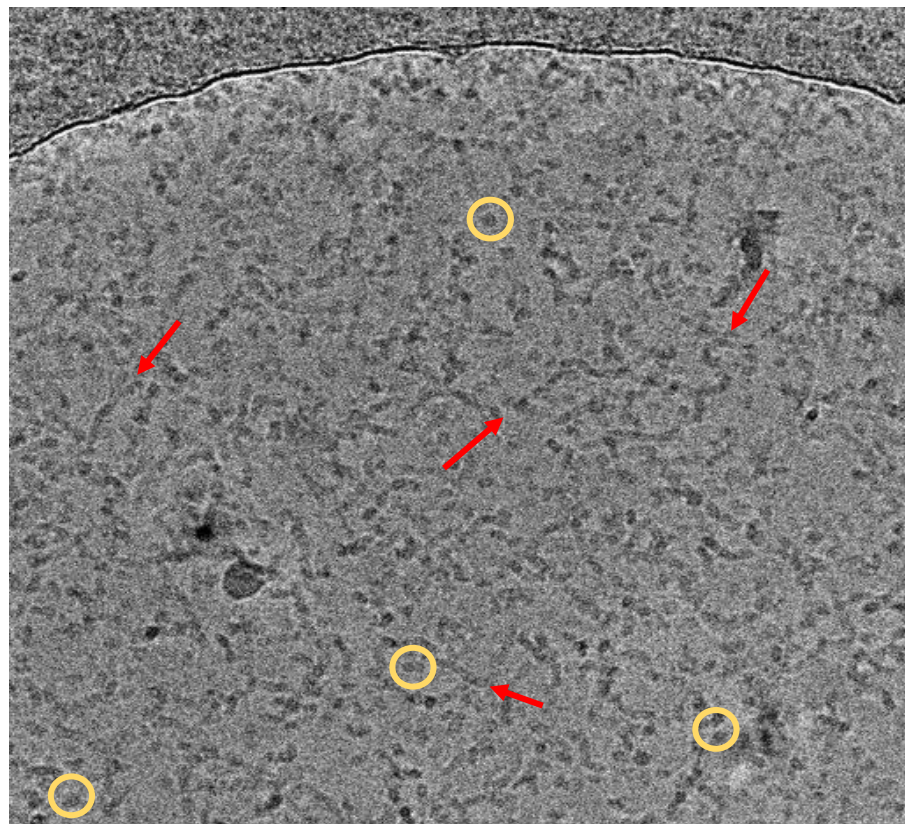
Figure 8.7: Representative micrograph of cryo-EM grids of 2.14 mg/ml mABCG4 on Quantifoil 1.2/ 1.3 grid (2nd attempt). Micelles are indicated by red arrows.; while, yellow circles denote ABC-type particles.
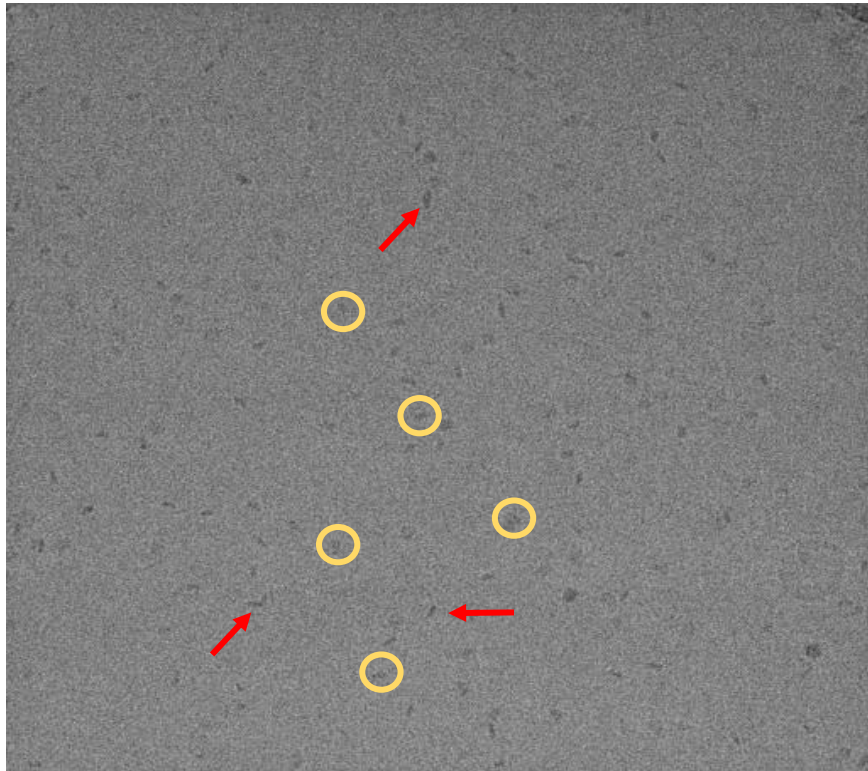
Figure 8.8: Representative micrograph of cryo-EM grids of 0.6 mg/ml TEV-treated mABCG4 on Quantifoil 1.2/ 1.3 grid (3rd attempt). Micelles are indicated by red arrows.; while, yellow circles denote ABC-type particles.
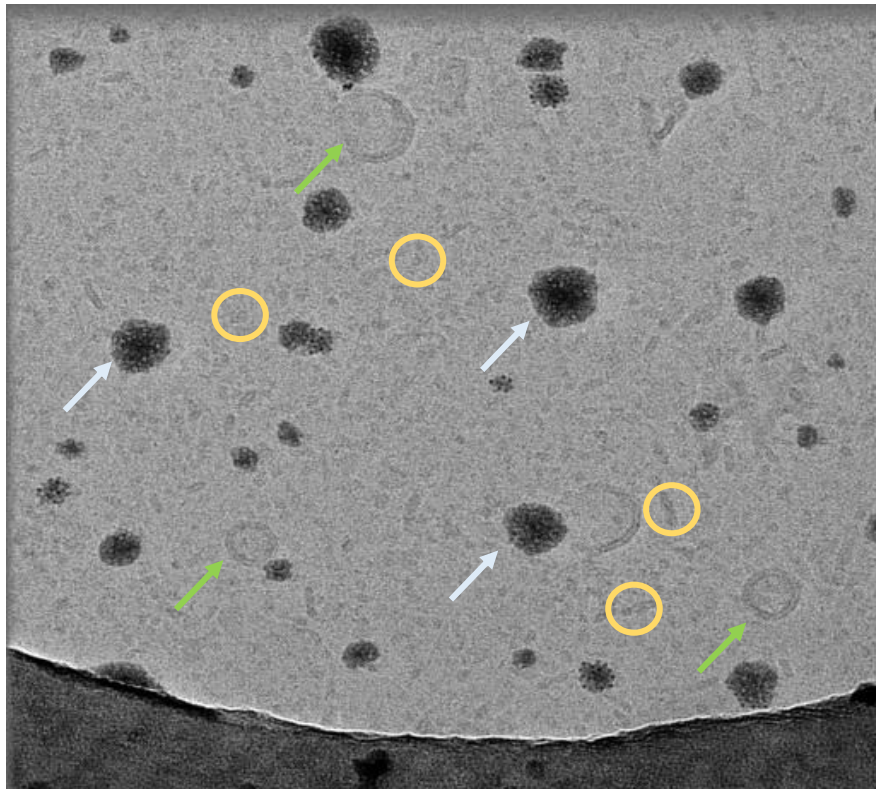
Figure 8.9: Representative micrograph of cryo-EM grids of 0.5 mg/ml mABCG4 bicelle/nanodisc (4th attempt). Yellow circles indicate ABC-type particles, whereas white arrows indicate ethane contamination. Green arrows point the empty bicelle.

Discussion and conclusion

The most promising target protein, ABCG4, was studied using cryo-electron microscopy to obtain a high resolution structure. We collected four cryo-EM test datasets in different conditions. Unfortunately, detergent micelle was mostly seen. Even though the sample was replaced in the buffer without detergent, it is possible that ultrafiltration using 100 kDa cutoff filters concentrated the detergent micelles in the protein samples rather than filtered them away. Studies from the previous have successfully used ultrafiltration columns to replace buffer (Thonghin et al., 2018b).

There may be several reasons for TEV treatment inefficiency. Because ABCG4 is a homodimer, which implies that its functional unit contains two TEV cleavage sites (as highlighted below), TEV may cleave at any site, leaving the remaining ABCG4 with His and binding to the nickel column.

8His_ GFP_ TEV_ABCG4_ABCG4_8His_ GFP_ TEV

It has been suggested that reconstituting membrane proteins into nanodiscs will provide the optimum balance between preserving a lipid-like environment at the membrane protein's boundary and enabling the membrane proteins to exist as "single particles" that can be investigated using cryo-EM (Earl et al., 2017, Wu and Lander, 2020). However, in micrographs few ABC particles were seen after reconstitution of ABCG4 into nanodisc which suggests that a good structure is unlikely to result. The image processing was therefore terminated.

Chapter 09 General discussion, conclusion and future work

About two-third of known druggable targets in cells are membrane proteins and ~50% of drugs target membrane proteins (Lappano and Maggiolini, 2011, Sameshima et al., 2019, Terstappen and Reggiani, 2001b). In this regard, no discussion of biological membranes would be complete without mentioning ATP-binding cassette transporters (ABC). In all prokaryotes, as well as, in plants, fungi, yeast, and animals, ABC transporters are ubiquitous membrane-embedded proteins. The human genome has 49 ABC genes that are organised into eight subfamilies and given names through divergent evolution (Vasiliou et al., 2009). Given that at least 11 of these genes have already been linked to serious hereditary disorders, this emphasises the significance of ABC genes (see chapter 01). ABC transporters also play a role in the transit of most medicines and their metabolites through cell surface and cellular organelle membranes; therefore, defects in these genes can have a significant impact on cancer therapy, pharmacokinetics, and a variety of pharmacogenetic disorders (Vasiliou et al., 2009). Therefore, functional and structural research on ABC proteins may be essential to advance our understanding of the protein's structural and molecular characteristics. This will have a huge impact on numerous downstream research projects like protein modification and drug design based on structural principles. It is possible to resolve molecular structures to an atomic level using complex techniques like X-ray crystallography and cryoelectron microscopy (cryo-EM). However, it is challenging to express and purify membrane proteins (Lacapere et al., 2007, Carpenter et al., 2008). All ABC transporters structure on protein data bank is given in Appendix Table 9.1.

9.1 Selection of protein targets for high-resolution structural studies

Target selection for structural studies is critical. The selection of target proteins for high-resolution structural studies can be aided by *in-silico* investigations *i.e.* the success of crystallisation can be aided using crystallisation prediction methods before starting X-ray crystallography research with a particular protein. This study was initially aiming to use a crystallisation prediction tool to screen the entire ABC family except ABCE and ABCF in order to choose targets for crystal trials. The bioinformatics screening did not include the subfamilies ABCE and ABCF (because these proteins are engaged in DNA repair and chromosomal maintenance rather than transportation) (Rees et al., 2009, Higgins, 1992). For this purpose, TMCrys server applied to screen a number of orthologs of the eukaryotic ATP

binding cassette (ABC) family (see chapter 03). The TMCrys server is chosen since it is designed explicitly for transmembrane proteins. Ortholog selection for high-resolution structural study was made applying TMCrys. Four ABC transporters namely ATP binding cassette subfamily B, member 5 (ABCB5), ATP binding cassette family B, member 6 (ABCB6), ATP binding cassette family G, member 1 (ABCG1), and ATP binding cassette family G, member 4 (ABCG4) were selected for high-resolution structural studies. The selection of these ABC proteins was made for two reasons: (1) none of the ABCB5, ABCB6, ABCG1, or ABCG4 structures were present in the Protein Data Bank (PDB) when the project was started; and (2) the therapeutic importance of these proteins. Afterward, orthologs of these representative transporters with a high likelihood of expression, solubilisation, and purification were selected based on the bioinformatics scoring. For ABCB5, mouse ortholog, giant panda for ABCB6, little brown bat for ABCG1, and mouse for ABCG4 were selected for study. The amino acid identity of these orthologs relative to the human protein is given in table 9.1.

Table 9.1: Sequence identity of selected orthologs to human protein calculated by NCBI BLAST.

Protein name	Percent Identity to human protein
mABCB5	78.17 %
gpABCB6	87.63 %
bbABCG1	87.78%
mABCG4	96.59%

The strategy of using an ortholog was chosen because it could result in a protein sequence that is more stable, and accessible to expression, solubilisation, purification, and structural and functional research than the human protein due to natural divergence in amino acids.

9.2 Selection of *Saccharomyces cerevisiae* as an expression system

In this study, we have used a yeast expression system, in the past membrane proteins have been expressed in yeast (Vieira Gomes et al., 2018). Yeast cells possess many important features that are necessary for the production of eukaryotic membrane proteins. Yeast cells have the ability to do post-translational modifications. They also have highly regulated protein quality control system in the endoplasmic reticulum (ER) which ensures that only correctly folded protein exit the ER. Misfolded proteins are degraded by Endoplasmic-reticulum-associated degradation pathways (Brodsky and Skach, 2011). Absence of these

mechanisms in prokaryotes, therefore, makes the expression of eukaryotic membrane protein in the bacterial system very difficult. *Pichia pastoris* has the ability to produce large biomass when compared to *S. cervisae* but cloning in *pichia* is tedious and time-consuming. It is observed that yeast cells can produce enough proteins for structural studies (Jidenko et al., 2005). Yeast cultures are reasonably cost-effective and fast to grow in the lab. Yeast also has the ability to express high levels of recombinant proteins at lower temperatures which allows protein folding to occur at a steadier rate.

9.3 Construct design

Codon optimized mABCB5, gpABCB6 bbABCG1, and mABCG4 DNA constructs for *Saccharomyces cerevisiae* were synthesized. A yeast Kozak-like sequence (AAAACA) was also added before the start codon ATG (Kozak, 1986). It has been shown that employing a codon-optimized CFTR gene increases expression by around 4-fold when compared to wild-type CFTR and that adding a yeast Kozak-like sequence before the CFTR gene start codon increases expression by about 28-fold when compared to the gene lacking the Kozak-like sequence (Rimington1 et al., 2018). Furthermore, it has been demonstrated that the use of Kozak-like sequences enhances the production of various other recombinant proteins in heterologous expression systems (Alonso et al., 2002, Jäger et al., 2013). The green fluorescent protein (GFP) sequence and an octa-His tag are fused at the C-terminus of the mABCB5 and gpABCB6 constructs, whereas they are located at the N-terminus of bbABCG1 and mABCG4 due to reverse topology (see chapter 02). The GFP makes it reasonably simple to follow the expression and purification of proteins. The constructs were transformed into the *S. cerevisiae* strain, FGY217 (originally provided by Dr. David Drew, Imperial College of London) in which expression of mABCB5, gpABCB6, bbABCG1, and mABCG4 was under the control of the galactose inducible promoter, GAL1, which is heavily repressed in the presence of glucose (Drew et al., 2008). As a result, the induction time of recombinant proteins can be tightly controlled (Drew et al., 2008) (see chapter 04).

9.4 Expression and purification of target proteins

All of these four target proteins *i.e.* mABCB5, gpABCB6, bbABCG1, and mABCG4 were expressed in yeast using optimized yeast culture conditions to obtain a high level of expression (Rimington et al., 2018). Establishing the optimal post-induction cell harvest time and the optimal cell density at induction were important parameters for getting a good

yield (see chapter 4). The fluorescence microscopy data was very encouraging which showed that all of these four proteins had been correctly folded and successfully expressed in yeast and can be processed further for purification. The GFP fluorescence was more diffused and localised in mABCB5 and gpABCB6 expressing yeast cells than it was in bbABCG1 and mABCG4, which displayed punctate GFP localization. When comparing purity, bbABCG1 and mABCG4 are purer than mABCB5 and gpABCB6. It could be inferred from this that, in contrast to localised and diffused proteins, punctate localised proteins in yeast cells are well-behaved in purification process. In our study, it appears that purifying mABCB5 is somewhat challenging. All mABCB5 purification batches produced contaminated mABCB5 (see chapter 05). It might be because ABCB5 is a full-length transporters and needs more stringent conditions for purification. Some critical downstream purification steps that are specific to the *S. cerevisae* expression system, such as cell lysis and microsomes purification. In this study, we used the bead-beating method which is the most practical and effective in terms of equipment cost and ease. DDM was used for protein solubilisation and afterwards in all purification steps. The selection of detergents for solubilisation and purification is a critical step in the purification of any membrane protein. The reason for using DDM in this study is because Dodecyl-β-D-maltoside (DDM) is a non-ionic detergent that has been extensively used for both structural and functional studies of membrane proteins, which is why it was chosen for this study. (Oldham and Chen, 2011, Pinkett et al., 2007, Dawson and Locher, 2006, Gerber et al., 2008, Ward et al., 2007, Kadaba et al., 2008, Matsson et al., 2009, Lingam et al., 2017).

9.5 Protein characterisation and structural studies

Different approaches were used to characterise the biochemical and biological characteristics of all four purified proteins. Dynamic light scattering (DLS), which is regarded as a non-destructive method and suitable for membrane protein study because it takes comparatively less sample to be studied, was used to confirm monodispersity and aggregation of enriched proteins (Murphy, 1997). The DLS profiles indicated that samples of all four e proteins in apo-state had a mixed population of monomers with ~10 nm hydrodynamic diameter and big particles (such as oligomers or aggregates) with ~100 nm hydrodynamic diameter (see details in section 6.4). Additionally, the presence of nucleotides was tested for any effects on the DLS measurement, and this revealed that there was no noticeable impact on mass and intensity distribution of protein particles in the presence of

nucleotides. This might be because the particle population that underwent conformational changes in the presence of nucleotides may have been limited (or absent altogether), and DLS was unable to identify any alteration in the hydrodynamic diameter of protein particles. Furthermore, the DLS signal was likely dominated by a signal from the larger population. Our findings were consistent with DLS research on P-gp, which reported that no nucleotide-induced conformational changes could be detected by DLS (Thonghin et al., 2018b). However, there hasn't been any information reported on using this method to analyse the aggregation, monodispersity, and conformational changes of mABCG4, gpABCB6, bbABCG1, and mABCB5.

The ATPase activity of mABCB5, gpABCB6, bbABCG1, and mABCG4 may also use as a suitable benchmark of the quality of a purified protein sample (Pollock et al., 2014b, Thonghin, 2019) where substrate-stimulated rates of ATPase activity may be interpreted as a sign of functional protein. In this study, ATPase activity was determined *via* Chifflet assay (Chifflet et al., 1988). All of the four target proteins (mABCB5, gpABCB6, bbABCG1, and mABCG4) exhibited basal as well as substrate-stimulated ATPase activities in a detergent environment (discussed in section 6.2). The basal and substrate-stimulated ATPase activity of some target proteins appeared to be comparable to that of several previous studies (e.g. gpABCB6 basal activity of 24.83 nmol/min/mg and mABCG4 basal activity of 21.69 nmol/min/mg) (Song et al., 2021, Cserepes et al., 2004a, Kim and Chen, 2018, Loo et al., 2012). This indicated that all of the four purified proteins exhibited activity under the tested conditions to a degree comparable to that of previously published research (Kim and Chen, 2018, Song et al., 2021, Thonghin, 2019). These results were a useful check that the proteins expressed in yeast and purified in DDM were active.

mABCB5, gpABCB6, bbABCG1, and mABCG4 protein samples were also examined for their thermal stability by intrinsic tryptophan fluorescence and a fluorescent assay with a thiol-specific dye, CPM, in the absence and presence of nucleotides (Hannemann et al., 2002, Alexandrov et al., 2008). It has been demonstrated that nucleotide binding stabilises NBDs; therefore, a direct stabilising action of nucleotides on the full-length protein may take place in a conformation-independent manner (Altieri et al., 2008). The CPM assay showed that the melting temperatures of the proteins increased in the presence of nucleotides. This indicates that additional energy is needed for the protein to unfold to the same extent compared to the apo-form, indicating an improvement in protein stability. This is consistent with thermal studies conducted on P-gp where a T_m increase upon nucleotide binding was reported which suggest the stabilising effect of nucleotide on P-gp (Celej et al., 2003, Pollock

et al., 2014b). There were small variations in the way target proteins were first labelled by CPM and in their unfolding transition patterns, highlighting some variation in their structural integrity (see details in chapter 06).

Tyrosine kinase inhibitors were shown to have stabilising effects on mABCG4 according to a membrane thermal shift experiment (see section 6.5.3). In the future this assay could be combined with docking studies to investigate this stabilising mechanism in more detail. After identifying the precise residues (by molecular docking) that participate the interaction of tyrosine kinase inhibitors with mABCG4, mutations might be produced in the lab, and afterwards can be verified by comparing to wild type mABCG4 using the same membrane thermal shift assay.

Low-resolution methods such as small-angle X-ray scattering were explored to shed light on the structural characteristics. Unfortunately, aggregation was seen in our measurements, and the data were unsuitable for ab-initio modelling (see details in chapter 05). With improvement in preparation, ab-initio models can be made which can depict the gross landscape of molecular dynamics in an aqueous solution. In cryoEM experiments conducted on ABCG4, the difficulty in preparing a high-quality grid was mainly to the removal of micelles and detergent. We found a concentration of micelles during cryo-EM, despite the fact that cut-off filters are often effective at eliminating detergents during buffer exchange.

9.6 Conclusion

Overall, the bioinformatics approach we used in this study was productive during all of the preliminary stages of research. The transformation and expression of each chosen ortholog was successful in yeast cells. All target proteins were purified and characterised which demonstrates that all purified proteins were functionally active with Vmax in the following order:

bbABCG1> gpABCB6> mABCG4> mABCB5

mABCB5 and gpABCB6 appeared to be less pure as compared to bbABCG1 and mABCG4. Furthermore, bbABCG1 size exclusion chromatogram revealed its elution in void volume indicating that the majority of the protein was aggregated. This left us with only one target protein, mABCG4, for high resolution structural study. Although no structure has been determined in this study, it does provide information on the expression of mABCB5, gpABCB6, bbABCG1, and mABCG4 proteins in yeast cells. Additionally, it sheds light on

the purification conditions of these proteins as well as adding knowledge on the behaviour of these proteins which could serve as the basis for further research in this area.

9.7 Future work

The research in this thesis describes the use of computational tool for screening the target proteins for high-resolution structural studies and afterwards expression and purification of mABCB5, gpABCB6, bbABCG1, and mABCG4 proteins in FGY217 yeast strain for functional and structural analysis. The adoption of different yeast strain for protein expression is another potential direction for research. Using yeast strains that have diminished, or are even defective in components of the degradation and/or aggregation pathways may help to further optimise protein expression. There are many more tags that might be evaluated in order to purify the protein; the ones that were investigated in this thesis are just a good subset of all conceivable affinity tags. Different detergents can be screened for microsomes solubilisation that may improve the protein yield. Purification protocol can be improved or purification can be performed using GFP resin that might lead to greater purity as well as yield of protein, that would enable a greater range of biochemical and biophysical analysis to be carried out, so expanding our knowledge.

CPM thermal stability assay described in chapter 06 of the thesis could form the platform for screening different compounds for all of the proteins that have stabilising effect on purified proteins that could go into the crystal trials in future or cryo-EM studies. With this technique, numerous compounds could be screened simultaneously utilising direct comparisons to controls in multi-well plates, and only tiny amounts of protein is needed. Similarly, all proteins can be subjected to the membrane thermal shift experiment (see section 6.5.3) for compound screening. The advantage of this technique is this does not require purified protein and can be performed on the protein containing membrane vesicles. In terms of high-resolution structural studies, crystal experiments can be carried out after TEV treatment and obtaining a protein that is ~99.95% homogeneous. The crystallisation process still requires intensive optimization of a number of variables, including detergent types and contents, protein concentrations, protein-to-lipid ratios, and screening reagents. Cryo-EM experiments were carried out primarily on mABCG4 in this study, but they can be done on the other target proteins as well after enhancing the protein purity level. Furthermore, for mABCG4, following the adjustment of the membrane scaffold protein to mABCG4 ratio for nanodisc production, cryo experiments can be carried out.

REFERENCES

ABDUL JALIL, Y., RITZ, V., JAKIMENKO, A., SCHMITZ-SALUE, C., SIEBERT, H., AWUAH, D., KOTTHAUS, A., KIETZMANN, T., ZIEMANN, C. & HIRSCH-ERNST, K. I. 2008. Vesicular localization of the rat ATP-binding cassette half-transporter rAbcb6. *American Journal of Physiology-Cell Physiology,* 294**,** C579-C590.

ABRÀMOFF, M. D., MAGALHÃES, P. J. & RAM, S. J. 2004. Image processing with ImageJ. *Biophotonics international,* 11**,** 36-42.

ADAM, H. 1965. Adenosine-5′-triphosphate: Determination with phosphoglycerate kinase. *Methods of enzymatic analysis.* Elsevier.

AIVALIOTIS, M., SAMOLIS, P., NEOFOTISTOU, E., REMIGY, H., RIZOS, A. K. & TSIOTIS, G. 2003. Molecular size determination of a membrane protein in surfactants by light scattering. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1615**,** 69-76.

ALEXANDROV, A. I., MILENI, M., CHIEN, E. Y., HANSON, M. A. & STEVENS, R. C. 2008. Microscale fluorescent thermal stability assay for membrane proteins. *Structure,* 16**,** 351-359.

ALMEIDA, J. G., PRETO, A. J., KOUKOS, P. I., BONVIN, A. M. & MOREIRA, I. S. 2017. Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1859**,** 2021-2039.

ALONSO, S., IZETA, A., SOLA, I. & ENJUANES, L. 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *Journal of virology,* 76**,** 1293-1308.

ALTIERI, S. L., CLAYTON, G. M., SILVERMAN, W. R., OLIVARES, A. O., ENRIQUE, M., THOMAS, L. R. & MORAIS-CABRAL, J. H. 2008. Structural and energetic analysis of activation by a cyclic nucleotide binding domain. *Journal of molecular biology,* 381**,** 655-669.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research,* 25**,** 3389-3402.

ANANDAN, A. & VRIELINK, A. 2016. Detergents in membrane protein purification and crystallisation. *The next generation in membrane protein structure determination.* Springer.

ANDOLFO, I., ALPER, S. L., DELAUNAY, J., AURIEMMA, C., RUSSO, R., ASCI, R., ESPOSITO, M. R., SHARMA, A. K., SHMUKLER, B. E. & BRUGNARA, C. 2013. Missense mutations in the ABCB6 transporter cause dominant familialpseudohyperkalemia. *American Journal of Hematology,* 88**,** 66-72.

ANDOLFO, I., RUSSO, R., MANNA, F., DE ROSA, G., GAMBALE, A., ZOUWAIL, S., DETTA, N., PARDO, C. L., ALPER, S. L. & BRUGNARA, C. 2016. Functional characterization of novel ABCB6 mutations and their clinical implications in familial pseudohyperkalemia. *Haematologica,* 101**,** 909.

ANDRÉLL, J. & TATE, C. G. 2013. Overexpression of membrane proteins in mammalian cells for structural studies. *Molecular membrane biology,* 30**,** 52-63.

ANREDDY, N., GUPTA, P., KATHAWALA, R. J., PATEL, A., WURPEL, J. N. & CHEN, Z.-S. 2014. Tyrosine kinase inhibitors as reversal agents for ABC transporter mediated drug resistance. *Molecules,* 19**,** 13848-13877.

ARACHEA, B. T., SUN, Z., POTENTE, N., MALIK, R., ISAILOVIC, D. & VIOLA, R. E. 2012. Detergent selection for enhanced extraction of membrane proteins. *Protein expression and purification,* 86**,** 12-20.

ARCE, C., PÉREZ-PLASENCIA, C., GONZÁLEZ-FIERRO, A., DE LA CRUZ-HERNÁNDEZ, E., REVILLA-VÁZQUEZ, A., CHÁVEZ-BLANCO, A., TREJO-BECERRIL, C., PÉREZ-CÁRDENAS, E., TAJA-CHAYEB, L. & BARGALLO, E. 2006. A proof-of-principle study of epigenetic therapy added to neoadjuvant doxorubicin cyclophosphamide for locally advanced breast cancer. *PloS one,* 1**,** e98.

ASHANI, Y. & CATRAVAS, G. N. 1980. Highly reactive impurities in Triton X-100 and Brij 35: partial characterization and removal. *Analytical Biochemistry,* 109**,** 55-62.

ASHOK, Y., NANEKAR, R. & JAAKOLA, V.-P. 2015. Defining thermostability of membrane proteins by western blotting. *Protein Engineering, Design & Selection,* 28**,** 539-542.

AWAYN, N., ROSENBERG, M., KAMIS, A., ALEKSANDROV, L., RIORDAN, J. & FORD, R. 2005. Crystallographic and single-particle analyses of native-and nucleotide-bound forms of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Portland Press Ltd.

AZEGAMI, M. & IWAI, K. 1964. Specific Modification of Nucleic Aicds and Their Constituents with Trinitrophenyl Group. *The Journal of Biochemistry,* 55**,** 346-348.

BAGSHAW, R. D., MAHURAN, D. J. & CALLAHAN, J. W. 2005. A proteomic analysis of lysosomal integral membrane proteins reveals the diverse composition of the organelle. *Molecular & Cellular Proteomics,* 4**,** 133-143.

BAI, J., SWARTZ, D. J., PROTASEVICH, I. I., BROUILLETTE, C. G., HARRELL, P. M., HILDEBRANDT, E., GASSER, B., MATTANOVICH, D., WARD, A. & CHANG, G. 2011. A gene optimization strategy that enhances production of fully functional P-glycoprotein in Pichia pastoris. *PloS one,* 6**,** e22577.

BAN, C. & YANG, W. 1998. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell,* 95**,** 541-552.

BARITAKI, S., TZANAKAKIS, G. N., ALIFRAGIS, J., ZAFIROPOULOS, A., TASHMUKHAMEDOV, R. I., TSATSAKIS, A., SHTILMAN, M. I., RIZOS, A. K. & KRAMBOVITIS, E. 2002. Light scattering and in vitro biocompatibility studies of poly (vinyl pyrrolidone) derivatives with amino-acid-dependent groups. *Journal of Biomedical Materials Research: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials,* 63**,** 830-837.

BASAN, M., HUI, S., OKANO, H., ZHANG, Z., SHEN, Y., WILLIAMSON, J. R. & HWA, T. 2015. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature,* 528**,** 99-104.

BAYBURT, T. H., GRINKOVA, Y. V. & SLIGAR, S. G. 2002. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano letters,* 2**,** 853-856.

BAYBURT, T. H. & SLIGAR, S. G. 2010. Membrane protein assembly into Nanodiscs. *FEBS letters,* 584**,** 1721-1727.

BAYKOV, A., EVTUSHENKO, O. & AVAEVA, S. 1988. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Analytical biochemistry,* 171**,** 266-270.

BEIS, K. 2015. Structural basis for the mechanism of ABC transporters. *Biochemical Society Transactions,* 43**,** 889-893.

BERETTA, G. L., CASSINELLI, G., PENNATI, M., ZUCO, V. & GATTI, L. 2017. Overcoming ABC transporter-mediated multidrug resistance: The dual role of

tyrosine kinase inhibitors as multitargeting agents. *European journal of medicinal chemistry,* 142**,** 271-289.

BERNADÓ, P., MYLONAS, E., PETOUKHOV, M. V., BLACKLEDGE, M. & SVERGUN, D. I. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society,* 129**,** 5656-5664.

BERNAUDAT, F., FRELET-BARRAND, A., POCHON, N., DEMENTIN, S., HIVIN, P., BOUTIGNY, S., RIOUX, J.-B., SALVI, D., SEIGNEURIN-BERNY, D. & RICHAUD, P. 2011. Heterologous expression of membrane proteins: choosing the appropriate host. *PloS one,* 6**,** e29191.

BIRCH, J., CHERUVARA, H., GAMAGE, N., HARRISON, P. J., LITHGO, R. & QUIGLEY, A. 2020. Changes in membrane protein structural biology. *Biology,* 9**,** 401.

BLEES, A., REICHEL, K., TROWITZSCH, S., FISETTE, O., BOCK, C., ABELE, R., HUMMER, G., SCHÄFER, L. V. & TAMPÉ, R. 2015. Assembly of the MHC I peptide-loading complex determined by a conserved ionic lock-switch. *Scientific reports,* 5**,** 1-11.

BOJANIC, D. D., TARR, P. T., GALE, G. D., SMITH, D. J., BOK, D., CHEN, B., NUSINOWITZ, S., LÖVGREN-SANDBLOM, A., BJÖRKHEM, I. & EDWARDS, P. A. 2010. Differential expression and function of ABCG1 and ABCG4 during development and aging [S]. *Journal of lipid research,* 51**,** 169-181.

BOOTH, D. S., AVILA-SAKAR, A. & CHENG, Y. 2011. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *JoVE (Journal of Visualized Experiments)***,** e3227.

BOSWELL-CASTEEL, R. C., FUKUDA, Y. & SCHUETZ, J. D. 2018. ABCB6, an ABC transporter impacting drug response and disease. *The AAPS journal,* 20**,** 1-10.

BREA-CALVO, G., SIENDONES, E., SÁNCHEZ-ALCÁZAR, J. A., DE CABO, R. & NAVAS, P. 2009. Cell survival from chemotherapy depends on NF-κB transcriptional up-regulation of coenzyme Q biosynthesis. *PloS one,* 4**,** e5301.

BRENNER, S. & HORNE, R. 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et biophysica acta,* 34**,** 103-110.

BRITTON, Z., YOUNG, C., CAN, Ö., MCNEELY, P., NARANJO, A. & ROBINSON, A. S. 2011. Membrane protein expression in Saccharomyces cerevisiae. *Production of Membrane Proteins: Strategies for Expression and Isolation***,** 37-73.

BRODSKY, J. L. & SKACH, W. R. 2011. Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems. *Current opinion in cell biology,* 23**,** 464-475.

CAFFREY, M. 2003. Membrane protein crystallization. *Journal of structural biology,* 142**,** 108-132.

CAI, S., HE, F., SAMRA, H. S., DE LA MAZA, L. M., BOTTAZZI, M. E., JOSHI, S. B. & MIDDAUGH, C. R. 2009. Biophysical and stabilization studies of the Chlamydia trachomatis mouse pneumonitis major outer membrane protein. *Molecular pharmaceutics,* 6**,** 1553-1561.

CAMPBELL, J. D., DEOL, S. S., ASHCROFT, F. M., KERR, I. D. & SANSOM, M. S. 2004. Nucleotide-dependent conformational changes in HisP: molecular dynamics simulations of an ABC transporter nucleotide-binding domain. *Biophysical journal,* 87**,** 3703-3715.

CARPENTER, E. P., BEIS, K., CAMERON, A. D. & IWATA, S. 2008. Overcoming the challenges of membrane protein crystallography. *Current opinion in structural biology,* 18**,** 581-586.

CELEJ, M. S., MONTICH, G. G. & FIDELIO, G. D. 2003. Protein stability induced by ligand binding correlates with changes in protein flexibility. *Protein science,* 12**,** 1496-1506.

CHAVAN, H., KHAN, M. M. T., TEGOS, G. & KRISHNAMURTHY, P. 2013. Efficient purification and reconstitution of ATP binding cassette transporter B6 (ABCB6) for functional and structural studies. *Journal of Biological Chemistry,* 288**,** 22658-22669.

CHAVAN, H., ORUGANTI, M. & KRISHNAMURTHY, P. 2011. The ATP-binding cassette transporter ABCB6 is induced by arsenic and protects against arsenic cytotoxicity. *Toxicological Sciences,* 120**,** 519-528.

CHEN, K. G., SZAKÁCS, G., ANNEREAU, J. P., ROUZAUD, F., LIANG, X. J., VALENCIA, J. C., NAGINENI, C. N., HOOKS, J. J., HEARING, V. J. & GOTTESMAN, M. M. 2005. Principal expression of two mRNA isoforms (ABCB 5α and ABCB 5β) of the ATP-binding cassette transporter gene ABCB 5 in melanoma cells and melanocytes. *Pigment cell research,* 18**,** 102-112.

CHEN, Z.-P., XU, D., WANG, L., MAO, Y.-X., LI, Y., CHENG, M.-T., ZHOU, C.-Z., HOU, W.-T. & CHEN, Y. 2022. Structural basis of substrate recognition and translocation by human very long-chain fatty acid transporter ABCD1. *Nature Communications,* 13**,** 3299.

CHEUNG, P. F. Y., CHENG, C. K. C., WONG, N. C. L., HO, J. C. Y., YIP, C. W., LUI, V. C. H., CHEUNG, A. N. Y., FAN, S. T. & CHEUNG, S. T. 2011a. Granulin-epithelin precursor is an oncofetal protein defining hepatic cancer stem cells. *PloS one,* 6**,** e28246.

CHEUNG, S. T., CHEUNG, P. F., CHENG, C. K., WONG, N. C. & FAN, S. T. 2011b. Granulin-epithelin precursor and ATP-dependent binding cassette (ABC) B5 regulate liver cancer cell chemoresistance. *Gastroenterology,* 140**,** 344-355. e2.

CHIFFLET, S., TORRIGLIA, A., CHIESA, R. & TOLOSA, S. 1988. A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. *Analytical biochemistry,* 168**,** 1-4.

CHIU, W., MCGOUGH, A., SHERMAN, M. B. & SCHMID, M. F. 1999. High-resolution electron cryomicroscopy of macromolecular assemblies. *Trends in Cell Biology,* 9**,** 154-159.

COWIESON, N. P., EDWARDS-GAYLE, C. J., INOUE, K., KHUNTI, N. S., DOUTCH, J., WILLIAMS, E., DANIELS, S., PREECE, G., KRUMPA, N. A. & SUTTER, J. P. 2020. Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source. *Journal of Synchrotron Radiation,* 27**,** 1438-1446.

CRAMER, P. 2021. AlphaFold2 and the future of structural biology. *Nature Structural & Molecular Biology,* 28**,** 704-705.

CSEREPES, J., SZENTPÉTERY, Z., SERES, L., ÖZVEGY-LACZKA, C., LANGMANN, T., SCHMITZ, G., GLAVINAS, H., KLEIN, I., HOMOLYA, L. & VÁRADI, A. 2004a. Functional expression and characterization of the human ABCG1 and ABCG4 proteins: indications for heterodimerization. *Biochemical and biophysical research communications,* 320**,** 860-867.

CSEREPES, J., SZENTPÉTERY, Z., SERES, L., ÖZVEGY-LACZKA, C., LANGMANN, T., SCHMITZ, G., GLAVINAS, H., KLEIN, I., HOMOLYA, L., VÁRADI, A., SARKADI, B. & ELKIND, N. B. 2004b. Functional expression and characterization of the human ABCG1 and ABCG4 proteins: indications for heterodimerization. *Biochemical and Biophysical Research Communications,* 320**,** 860-867.

CUATRECASAS, P. 1974. Membrane receptors. *Annual review of biochemistry,* 43**,** 169-214.

CUI, Y.-X., XIA, X.-Y., ZHOU, Y., GAO, L., SHANG, X.-J., NI, T., WANG, W.-P., FAN, X.-B., YIN, H.-L. & JIANG, S.-J. 2013. Novel mutations of ABCB6 associated with autosomal dominant dyschromatosis universalis hereditaria. *PLoS One,* 8**,** e79808.

DASSA, E. & BOUIGE, P. 2001. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Research in microbiology,* 152**,** 211-229.

DAVEY, J. 2004. G-protein-coupled receptors: new approaches to maximise the impact of GPCRS in drug discovery. *Expert opinion on therapeutic targets,* 8**,** 165-170.

DAWSON, R. J., HOLLENSTEIN, K. & LOCHER, K. P. 2007. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Molecular microbiology,* 65**,** 250-257.

DAWSON, R. J. & LOCHER, K. P. 2006. Structure of a bacterial multidrug ABC transporter. *Nature,* 443**,** 180-185.

DE AGUIAR VALLIM, T. Q., LEE, E., MERRIOTT, D. J., GOULBOURNE, C. N., CHENG, J., CHENG, A., GONEN, A., ALLEN, R. M., PALLADINO, E. N. & FORD, D. A. 2017. ABCG1 regulates pulmonary surfactant metabolism in mice and men [S]. *Journal of lipid research,* 58**,** 941-954.

DEAN, M., HAMON, Y. & CHIMINI, G. 2001. The human ATP-binding cassette (ABC) transporter superfamily. *Journal of lipid research,* 42**,** 1007-1017.

DECERO, S. A., WINSLOW, C. H. & COBURN, J. 2020. Method to Overcome Inefficiencies in Site-Directed Mutagenesis of A/T-Rich DNA. *Journal of Biomolecular Techniques: JBT,* 31**,** 94.

DEMARIA, A. N. 2003. A structure for deoxyribose nucleic acid. American College of Cardiology Foundation Washington, DC.

DEMIREL, Ö., BANGERT, I., TAMPÉ, R. & ABELE, R. 2010. Tuning the Cellular Trafficking of the Lysosomal Peptide Transporter TAPL by its N-terminal Domain. *Traffic,* 11**,** 383-393.

DENISOV, I. G., GRINKOVA, Y. V., LAZARIDES, A. A. & SLIGAR, S. G. 2004. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *Journal of the American Chemical Society,* 126**,** 3477-3487.

DO, T. M., NOEL-HUDSON, M.-S., RIBES, S., BESENGEZ, C., SMIRNOVA, M., CISTERNINO, S., BUYSE, M., CALON, F., CHIMINI, G. & CHACUN, H. 2012. ABCG2-and ABCG4-mediated efflux of amyloid-β peptide 1-40 at the mouse blood-brain barrier. *Journal of Alzheimer's Disease,* 30**,** 155-166.

DODACKI, A., WORTMAN, M., SAUBAMÉA, B., CHASSEIGNEAUX, S., NICOLIC, S., PRINCE, N., LOCHUS, M., RAVEU, A.-L., DECLÈVES, X. & SCHERRMANN, J.-M. 2017. Expression and function of Abcg4 in the mouse blood-brain barrier: role in restricting the brain entry of amyloid-β peptide. *Scientific reports,* 7**,** 1-14.

DONIACH, S. 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews,* 101**,** 1763-1778.

DREW, D., LERCH, M., KUNJI, E., SLOTBOOM, D.-J. & DE GIER, J.-W. 2006. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature methods,* 3**,** 303-313.

DREW, D., NEWSTEAD, S., SONODA, Y., KIM, H., VON HEIJNE, G. & IWATA, S. 2008. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nature protocols,* 3**,** 784-798.

DREW, D. E., VON HEIJNE, G., NORDLUND, P. & DE GIER, J.-W. L. 2001. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS letters,* 507**,** 220-224.

DUPEUX, F., RÖWER, M., SEROUL, G., BLOT, D. & MÁRQUEZ, J. A. 2011. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallographica Section D: Biological Crystallography,* 67**,** 915-919.

EARL, L. A., FALCONIERI, V., MILNE, J. L. & SUBRAMANIAM, S. 2017. Cryo-EM: beyond the microscope. *Current opinion in structural biology,* 46**,** 71-78.

EMADI-KONJIN, H.-P., ZHANG, H., ANANDAN, V., SUN, D., SCHUETZ, J. & FURUYA, K. N. 2002. Isolation of a genomic clone containing the promoter region of the human ATP binding cassette (ABC) transporter, ABCB6. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1574**,** 117-130.

ENGEL, T., LORKOWSKI, S., LUEKEN, A., RUST, S., SCHLÜTER, B., BERGER, G., CULLEN, P. & ASSMANN, G. 2001. The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-derived macrophages. *Biochemical and biophysical research communications,* 288**,** 483-488.

ESWAR, N., ERAMIAN, D., WEBB, B., SHEN, M.-Y. & SALI, A. 2008. Protein structure modeling with MODELLER. *Structural proteomics.* Springer.

ESWAR, N., WEBB, B., MARTI-RENOM, M. A., MADHUSUDHAN, M., ERAMIAN, D., SHEN, M. Y., PIEPER, U. & SALI, A. 2006. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics,* 15**,** 5.6. 1-5.6. 30.

FERNÁNDEZ, F. J., LÓPEZ-ESTEPA, M., QUEROL-GARCÍA, J. & VEGA, M. C. 2016. Production of protein complexes in non-methylotrophic and methylotrophic Yeasts: nonmethylotrophic and methylotrophic Yeasts. *Advanced Technologies for Protein Complex Production and Characterization***,** 137-153.

FIGLER, R. A., OMOTE, H., NAKAMOTO, R. K. & AL-SHAWI, M. K. 2000. Use of chemical chaperones in the yeast Saccharomyces cerevisiae to enhance heterologous membrane protein expression: high-yield expression and purification of human P-glycoprotein. *Archives of biochemistry and biophysics,* 376**,** 34-46.

FOCACCETTI, C., BRUNO, A., MAGNANI, E., BARTOLINI, D., PRINCIPI, E., DALLAGLIO, K., BUCCI, E. O., FINZI, G., SESSA, F. & NOONAN, D. M. 2015. Effects of 5-fluorouracil on morphology, cell cycle, proliferation, apoptosis, autophagy and ROS production in endothelial cells and cardiomyocytes. *PloS one,* 10**,** e0115686.

FÖRSTER, S., APOSTOL, L. & BRAS, W. 2010. Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography,* 43**,** 639-646.

FRANCIS, D. M. & PAGE, R. 2010. Strategies to optimize protein expression in E. coli. *Current protocols in protein science,* 61**,** 5.24. 1-5.24. 29.

FRANK, G. A., SHUKLA, S., RAO, P., BORGNIA, M. J., BARTESAGHI, A., MERK, A., MOBIN, A., ESSER, L., EARL, L. A. & GOTTESMAN, M. M. 2016. Cryo-EM analysis of the conformational landscape of human P-glycoprotein (ABCB1) during its catalytic cycle. *Molecular pharmacology,* 90**,** 35-41.

FRANK, J. 2006. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*, Oxford university press.

FRANK, N. Y., MARGARYAN, A., HUANG, Y., SCHATTON, T., WAAGA-GASSER, A. M., GASSER, M., SAYEGH, M. H., SADEE, W. & FRANK, M. H. 2005. ABCB5-mediated doxorubicin transport and chemoresistance in human malignant melanoma. *Cancer research,* 65**,** 4320-4333.

FRANK, N. Y., PENDSE, S. S., LAPCHAK, P. H., MARGARYAN, A., SHLAIN, D., DOEING, C., SAYEGH, M. H. & FRANK, M. H. 2003. Regulation of progenitor cell fusion by ABCB5 P-glycoprotein, a novel human ATP-binding cassette transporter. *Journal of Biological Chemistry,* 278**,** 47156-47165.

FRANKE, D., PETOUKHOV, M., KONAREV, P., PANJKOVICH, A., TUUKKANEN, A., MERTENS, H., KIKHNEY, A., HAJIZADEH, N., FRANKLIN, J. & JEFFRIES, C. 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography,* 50**,** 1212-1225.

FREY, S., SOTT, K., SMEDH, M., MILLAT, T., DAHL, P., WOLKENHAUER, O. & GOKSÖR, M. 2011. A mathematical analysis of nuclear intensity dynamics for Mig1-GFP under consideration of bleaching effects and background noise in Saccharomyces cerevisiae. *Molecular BioSystems,* 7**,** 215-223.

FUKUDA, Y., CHEONG, P. L., LYNCH, J., BRIGHTON, C., FRASE, S., KARGAS, V., RAMPERSAUD, E., WANG, Y., SANKARAN, V. G. & YU, B. 2016. The severity of hereditary porphyria is modulated by the porphyrin exporter and Lan antigen ABCB6. *Nature communications,* 7**,** 1-9.

GARIDEL, P., HEGYI, M., BASSARAB, S. & WEICHEL, M. 2008. A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy. *Biotechnology Journal: Healthcare Nutrition Technology,* 3**,** 1201-1211.

GELISSEN, I. C., CARTLAND, S., BROWN, A. J., SANDOVAL, C., KIM, M., DINNES, D. L., LEE, Y., HSIEH, V., GAUS, K. & KRITHARIDES, L. 2010. Expression and stability of two isoforms of ABCG1 in human vascular cells. *Atherosclerosis,* 208**,** 75-82.

GELISSEN, I. C., SHARPE, L. J., SANDOVAL, C., RAO, G., KOCKX, M., KRITHARIDES, L., JESSUP, W. & BROWN, A. J. 2012. Protein kinase A modulates the activity of a major human isoform of ABCG1. *Journal of lipid research,* 53**,** 2133-2140.

GERBER, S., COMELLAS-BIGLER, M., GOETZ, B. A. & LOCHER, K. P. 2008. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science,* 321**,** 246-250.

GHISAIDOOBE, A. B. & CHUNG, S. J. 2014. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences,* 15**,** 22518-22538.

GIETZ, R. D. & WOODS, R. A. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in enzymology.* Elsevier.

GOLDBURG, W. I. 1999. Dynamic light scattering. *American Journal of Physics,* 67**,** 1152-1160.

GOLDIE, K. N., ABEYRATHNE, P., KEBBEL, F., CHAMI, M., RINGLER, P. & STAHLBERG, H. 2014. Cryo-electron microscopy of membrane proteins. *Electron Microscopy: Methods and Protocols***,** 325-341.

GORBULEV, S., ABELE, R. & TAMPÉ, R. 2001. Allosteric crosstalk between peptide-binding, transport, and ATP hydrolysis of the ABC transporter TAP. *Proceedings of the National Academy of Sciences,* 98**,** 3732-3737.

GOSE, T., SHAFI, T., FUKUDA, Y., DAS, S., WANG, Y., ALLCOCK, A., GAVAN MCHARG, A., LYNCH, J., CHEN, T. & TAMAI, I. 2020. ABCG2 requires a single aromatic amino acid to "clamp" substrates and inhibitors into the binding pocket. *The FASEB Journal,* 34**,** 4890-4903.

GOTTESMAN, M. M., HRYCYNA, C., SCHOENLEIN, P. V., GERMANN, U. & PASTAN, I. 1995. Genetic analysis of the multidrug transporter. *Annual review of genetics,* 29**,** 607-649.

GRONINGER, E., BOER, M.-D., DE GRAAF, S., KAMPS, W. & DE BONT, E. 2002. Vincristine induced apoptosis in acute lymphoblastic leukaemia cells: a mitochondrial controlled pathway regulated by reactive oxygen species? *International journal of oncology,* 21**,** 1339-1345.

GUAN, L., SMIRNOVA, I. N., VERNER, G., NAGAMORI, S. & KABACK, H. R. 2006. Manipulating phospholipids for crystallization of a membrane transport protein. *Proceedings of the National Academy of Sciences,* 103**,** 1723-1726.

GUARNIERI, M. T., BLAGG, B. S. & ZHAO, R. 2011. A high-throughput TNP-ATP displacement assay for screening inhibitors of ATP-binding in bacterial histidine kinases. *Assay and drug development technologies,* 9**,** 174-183.

GUINIER, A. & FOURNET, G. 1955. Small angle X-ray scattering. *New York.*

GUNDRY, R. L., WHITE, M. Y., MURRAY, C. I., KANE, L. A., FU, Q., STANLEY, B. A. & VAN EYK, J. E. 2010. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Current protocols in molecular biology,* 90**,** 10.25. 1-10.25. 23.

HAFFKE, M., MENZEL, A., CARIUS, Y., JAHN, D. & HEINZ, D. W. 2010. Structures of the nucleotide-binding domain of the human ABCB6 transporter and its complexes with nucleotides. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 979-987.

HANIKENNE, M., KRÄMER, U., DEMOULIN, V. & BAURAIN, D. 2005. A comparative inventory of metal transporters in the green alga Chlamydomonas reinhardtii and the red alga Cyanidioschizon merolae. *Plant Physiology,* 137**,** 428-446.

HANNEMANN, F., BERA, A. K., FISCHER, B., LISUREK, M., TEUCHNER, K. & BERNHARDT, R. 2002. Unfolding and conformational studies on bovine adrenodoxin probed by engineered intrinsic tryptophan fluorescence. *Biochemistry,* 41**,** 11008-11016.

HEFTBERGER, P., KOLLMITZER, B., HEBERLE, F. A., PAN, J., RAPPOLT, M., AMENITSCH, H., KUČERKA, N., KATSARAS, J. & PABST, G. 2014. Global small-angle X-ray scattering data analysis for multilamellar vesicles: the evolution of the scattering density profile model. *Journal of applied crystallography,* 47**,** 173-180.

HEGEDŰS, T., ŐRFI, L., SEPRŐDI, A., VÁRADI, A., SARKADI, B. & KÉRI, G. 2002. Interaction of tyrosine kinase inhibitors with the human multidrug transporter proteins, MDR1 and MRP1. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1587**,** 318-325.

HEGYI, Z. & HOMOLYA, L. 2016. Functional cooperativity between ABCG4 and ABCG1 isoforms. *PLoS One,* 11**,** e0156516.

HELIAS, V., SAISON, C., BALLIF, B. A., PEYRARD, T., TAKAHASHI, J., TAKAHASHI, H., TANAKA, M., DEYBACH, J.-C., PUY, H. & LE GALL, M. 2012. ABCB6 is dispensable for erythropoiesis and specifies the new blood group system Langereis. *Nature genetics,* 44**,** 170-173.

HESPER, B. & HOGEWEG, P. 1970. Bioinformatica: een werkconcept. *Kameleon,* 1**,** 28-29.

HIGGINS, C. F. 1992. ABC transporters: from microorganisms to man. *Annual review of cell biology,* 8**,** 67-113.

HIGGINS, C. F. & LINTON, K. J. 2004. The ATP switch model for ABC transporters. *Nature structural & molecular biology,* 11**,** 918-926.

HIRATSUKA, T. 1975. 2'(or 3')-O-(2, 4, 6-trinitrophenyl) adenosine 5'-triphosphate as a probe for the binding site of heavy meromyosin ATPase. *The Journal of Biochemistry,* 78**,** 1135-1147.

HIRATSUKA, T. & UCHIDA, K. 1973. Preparation and properties of 2′(or 3′)-O-(2, 4, 6-trinitrophenyl) adenosine 5′-triphosphate, an analog of adenosine triphosphate. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 320**,** 635-647.

HIRAYAMA, H., KIMURA, Y., KIOKA, N., MATSUO, M. & UEDA, K. 2013. ATPase activity of human ABCG1 is stimulated by cholesterol and sphingomyelin [S]. *Journal of lipid research,* 54**,** 496-502.

HOFACKER, M., GOMPF, S., ZUTZ, A., PRESENTI, C., HAASE, W., VAN DER DOES, C., MODEL, K. & TAMPÉ, R. 2007. Structural and functional fingerprint of the mitochondrial ATP-binding cassette transporter Mdl1 from Saccharomyces cerevisiae. *Journal of Biological Chemistry,* 282**,** 3951-3961.

HOFFMAN, H. 1978. The dynamics of micelle formation. *Berichte der Bunsengesellschaft für physikalische Chemie,* 82**,** 988-1001.

HOGEWEG, P. 2011. The roots of bioinformatics in theoretical biology. *PLoS computational biology,* 7**,** e1002021.

HONG, H., JOH, N. H., BOWIE, J. U. & TAMM, L. K. 2009. Methods for measuring the thermodynamic stability of membrane proteins. *Methods in enzymology,* 455**,** 213-236.

HSIEH, V., KIM, M.-J., GELISSEN, I. C., BROWN, A. J., SANDOVAL, C., HALLAB, J. C., KOCKX, M., TRAINI, M., JESSUP, W. & KRITHARIDES, L. 2014. Cellular cholesterol regulates ubiquitination and degradation of the cholesterol export proteins ABCA1 and ABCG1. *Journal of Biological Chemistry,* 289**,** 7524-7536.

HUANG, P., STROFFEKOVA, K., CUPPOLETTI, J., MAHANTY, S. K. & SCARBOROUGH, G. A. 1996. Functional expression of the cystic fibrosis transmembrane conductance regulator in yeast. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1281**,** 80-90.

HUANG, Y., ANDERLE, P., BUSSEY, K. J., BARBACIORU, C., SHANKAVARAM, U., DAI, Z., REINHOLD, W. C., PAPP, A., WEINSTEIN, J. N. & SADÉE, W. 2004. Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer research,* 64**,** 4294-4301.

ILARI, A. & SAVINO, C. 2008. Protein structure determination by x-ray crystallography. *Bioinformatics***,** 63-87.

JACKSON, S. M., MANOLARIDIS, I., KOWAL, J., ZECHNER, M., TAYLOR, N. M., BAUSE, M., BAUER, S., BARTHOLOMAEUS, R., BERNHARDT, G. & KOENIG, B. 2018a. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nature structural & molecular biology,* 25**,** 333-340.

JACKSON, S. M., MANOLARIDIS, I., KOWAL, J., ZECHNER, M., TAYLOR, N. M. I., BAUSE, M., BAUER, S., BARTHOLOMAEUS, R., BERNHARDT, G., KOENIG, B., BUSCHAUER, A., STAHLBERG, H., ALTMANN, K. H. & LOCHER, K. P. 2018b. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol,* 25**,** 333-340.

JAFARI, R., ALMQVIST, H., AXELSSON, H., IGNATUSHCHENKO, M., LUNDBÄCK, T., NORDLUND, P. & MOLINA, D. M. 2014. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nature protocols,* 9**,** 2100-2122.

JÄGER, V., BÜSSOW, K., WAGNER, A., WEBER, S., HUST, M., FRENZEL, A. & SCHIRRMANN, T. 2013. High level transient production of recombinant antibodies and antibody fusion proteins in HEK293 cells. *BMC biotechnology,* 13**,** 1-20.

JANUCHOWSKI, R., ZAWIERUCHA, P., ANDRZEJEWSKA, M., RUCIŃSKI, M. & ZABEL, M. 2013. Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines. *Biomedicine & pharmacotherapy,* 67**,** 240-245.

JOHNSON, M., ZARETSKAYA, I., RAYTSELIS, Y., MEREZHUK, Y., MCGINNIS, S. & MADDEN, T. L. 2008. NCBI BLAST: a better web interface. *Nucleic acids research,* 36**,** W5-W9.

JONES, P. & GEORGE, A. 2004. The ABC transporter structure and mechanism: perspectives on recent research. *Cellular and Molecular Life Sciences CMLS,* 61**,** 682-699.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., TUNYASUVUNAKOOL, K., RONNEBERGER, O., BATES, R., ŽÍDEK, A. & BRIDGLAND, A. 2020. AlphaFold 2. *In Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book.*

KADABA, N. S., KAISER, J. T., JOHNSON, E., LEE, A. & REES, D. C. 2008. The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation. *Science,* 321**,** 250-253.

KAMPJUT, D., STEINER, J. & SAZANOV, L. A. 2021. Cryo-EM grid optimization for membrane proteins. *Iscience,* 24**,** 102139.

KAWANOBE, T., KOGURE, S., NAKAMURA, S., SATO, M., KATAYAMA, K., MITSUHASHI, J., NOGUCHI, K. & SUGIMOTO, Y. 2012. Expression of human ABCB5 confers resistance to taxanes and anthracyclines. *Biochemical and biophysical research communications,* 418**,** 736-741.

KEAN, J., CLEVERLEY, R. M., O'RYAN, L., FORD, R. C., PRINCE, S. M. & DERRICK, J. P. 2008. Characterization of a CorA Mg2+ transport channel from Methanococcus jannaschii using a Thermofluor-based stability assay. *Molecular membrane biology,* 25**,** 653-661.

KELLER, B. O., SUI, J., YOUNG, A. B. & WHITTAL, R. M. 2008. Interferences and contaminants encountered in modern mass spectrometry. *Analytica chimica acta,* 627**,** 71-81.

KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R., WYCKOFF, H. & PHILLIPS, D. C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* 181**,** 662-666.

KENIYA, M. V., HOLMES, A. R., NIIMI, M., LAMPING, E., GILLET, J.-P., GOTTESMAN, M. M. & CANNON, R. D. 2014. Drug resistance is conferred on the model yeast Saccharomyces cerevisiae by expression of full-length melanoma-associated human ATP-binding cassette transporter ABCB5. *Molecular pharmaceutics,* 11**,** 3452-3462.

KERR, I. D., HAIDER, A. J. & GELISSEN, I. C. 2011. The ABCG family of membrane-associated transporters: you don't have to be big to be mighty. *British journal of pharmacology,* 164**,** 1767-1779.

KIKHNEY, A. G. & SVERGUN, D. I. 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters,* 589**,** 2570-2577.

KIM, S., LEE, S. S., PARK, J. G., KIM, J. W., JU, S., CHOI, S. H., KIM, S., KIM, N. J., HONG, S. & KANG, J. Y. 2022. Structural Insights into Porphyrin Recognition by the Human ATP-Binding Cassette Transporter ABCB6. *Molecules and cells,* 45**,** 575-587.

KIM, S., SHARMA, A. K. & VATAMANIUK, O. K. 2018. N-terminal extension and C-terminal domains are required for ABCB6/HMT-1 protein interactions, function in

cadmium detoxification, and localization to the endosomal-recycling system in Caenorhabditis elegans. *Frontiers in physiology,* 9**,** 885.

KIM, Y. & CHEN, J. 2018. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. *Science,* 359**,** 915-919.

KISS, K., BROZIK, A., KUCSMA, N., TOTH, A., GERA, M., BERRY, L., VALLENTIN, A., VIAL, H., VIDAL, M. & SZAKACS, G. 2012. Shifting the paradigm: the putative mitochondrial protein ABCB6 resides in the lysosomes of cells and in the plasma membrane of erythrocytes. *PloS one,* 7**,** e37378.

KISS, K., KUCSMA, N., BROZIK, A., TUSNADY, G. E., BERGAM, P., VAN NIEL, G. & SZAKACS, G. 2015. Role of the N-terminal transmembrane domain in the endo-lysosomal targeting and function of the human ABCB6 protein. *Biochemical Journal,* 467**,** 127-139.

KLAPPE, K., HUMMEL, I., HOEKSTRA, D. & KOK, J. W. 2009. Lipid dependence of ABC transporter localization and function. *Chemistry and physics of lipids,* 161**,** 57-64.

KOHLSTAEDT, M., VON DER HOCHT, I., HILBERS, F., THIELMANN, Y. & MICHEL, H. 2015. Development of a Thermofluor assay for stability determination of membrane proteins using the Na+/H+ antiporter NhaA and cytochrome c oxidase. *Acta Crystallographica Section D: Biological Crystallography,* 71**,** 1112-1122.

KONAREV, P. V., VOLKOV, V. V., SOKOLOVA, A. V., KOCH, M. H. & SVERGUN, D. I. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography,* 36**,** 1277-1282.

KOTOV, V., BARTELS, K., VEITH, K., JOSTS, I., SUBHRAMANYAM, U. K. T., GÜNTHER, C., LABAHN, J., MARLOVITS, T. C., MORAES, I. & TIDOW, H. 2019. High-throughput stability screening for detergent-solubilized membrane proteins. *Scientific reports,* 9**,** 1-19.

KOUTSIOUBAS, A. 2017. Low-resolution structure of detergent-solubilized membrane proteins from small-angle scattering data. *Biophysical journal,* 113**,** 2373-2382.

KOZAK, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell,* 44**,** 283-292.

KRIEGER, E., NABUURS, S. B. & VRIEND, G. 2003. Homology modeling. *Methods of biochemical analysis,* 44**,** 509-524.

KRISHNAMURTHY, P. C., DU, G., FUKUDA, Y., SUN, D., SAMPATH, J., MERCER, K. E., WANG, J., SOSA-PINEDA, B., MURTI, K. G. & SCHUETZ, J. D. 2006. Identification of a mammalian mitochondrial porphyrin transporter. *Nature,* 443**,** 586-589.

KUBICEK, J., BLOCK, H., MAERTENS, B., SPRIESTERSBACH, A. & LABAHN, J. 2014. Expression and purification of membrane proteins. *Methods in enzymology.* Elsevier.

KUHNKE, G., NEUMANN, K., MÜHLENHOFF, U. & LILL, R. 2006. Stimulation of the ATPase activity of the yeast mitochondrial ABC transporter Atm1p by thiol compounds. *Molecular membrane biology,* 23**,** 173-184.

KUNTAL, B. K., APAROY, P. & REDDANNA, P. 2010. EasyModeller: A graphical interface to MODELLER. *BMC research notes,* 3**,** 226.

KURGAN, L. & MIZIANTY, M. J. 2009. Sequence-based protein crystallization propensity prediction for structural genomics: review and comparative analysis. *Natural Science,* 1**,** 93.

KURGAN, L., RAZIB, A. A., AGHAKHANI, S., DICK, S., MIZIANTY, M. & JAHANDIDEH, S. 2009. CRYSTALP2: sequence-based protein crystallization propensity prediction. *BMC structural biology,* 9**,** 1-14.

LABROU, N. E. 2014. Protein purification: an overview. *Protein Downstream Processing*, 3-10.

LACAPERE, J.-J., PEBAY-PEYROULA, E., NEUMANN, J.-M. & ETCHEBEST, C. 2007. Determining membrane protein structures: still a challenge! *Trends in biochemical sciences,* 32, 259-270.

LACONTE, L. E., SRIVASTAVA, S. & MUKHERJEE, K. 2017. Probing protein kinase-ATP interactions using a fluorescent ATP analog. *Proteomics for Drug Discovery.* Springer.

LAPPANO, R. & MAGGIOLINI, M. 2011. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature reviews Drug discovery,* 10, 47-60.

LE MAIRE, M., CHAMPEIL, P. & MÖLLER, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1508, 86-111.

LEE, J.-Y., KINCH, L. N., BOREK, D. M., WANG, J., WANG, J., URBATSCH, I. L., XIE, X.-S., GRISHIN, N. V., COHEN, J. C. & OTWINOWSKI, Z. 2016. Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature,* 533, 561-564.

LEE, S. C., BENNETT, B. C., HONG, W.-X., FU, Y., BAKER, K. A., MARCOUX, J., ROBINSON, C. V., WARD, A. B., HALPERT, J. R. & STEVENS, R. C. 2013. Steroid-based facial amphiphiles for stabilization and crystallization of membrane proteins. *Proceedings of the National Academy of Sciences,* 110, E1203-E1211.

LEHNE, G., GRASMO-WENDLER, U.-H., BERNER, J.-M., MEZA-ZEPEDA, L. A., ADAMSEN, B. L., FLACK, A., REINER, A., CLAUSEN, O. P. F., HOVIG, E. & MYKLEBOST, O. 2009. Upregulation of stem cell genes in multidrug resistant K562 leukemia cells. *Leukemia research,* 33, 1379-1385.

LERNER-MARMAROSH, N., GIMI, K., URBATSCH, I. L., GROS, P. & SENIOR, A. E. 1999. Large scale purification of detergent-soluble p-glycoprotein frompichia pastoris cells and characterization of nucleotide binding properties of wild-type, walker a, and walker b mutant proteins. *Journal of Biological Chemistry,* 274, 34711-34718.

LI, C., RAMJEESINGH, M., WANG, W., GARAMI, E., HEWRYK, M., LEE, D., ROMMENS, J. M., GALLEY, K. & BEAR, C. E. 1996. ATPase activity of the cystic fibrosis transmembrane conductance regulator. *Journal of Biological Chemistry,* 271, 28463-28468.

LI, N., WU, J.-X., DING, D., CHENG, J., GAO, N. & CHEN, L. 2017. Structure of a pancreatic ATP-sensitive potassium channel. *Cell,* 168, 101-110. e10.

LIAO, Y., YUAN, Q., TORRES, J., TAM, J. & LIU, D. 2006. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology,* 349, 264-275.

LICHTY, J. J., MALECKI, J. L., AGNEW, H. D., MICHELSON-HOROWITZ, D. J. & TAN, S. 2005. Comparison of affinity tags for protein purification. *Protein expression and purification,* 41, 98-105.

LIN, S.-H. & GUIDOTTI, G. 2009. Purification of membrane proteins. *Methods in enzymology.* Elsevier.

LINGAM, S., THONGHIN, N. & FORD, R. C. 2017. Investigation of the effects of the CFTR potentiator ivacaftor on human P-glycoprotein (ABCB1). *Scientific reports,* 7, 1-7.

LINKE, D. 2009. Detergents: an overview. *Methods in enzymology,* 463, 603-617.

LIPFERT, J. & DONIACH, S. 2007. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.,* 36, 307-327.

LIU, Z.-Q., MAHMOOD, T. & YANG, P.-C. 2014. Western blot: technique, theory and trouble shooting. *North American journal of medical sciences,* 6**,** 160.

LOCHER, K. P. 2004. Structure and mechanism of ABC transporters. *Current opinion in structural biology,* 14**,** 426-431.

LOCHER, K. P. 2009. Structure and mechanism of ATP-binding cassette transporters. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 239-245.

LOCHER, K. P. 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature structural & molecular biology,* 23**,** 487-493.

LOO, T. W., BARTLETT, M. C., DETTY, M. R. & CLARKE, D. M. 2012. The ATPase activity of the P-glycoprotein drug pump is highly activated when the N-terminal and central regions of the nucleotide-binding domains are linked closely together. *Journal of Biological Chemistry,* 287**,** 26806-26816.

LOUGHRAN, S. T. & WALLS, D. 2011. Purification of poly-histidine-tagged proteins. *Protein Chromatography.* Springer.

LYNCH, J., FUKUDA, Y., KRISHNAMURTHY, P., DU, G. & SCHUETZ, J. D. 2009. Cell survival under stress is enhanced by a mitochondrial ATP-binding cassette transporter that regulates hemoproteins. *Cancer research,* 69**,** 5560-5567.

MA, P., WEICHERT, D., ALEKSANDROV, L. A., JENSEN, T. J., RIORDAN, J. R., LIU, X., KOBILKA, B. K. & CAFFREY, M. 2017. The cubicon method for concentrating membrane proteins in the cubic mesophase. *nature protocols,* 12**,** 1745-1762.

MAHMOOD, T. & YANG, P.-C. 2012. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences,* 4**,** 429.

MALAWSKI, G. A., HILLIG, R. C., MONTECLARO, F., EBERSPAECHER, U., SCHMITZ, A. A., CRUSIUS, K., HUBER, M., EGNER, U., DONNER, P. & MÜLLER-TIEMANN, B. 2006. Identifying protein construct variants with increased crystallization propensity—A case study. *Protein science,* 15**,** 2718-2728.

MANALASTAS-CANTOS, K., KONAREV, P. V., HAJIZADEH, N. R., KIKHNEY, A. G., PETOUKHOV, M. V., MOLODENSKIY, D. S., PANJKOVICH, A., MERTENS, H. D., GRUZINOV, A. & BORGES, C. 2021. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography,* 54**,** 343-355.

MAO, Q. & SCARBOROUGH, G. A. 1997. Purification of functional human P-glycoprotein expressed in Saccharomyces cerevisiae. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1327**,** 107-118.

MARTIN, G. M., YOSHIOKA, C., REX, E. A., FAY, J. F., XIE, Q., WHORTON, M. R., CHEN, J. Z. & SHYNG, S.-L. 2017. Cryo-EM structure of the ATP-sensitive potassium channel illuminates mechanisms of assembly and gating. *elife,* 6**,** e24149.

MARTINEZ MOLINA, D. & NORDLUND, P. 2016. The cellular thermal shift assay: a novel biophysical assay for in situ drug target engagement and mechanistic biomarker studies. *Annual review of pharmacology and toxicology,* 56**,** 141-161.

MASTROTTO, F., FRANCINI, N., ALLEN, S., VAN DER WALLE, C., STOLNIK, S. & MANTOVANI, G. 2018. Synthetic glycopolymers as modulators of protein aggregation: influences of chemical composition, topology and concentration. *Journal of Materials Chemistry B,* 6**,** 1044-1054.

MATSSON, P., PEDERSEN, J. M., NORINDER, U., BERGSTRÖM, C. A. & ARTURSSON, P. 2009. Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP and MRP2 among registered drugs. *Pharmaceutical research,* 26**,** 1816-1831.

MENG, X. 2017. *Thermal stability and structural studies of the human Cystic Fibrosis Transmembrane Conductance Regulator.* University of Manchester.

MENG, X., CLEWS, J., KARGAS, V., WANG, X. & FORD, R. C. 2017. The cystic fibrosis transmembrane conductance regulator (CFTR) and its stability. *Cellular and Molecular Life Sciences,* 74**,** 23-38.

MINJIE, G. & ZHONGPING, S. 2013. Process control and optimization for heterologous protein production by methylotrophic Pichia pastoris. *Chinese Journal of Chemical Engineering,* 21**,** 216-226.

MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology,* 260**,** 289-298.

MISTELI, T. & SPECTOR, D. L. 1997. Applications of the green fluorescent protein in cell biology and biotechnology. *Nature biotechnology,* 15**,** 961-964.

MIZIANTY, M. J. & KURGAN, L. 2011. Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics,* 27**,** i24-i33.

MOITRA, K., SCALLY, M., MCGEE, K., LANCASTER, G., GOLD, B. & DEAN, M. 2011. Molecular evolutionary analysis of ABCB5: the ancestral gene is a full transporter with potentially deleterious single nucleotide polymorphisms. *PLoS One,* 6**,** e16318.

MOLINA, D. M., JAFARI, R., IGNATUSHCHENKO, M., SEKI, T., LARSSON, E. A., DAN, C., SREEKUMAR, L., CAO, Y. & NORDLUND, P. 2013. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science,* 341**,** 84-87.

MOLODENSKIY, D. S., MERTENS, H. D. & SVERGUN, D. I. 2020. An automated data processing and analysis pipeline for transmembrane proteins in detergent solutions. *Scientific reports,* 10**,** 1-11.

MONTECINOS-FRANJOLA, F., BAUER, B. L., MEARS, J. A. & RAMACHANDRAN, R. 2020. GFP fluorescence tagging alters dynamin-related protein 1 oligomerization dynamics and creates disassembly-refractory puncta to mediate mitochondrial fission. *Scientific reports,* 10**,** 1-16.

MORAES, I., EVANS, G., SANCHEZ-WEATHERBY, J., NEWSTEAD, S. & STEWART, P. D. S. 2014. Membrane protein structure determination—the next generation. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1838**,** 78-87.

MUNTEANU, B., BRAUN, M. & BOONROD, K. 2012. Improvement of PCR reaction conditions for site-directed mutagenesis of big plasmids. *Journal of Zhejiang University Science B,* 13**,** 244-247.

MURATA, K. & WOLF, M. 2018. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1862**,** 324-334.

MURPHY, R. M. 1997. Static and dynamic light scattering of biological macromolecules: what can we learn? *Current Opinion in Biotechnology,* 8**,** 25-30.

NANDIGAMA, K., LUSVARGHI, S., SHUKLA, S. & AMBUDKAR, S. V. 2019. Large-scale purification of functional human P-glycoprotein (ABCB1). *Protein expression and purification,* 159**,** 60-68.

NEWSTEAD, S., KIM, H., VON HEIJNE, G., IWATA, S. & DREW, D. 2007. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences,* 104**,** 13936-13941.

O'RYAN, L., RIMINGTON, T., CANT, N. & FORD, R. C. 2012a. Expression and purification of the cystic fibrosis transmembrane conductance regulator protein in Saccharomyces cerevisiae. *JoVE (Journal of Visualized Experiments)*, e3860.

O'RYAN, L., RIMINGTON, T. L., CANT, N. & FORD, R. C. 2012b. Expression and Purification of the Cystic Fibrosis Transmembrane Conductance Regulator Protein in Saccharomyces cerevisiae. *Journal of Visualized Experiments : JoVE*.

OBERG, K., CHRUNYK, B. A., WETZEL, R. & FINK, A. L. 1994. Native-like secondary structure in interleukin-1. beta. Inclusion bodies by attenuated total reflectance FTIR. *Biochemistry,* 33, 2628-2634.

OHI, M., LI, Y., CHENG, Y. & WALZ, T. 2004. Negative staining and image classification—powerful tools in modern electron microscopy. *Biological procedures online,* 6, 23-34.

OLDFIELD, S., LOWRY, C. A., RUDDICK, J. & LIGHTMAN, S. L. 2002. ABCG4: a novel human white family ABC-transporter expressed in the brain and eye. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1591, 175-179.

OLDHAM, M. L. & CHEN, J. 2011. Snapshots of the maltose transporter during ATP hydrolysis. *Proceedings of the National Academy of Sciences,* 108, 15152-15156.

OLDHAM, M. L., DAVIDSON, A. L. & CHEN, J. 2008. Structural insights into ABC transporter mechanism. *Current opinion in structural biology,* 18, 726-733.

OLDHAM, M. L., HITE, R. K., STEFFEN, A. M., DAMKO, E., LI, Z., WALZ, T. & CHEN, J. 2016. A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. *Nature,* 529, 537-540.

OLIVER, R. M. 1973. [26] Negative stain electron microscopy of protein macromolecules. *Methods in enzymology.* Elsevier.

ORTIZ, D., KREPPEL, L., SPEISER, D., SCHEEL, G., MCDONALD, G. & OW, D. 1992. Heavy metal tolerance in the fission yeast requires an ATP-binding cassette-type vacuolar membrane transporter. *The EMBO journal,* 11, 3491-3499.

OVERTON, I. M. & BARTON, G. J. 2006. A normalised scale for structural genomics target ranking: the OB-Score. *FEBS letters,* 580, 4005-4009.

PAPISH, A. L., TARI, L. W. & VOGEL, H. J. 2002. Dynamic light scattering study of calmodulin–target peptide complexes. *Biophysical journal,* 83, 1455-1464.

PARK, S., SHIMIZU, C., SHIMOYAMA, T., TAKEDA, M., ANDO, M., KOHNO, T., KATSUMATA, N., KANG, Y.-K., NISHIO, K. & FUJIWARA, Y. 2006. Gene expression profiling of ATP-binding cassette (ABC) transporters as a predictor of the pathologic response to neoadjuvant chemotherapy in breast cancer patients. *Breast cancer research and treatment,* 99, 9-17.

PARKER, J. L. & NEWSTEAD, S. 2012. Current trends in α-helical membrane protein crystallization: an update. *Protein Science,* 21, 1358-1365.

PATERSON, J. K., SHUKLA, S., BLACK, C. M., TACHIWADA, T., GARFIELD, S., WINCOVITCH, S., ERNST, D. N., AGADIR, A., LI, X. & AMBUDKAR, S. V. 2007. Human ABCB6 localizes to both the outer mitochondrial membrane and the plasma membrane. *Biochemistry,* 46, 9443-9452.

PATTERSON, G. H., KNOBEL, S. M., SHARIF, W. D., KAIN, S. R. & PISTON, D. W. 1997. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical journal,* 73, 2782-2790.

PEITSCH, M. C. 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Chemical Design Automation News,* 1, 13-14.

PELIKAN, M., HURA, G. L. & HAMMEL, M. 2009. Structure and flexibility within proteins as identified through small angle X-ray scattering. *General physiology and biophysics,* 28, 174.

PÉREZ, J. & KOUTSIOUBAS, A. 2015. Memprot: a program to model the detergent corona around a membrane protein based on SEC–SAXS data. *Acta Crystallographica Section D: Biological Crystallography,* 71**,** 86-93.

PETOUKHOV, M. V. & SVERGUN, D. I. 2013. Applications of small-angle X-ray scattering to biomacromolecular solutions. *The international journal of biochemistry & cell biology,* 45**,** 429-437.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry,* 25**,** 1605-1612.

PINKETT, H. W., LEE, A. T., LUM, P., LOCHER, K. P. & REES, D. C. 2007. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science,* 315**,** 373-377.

PLATT, G., POSTIS, V., HAO, Z., PALMER, T. & BALDWIN, S. Measuring the conformational stability of membrane proteins using the UNit.

POLIREDDY, K., KHAN, M. M. T., CHAVAN, H., YOUNG, S., MA, X., WALLER, A., GARCIA, M., PEREZ, D., CHAVEZ, S. & STROUSE, J. J. 2012. A novel flow cytometric HTS assay reveals functional modulators of ATP binding cassette transporter ABCB6. *PloS one,* 7**,** e40005.

POLLOCK, N., CANT, N., RIMINGTON, T. & FORD, R. C. 2014a. Purification of the cystic fibrosis transmembrane conductance regulator protein expressed in Saccharomyces cerevisiae. *JoVE (Journal of Visualized Experiments)***,** e51447.

POLLOCK, N. L., MCDEVITT, C. A., COLLINS, R., NIESTEN, P. H., PRINCE, S., KERR, I. D., FORD, R. C. & CALLAGHAN, R. 2014b. Improving the stability and function of purified ABCB1 and ABCA4: the influence of membrane lipids. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1838**,** 134-147.

PRIVÉ, G. G. 2007. Detergents for the stabilization and crystallization of membrane proteins. *Methods,* 41**,** 388-397.

PULLIKUTH, A. K. & GILL, S. S. 2002. In vivo membrane trafficking role for an insect N-ethylmaleimide-sensitive factor which is developmentally regulated in endocrine cells. *Journal of experimental biology,* 205**,** 911-926.

QU, P., TANG, H., GAO, Y., ZHANG, L. & WANG, S. 2010. Polyethersulfone composite membrane blended with cellulose fibrils. *BioResources,* 5**,** 2323-2336.

RAAIJMAKERS, J. G. 1987. Statistical analysis of the Michaelis-Menten equation. *Biometrics***,** 793-803.

RABBANI, G., KAUR, J., AHMAD, E., KHAN, R. H. & JAIN, S. 2014. Structural characteristics of thermostable immunogenic outer membrane protein from Salmonella enterica serovar Typhi. *Applied microbiology and biotechnology,* 98**,** 2533-2543.

RAHIMZADEH, M., SADEGHIZADEH, M., NAJAFI, F., ARAB, S. & MOBASHERI, H. 2016. Impact of heat shock step on bacterial transformation efficiency. *Molecular biology research communications,* 5**,** 257.

RANALLETTA, M., WANG, N., HAN, S., YVAN-CHARVET, L., WELCH, C. & TALL, A. R. 2006. Decreased atherosclerosis in low-density lipoprotein receptor knockout mice transplanted with Abcg1−/− bone marrow. *Arteriosclerosis, thrombosis, and vascular biology,* 26**,** 2308-2315.

RASBAND, W. S. 2011. Imagej, us national institutes of health, bethesda, maryland, usa. *http://imagej. nih. gov/ij/.*

REBBEOR, J. F. & SENIOR, A. E. 1998. Effects of cardiovascular drugs on ATPase activity of P-glycoprotein in plasma membranes and in purified reconstituted form. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1369**,** 85-93.

REES, D. C., JOHNSON, E. & LEWINSON, O. 2009. ABC transporters: the power to change. *Nature reviews Molecular cell biology,* 10**,** 218-227.

REINHARD, F., EBERHARD, D., WERNER, T., FRANKEN, H., CHILDS, D., DOCE, C., SAVITSKI, M. F., HUBER, W., BANTSCHEFF, M. & SAVITSKI, M. M. 2015. Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nature methods,* 12**,** 1129-1131.

RIGAUD, J.-L. & LÉVY, D. 2003. Reconstitution of membrane proteins into liposomes. *Methods in enzymology.* Elsevier.

RIGAUD, J.-L., PITARD, B. & LEVY, D. 1995. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics,* 1231**,** 223-246.

RIMINGTON1, 2, T. L., PURNA, B. M., POLLOCK, N., KAPPES, J. C., FORD, R. C. & URBATSCH, I. 2018. A gene optimisation approach to enhance expression of human CFTR in Saccharomyces cerevisiae. *Tracy L Rimington***,** 87.

RIMINGTON, T. L., CANT, N., O'RYAN, L. & FORD, R. C. 2018. Optimisation of Saccharomyces cerevisiae culture conditions for expression of recombinant Cystic Fibrosis Transmembrane conductance Regulator (CFTR). *Tracy L Rimington***,** 33.

RITCHIE, T., GRINKOVA, Y., BAYBURT, T., DENISOV, I., ZOLNERCIKS, J., ATKINS, W. & SLIGAR, S. 2009. Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods in enzymology,* 464**,** 211-231.

ROSSMAN, M., ARNOLD, E., RUECKERT, R., SHERRY, B., VRIEND, G. & ERICKSON, J. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature (London),* 317**,** 145-153.

ROST, B. 1997. Protein structures sustain evolutionary drift. *Folding and Design,* 2**,** S19-S24.

RUBINSTEIN, J. L. 2007. Structural analysis of membrane protein complexes by single particle electron microscopy. *Methods,* 41**,** 409-416.

SÁFÁR, Z., VASKÓ, B., K RITCHIE, T., IMRE, G., MOGYORÓSI, K., RAJNAI, Z., FEKETE, Z., SZERÉMY, P., MUKA, L. & K ZOLNERCIKS, J. 2016. Investigating ABCB1-mediated drug-drug interactions: considerations for in vitro and in vivo assay design. *Current Drug Metabolism,* 17**,** 430-455.

SAIDIJAM, M., PSAKIS, G., CLOUGH, J. L., MEULLER, J., SUZUKI, S. I., HOYLE, C. J., PALMER, S. L., MORRISON, S. M., POS, M. K. & ESSENBERG, R. C. 2003. Collection and characterisation of bacterial membrane proteins. *FEBS letters,* 555**,** 170-175.

ŠALI, A. & OVERINGTON, J. P. 1994. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Science,* 3**,** 1582-1596.

SAMBROOK, J. & RUSSELL, D. W. 2006. SDS-polyacrylamide gel electrophoresis of proteins. *CSH Protoc,* 2006**,** pdb. prot4540.

SAMESHIMA, T., YUKAWA, T., HIROZANE, Y., YOSHIKAWA, M., KATOH, T., HARA, H., YOGO, T., MIYAHISA, I., OKUDA, T. & MIYAMOTO, M. 2019. Small-scale panel comprising diverse gene family targets to evaluate compound promiscuity. *Chemical Research in Toxicology,* 33**,** 154-161.

SANO, O., TSUJITA, M., SHIMIZU, Y., KATO, R., KOBAYASHI, A., KIOKA, N., REMALEY, A. T., MICHIKAWA, M., UEDA, K. & MATSUO, M. 2016. ABCG1

and ABCG4 suppress γ-secretase activity and amyloid β production. *PLoS One,* 11**,** e0155400.

SAUNA, Z. E. & AMBUDKAR, S. V. 2001. Characterization of the catalytic cycle of ATP hydrolysis by human P-glycoprotein: the two ATP hydrolysis events in a single catalytic cycle are kinetically similar but affect different functional outcomes. *Journal of Biological Chemistry,* 276**,** 11653-11661.

SAURIN, W., HOFNUNG, M. & DASSA, E. 1999. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *Journal of molecular evolution,* 48**,** 22-41.

SCHATTON, T. & FRANK, M. H. 2008. Cancer stem cells and human malignant melanoma. *Pigment cell & melanoma research,* 21**,** 39-55.

SCHATTON, T., MURPHY, G. F., FRANK, N. Y., YAMAURA, K., WAAGA-GASSER, A. M., GASSER, M., ZHAN, Q., JORDAN, S., DUNCAN, L. M. & WEISHAUPT, C. 2008. Identification of cells initiating human melanomas. *Nature,* 451**,** 345-349.

SCHMIDT-KREY, I. & RUBINSTEIN, J. L. 2011. Electron cryomicroscopy of membrane proteins: specimen preparation for two-dimensional crystals and single particles. *Micron,* 42**,** 107-116.

SCHMIDT, F. 2004. Recombinant expression systems in the pharmaceutical industry. *Applied microbiology and biotechnology,* 65**,** 363-372.

SCHMIDT, T. G., KOEPKE, J., FRANK, R. & SKERRA, A. 1996. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *Journal of molecular biology,* 255**,** 753-766.

SCHNEIDMAN-DUHOVNY, D., HAMMEL, M. & SALI, A. 2010. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic acids research,* 38**,** W540-W544.

SCHRODT, S., KOCH, J. & TAMPÉ, R. 2006. Membrane topology of the transporter associated with antigen processing (TAP1) within an assembled functional peptide-loading complex. *Journal of Biological Chemistry,* 281**,** 6455-6462.

SCHWARTZ, M. S., BENCI, J. L., SELOTE, D. S., SHARMA, A. K., CHEN, A. G., DANG, H., FARES, H. & VATAMANIUK, O. K. 2010. Detoxification of multiple heavy metals by a half-molecule ABC transporter, HMT-1, and coelomocytes of Caenorhabditis elegans. *PloS one,* 5**,** e9564.

SEDDON, A. M., CURNOW, P. & BOOTH, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1666**,** 105-117.

SENIOR, A. E., AL-SHAWI, M. K. & URBATSCH, I. L. 1995. ATP hydrolysis by multidrug-resistance protein from Chinese hamster ovary cells. *Journal of bioenergetics and biomembranes,* 27**,** 31-36.

SHAFI, T. 2022. *Characterisation of drug translocation pathway in multidrug resistance proteins: P-glycoprotein and ABCG2 using structural approaches.* PhD, University of Manchester.

SHANER, N. C., PATTERSON, G. H. & DAVIDSON, M. W. 2007. Advances in fluorescent protein technology. *Journal of cell science,* 120**,** 4247-4260.

SHANER, N. C., STEINBACH, P. A. & TSIEN, R. Y. 2005. A guide to choosing fluorescent proteins. *Nature methods,* 2**,** 905-909.

SHANMUGARAJAH, K., LINKA, N., GRÄFE, K., SMITS, S. H., WEBER, A. P., ZEIER, J. & SCHMITT, L. 2019. ABCG1 contributes to suberin formation in Arabidopsis thaliana roots. *Scientific reports,* 9**,** 1-12.

SHAPIRO, A. L. & MAIZEL JR, J. V. 1969. Molecular weight estimation of polypeptides by SDS-polyacrylamide gel electrophoresis: further data concerning resolving power and general considerations. *Analytical biochemistry,* 29**,** 505-514.

SHAROM, F. J., RUSSELL, P. L., QU, Q. & LU, P. 2003. Fluorescence techniques for studying membrane transport proteins. *Membrane Transporters.* Springer.

SHINTRE, C. A., PIKE, A. C., LI, Q., KIM, J.-I., BARR, A. J., GOUBIN, S., SHRESTHA, L., YANG, J., BERRIDGE, G. & ROSS, J. 2013. Structures of ABCB10, a human ATP-binding cassette transporter in apo-and nucleotide-bound states. *Proceedings of the National Academy of Sciences,* 110**,** 9710-9715.

SIMPSON, B. W., OWENS, T. W., ORABELLA, M. J., DAVIS, R. M., MAY, J. M., TRAUGER, S. A., KAHNE, D. & RUIZ, N. 2016. Identification of residues in the lipopolysaccharide ABC transporter that coordinate ATPase activity with extractor function. *MBio,* 7**,** e01729-16.

SJOLLEMA, K. A., SCHNELL, U., KUIPERS, J., KALICHARAN, R. & GIEPMANS, B. N. 2012. Correlated light microscopy and electron microscopy. *Methods in cell biology,* 111**,** 157-173.

SKARDA, L., KOWAL, J. & LOCHER, K. P. 2021. Structure of the human cholesterol transporter ABCG1. *Journal of Molecular Biology,* 433**,** 167218.

SLABINSKI, L., JAROSZEWSKI, L., RYCHLEWSKI, L., WILSON, I. A., LESLEY, S. A. & GODZIK, A. 2007. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics,* 23**,** 3403-3405.

SMITH, B. 1984. SDS polyacrylamide gel electrophoresis of proteins. *Proteins.* Springer.

SMITH, P. C., KARPOWICH, N., MILLEN, L., MOODY, J. E., ROSEN, J., THOMAS, P. J. & HUNT, J. F. 2002. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Molecular cell,* 10**,** 139-149.

SMITH, S. M. 2011. Strategies for the purification of membrane proteins. *Methods Mol Biol,* 681**,** 485-96.

SMYTH, M. & MARTIN, J. 2000. x Ray crystallography. *Molecular Pathology,* 53**,** 8.

SONG, G., ZHANG, S., TIAN, M., ZHANG, L., GUO, R., ZHUO, W. & YANG, M. 2021. Molecular insights into the human ABCB6 transporter. *Cell Discovery,* 7**,** 1-11.

SONODA, Y., NEWSTEAD, S., HU, N.-J., ALGUEL, Y., NJI, E., BEIS, K., YASHIRO, S., LEE, C., LEUNG, J. & CAMERON, A. D. 2011. Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure,* 19**,** 17-25.

SOOKSA-NGUAN, T., YAKUBOV, B., KOZLOVSKYY, V. I., BARKUME, C. M., HOWE, K. J., THANNHAUSER, T. W., RUTZKE, M. A., HART, J. J., KOCHIAN, L. V. & REA, P. A. 2009. Drosophila ABC transporter, DmHMT-1, confers tolerance to cadmium: DmHMT-1 and its yeast homolog, SpHMT-1, are not essential for vacuolar phytochelatin sequestration. *Journal of biological chemistry,* 284**,** 354-362.

SORENSEN, C. & SHI, D. 2000. Guinier analysis for homogeneous dielectric spheres of arbitrary size. *Optics communications,* 178**,** 31-36.

STRUCTURAL GENOMICS, C. 2008. China Structural Genomics C. Northeast Structural Genomics C. Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, et al. *Nat Methods,* 5**,** 135-146.

STUREK, J. M., CASTLE, J. D., TRACE, A. P., PAGE, L. C., CASTLE, A. M., EVANS-MOLINA, C., PARKS, J. S., MIRMIRA, R. G. & HEDRICK, C. C. 2010. An intracellular role for ABCG1-mediated cholesterol transport in the regulated secretory pathway of mouse pancreatic β cells. *The journal of clinical investigation,* 120**,** 2575-2589.

SUN, Y., WANG, J., LONG, T., QI, X., DONNELLY, L., ELGHOBASHI-MEINHARDT, N., ESPARZA, L., COHEN, J. C., XIE, X.-S. & HOBBS, H. H. 2021. Molecular basis of cholesterol efflux via ABCG subfamily transporters. *Proceedings of the National Academy of Sciences,* 118**,** e2110483118.

SVERGUN, D. I., KOCH, M. H., TIMMINS, P. A. & MAY, R. P. 2013. *Small angle X-ray and neutron scattering from solutions of biological macromolecules*, Oxford University Press.

SWIFT, M. L. 1997. GraphPad prism, data analysis, and scientific graphing. *Journal of chemical information and computer sciences,* 37**,** 411-412.

SZOKA JR, F. & PAPAHADJOPOULOS, D. 1980. Comparative properties and methods of preparation of lipid vesicles (liposomes). *Annual review of biophysics and bioengineering,* 9**,** 467-508.

TALL, A. R. & WANG, N. 2022. New insights into cholesterol efflux via ABCA1. *Nature Cardiovascular Research,* 1**,** 198-199.

TANGELLA, L. P., AROOJ, M., DEPLAZES, E., GRAY, E. S. & MANCERA, R. L. 2021. Identification and characterisation of putative drug binding sites in human ATP-binding cassette B5 (ABCB5) transporter. *Computational and structural biotechnology journal,* 19**,** 691-704.

TATE, C. 2001. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS letters,* 504**,** 94-98.

TAYLOR, N. M., MANOLARIDIS, I., JACKSON, S. M., KOWAL, J., STAHLBERG, H. & LOCHER, K. P. 2017. Structure of the human multidrug transporter ABCG2. *Nature,* 546**,** 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**,** 211-222.

TERSTAPPEN, G. C. & REGGIANI, A. 2001a. In silico research in drug discovery. *Trends Pharmacol Sci,* 22**,** 23-6.

TERSTAPPEN, G. C. & REGGIANI, A. 2001b. In silico research in drug discovery. *Trends in pharmacological sciences,* 22**,** 23-26.

THEODOULOU, F. L. & KERR, I. D. 2015. ABC transporter research: going strong 40 years on. *Biochemical Society Transactions,* 43**,** 1033-1040.

THÉVENOT, C., GRASSL, B., BASTIAT, G. & BINANA, W. 2005. Aggregation number and critical micellar concentration of surfactant determined by time-dependent static light scattering (TDSLS) and conductivity. *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* 252**,** 105-111.

THONGHIN, N. 2019. *Structural studies of the multi-drug resistance protein P-glycoprotein (ABCB1)*, The University of Manchester (United Kingdom).

THONGHIN, N., COLLINS, R. F., BARBIERI, A., SHAFI, T., SIEBERT, A. & FORD, R. C. 2018a. 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**,** 1-11.

THONGHIN, N., COLLINS, R. F., BARBIERI, A., SHAFI, T., SIEBERT, A. & FORD, R. C. 2018b. Novel features in the structure of P-glycoprotein (ABCB1) in the post-hydrolytic state as determined at 7.9 Å resolution. *BMC structural biology* [Online], 18. Available: http://europepmc.org/abstract/MED/30545335

https://doi.org/10.1186/s12900-018-0098-z

https://europepmc.org/articles/PMC6293506

https://europepmc.org/articles/PMC6293506?pdf=render [Accessed 2018/12//].

THONGHIN, N., KARGAS, V., CLEWS, J. & FORD, R. C. 2018c. Cryo-electron microscopy of membrane proteins. *Methods,* 147**,** 176-186.

TSAI, A. M., UDOVIC, T. J. & NEUMANN, D. A. 2001. The inverse relationship between protein dynamics and thermal stability. *Biophysical Journal,* 81**,** 2339-2343.

TUSNÁDY, G. E., SARKADI, B., SIMON, I. & VÁRADI, A. 2006. Membrane topology of human ABC proteins. *FEBS letters,* 580**,** 1017-1022.

UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPF, C., SJÖSTEDT, E. & ASPLUND, A. 2015. Proteomics. Tissue-based map of the human proteome. *Science (New York, NY),* 347**,** 1260419-1260419.

ULRICH, D. L., LYNCH, J., WANG, Y., FUKUDA, Y., NACHAGARI, D., DU, G., SUN, D., FAN, Y., TSURKAN, L. & POTTER, P. M. 2012. ATP-dependent mitochondrial porphyrin importer ABCB6 protects against phenylhydrazine toxicity. *Journal of Biological Chemistry,* 287**,** 12679-12690.

UZMAN, A. 2001. Molecular Cell Biology (4th edition): Harvey Lodish, Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira, David Baltimore and James Darnell; Freeman & Co., New York, NY, 2000, 1084 pp., list price $102.25, ISBN 0-7167-3136-3. *Biochemistry and Molecular Biology Education,* 29**,** 126-128.

VARATHARAJAN, S., ABRAHAM, A., KARATHEDATH, S., GANESAN, S., LAKSHMI, K. M., ARTHUR, N., SRIVASTAVA, V. M., GEORGE, B., SRIVASTAVA, A. & MATHEWS, V. 2017. ATP-binding casette transporter expression in acute myeloid leukemia: association with in vitro cytotoxicity and prognostic markers. *Pharmacogenomics,* 18**,** 235-244.

VARGA, J. K. & TUSNÁDY, G. E. 2018. TMCrys: predict propensity of success for transmembrane protein crystallization. *Bioinformatics,* 34**,** 3126-3130.

VASILIOU, V., VASILIOU, K. & NEBERT, D. W. 2009. Human ATP-binding cassette (ABC) transporter family. *Human genomics,* 3**,** 1-10.

VATAMANIUK, O. K., BUCHER, E. A., SUNDARAM, M. V. & REA, P. A. 2005. CeHMT-1, a putative phytochelatin transporter, is required for cadmium tolerance in Caenorhabditis elegans. *Journal of Biological Chemistry,* 280**,** 23684-23690.

VAUGHAN, A. M. & ORAM, J. F. 2005. ABCG1 redistributes cell cholesterol to domains removable by high density lipoprotein but not by lipid-depleted apolipoproteins. *Journal of Biological Chemistry,* 280**,** 30150-30157.

VIEIRA GOMES, A. M., SOUZA CARMO, T., SILVA CARVALHO, L., MENDONÇA BAHIA, F. & PARACHIN, N. S. 2018. Comparison of yeasts as hosts for recombinant protein production. *Microorganisms,* 6**,** 38.

VOGELSTEIN, B. & GILLESPIE, D. 1979. Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences,* 76**,** 615-619.

VON STETTEN, D., NOIRCLERC-SAVOYE, M., GOEDHART, J., GADELLA, T. W. & ROYANT, A. 2012. Structure of a fluorescent protein from Aequorea victoria bearing the obligate-monomer mutation A206K. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications,* 68**,** 878-882.

WACKER, I., KAETHER, C., KROMER, A., MIGALA, A., ALMERS, W. & GERDES, H.-H. 1997. Microtubule-dependent transport of secretory vesicles visualized in real time with a GFP-tagged secretory protein. *Journal of cell science,* 110**,** 1453-1463.

WAGNER, S., BADER, M. L., DREW, D. & DE GIER, J.-W. 2006. Rationalizing membrane protein overexpression. *Trends in biotechnology,* 24**,** 364-371.

WALIAN, P., CROSS, T. A. & JAP, B. K. 2004. Structural genomics of membrane proteins. *Genome biology,* 5**,** 1-8.

WALLIN, E. & HEIJNE, G. V. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science,* 7**,** 1029-1038.

WANG, C., CAO, C., WANG, N., WANG, X., WANG, X. & ZHANG, X. C. 2020a. Cryo-electron microscopy structure of human ABCB6 transporter. *Protein Sci,* 29**,** 2363-2374.

WANG, C., CAO, C., WANG, N., WANG, X., WANG, X. & ZHANG, X. C. 2020b. Cryo-electron microscopy structure of human ABCB6 transporter. *Protein Science,* 29**,** 2363-2374.

WANG, C., CASTRO, A. F., WILKES, D. M. & ALTENBERG, G. A. 1999. Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein. *Biochemical Journal,* 338**,** 77-81.

WANG, H., FENG, L., WEBB, G. I., KURGAN, L., SONG, J. & LIN, D. 2018. Critical evaluation of bioinformatics tools for the prediction of protein crystallization propensity. *Briefings in bioinformatics,* 19**,** 838-852.

WANG, H., FENG, L., ZHANG, Z., WEBB, G. I., LIN, D. & SONG, J. 2016. Crysalis: an integrated server for computational analysis and design of protein crystallization. *Scientific reports,* 6**,** 1-14.

WANG, H., WANG, M., TAN, H., LI, Y., ZHANG, Z. & SONG, J. 2014. PredPPCrys: accurate prediction of sequence cloning, protein production, purification and crystallization propensity from protein sequences using multi-step heterogeneous feature fusion and selection. *PloS one,* 9**,** e105902.

WANG, L., HE, F., BU, J., LIU, X., DU, W., DONG, J., COONEY, J. D., DUBEY, S. K., SHI, Y. & GONG, B. 2012. ABCB6 mutations cause ocular coloboma. *The American Journal of Human Genetics,* 90**,** 40-48.

WANG, L. & SIGWORTH, F. J. 2006. Cryo-EM and single particles. *Physiology,* 21**,** 13-18.

WANG, R., QIN, Y. & LI, X. 2022. Structural basis of acyl-CoA transport across the peroxisomal membrane by human ABCD1. *Cell Research,* 32**,** 214-217.

WARD, A., REYES, C. L., YU, J., ROTH, C. B. & CHANG, G. 2007. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences,* 104**,** 19005-19010.

WARD, A. B., SZEWCZYK, P., GRIMARD, V., LEE, C.-W., MARTINEZ, L., DOSHI, R., CAYA, A., VILLALUZ, M., PARDON, E. & CREGGER, C. 2013. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *Proceedings of the National Academy of Sciences,* 110**,** 13386-13391.

WEBB, B. & SALI, A. 2016. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics,* 54**,** 5.6. 1-5.6. 37.

WEN, P.-C. & TAJKHORSHID, E. 2008. Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis. *Biophysical journal,* 95**,** 5100-5110.

WILSON, B. J., SAAB, K. R., MA, J., SCHATTON, T., PÜTZ, P., ZHAN, Q., MURPHY, G. F., GASSER, M., WAAGA-GASSER, A. M. & FRANK, N. Y. 2014. ABCB5 Maintains Melanoma-Initiating Cells through a Proinflammatory Cytokine Signaling CircuitABCB5 Maintains Melanoma–Initiating Cells. *Cancer research,* 74**,** 4196-4207.

WOODBURY, D. J., WHITT, E. C. & COFFMAN, R. E. 2021. A review of TNP-ATP in protein binding studies: benefits and pitfalls. *Biophysical Reports,* 1**,** 100012.

WOOLFORD, C., DANIELS, L., PARK, F., JONES, E., VAN ARSDELL, J. & INNIS, M. 1986. The PEP4 gene encodes an aspartyl protease implicated in the posttranslational regulation of Saccharomyces cerevisiae vacuolar hydrolases. *Molecular and Cellular Biology,* 6**,** 2500-2510.

WU, M. & LANDER, G. C. 2020. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Current opinion in structural biology,* 64**,** 9-16.

XU, D., LI, Y., YANG, F., SUN, C.-R., PAN, J., WANG, L., CHEN, Z.-P., FANG, S.-C., YAO, X. & HOU, W.-T. 2022. Structure and transport mechanism of the human cholesterol transporter ABCG1. *Cell Reports,* 38**,** 110298.

YAMASHITA, M., SHEPHERD, M., BOOTH, W. I., XIE, H., POSTIS, V., NYATHI, Y., TZOKOV, S. B., POOLE, R. K., BALDWIN, S. A. & BULLOUGH, P. A. 2014. Structure and function of the bacterial heterodimeric ABC transporter CydDC: Stimulation of ATPase activity by thiol and heme compounds. *Journal of Biological Chemistry,* 289**,** 23177-23188.

YANG, A., ALROSAN, A. Z., SHARPE, L. J., BROWN, A. J., CALLAGHAN, R. & GELISSEN, I. C. 2021. Regulation of ABCG4 transporter expression by sterols and LXR ligands. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1865**,** 129769.

YANG, Z., ZHOU, Q., MOK, L., SINGH, A., SWARTZ, D. J., URBATSCH, I. L. & BROUILLETTE, C. G. 2017. Interactions and cooperativity between P-glycoprotein structural domains determined by thermal unfolding provides insights into its solution structure and function. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1859**,** 48-60.

YASUI, K., MIHARA, S., ZHAO, C., OKAMOTO, H., SAITO-OHARA, F., TOMIDA, A., FUNATO, T., YOKOMIZO, A., NAITO, S. & IMOTO, I. 2004. Alteration in copy numbers of genes as a mechanism for acquired drug resistance. *Cancer research,* 64**,** 1403-1410.

YEH, A. P., MCMILLAN, A. & STOWELL, M. H. 2006. Rapid and simple protein-stability screens: application to membrane proteins. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 451-457.

YOO, J.-S., MOYER, B. D., BANNYKH, S., YOO, H.-M., RIORDAN, J. R. & BALCH, W. E. 2002. Non-conventional trafficking of the cystic fibrosis transmembrane conductance regulator through the early secretory pathway. *Journal of Biological Chemistry,* 277**,** 11401-11409.

YOSHIKAWA, M., YABUUCHI, H., KUROIWA, A., IKEGAMI, Y., SAI, Y., TAMAI, I., TSUJI, A., MATSUDA, Y., YOSHIDA, H. & ISHIKAWA, T. 2002. Molecular and cytogenetic characterization of the mouse ATP-binding cassette transporter Abcg4. *Gene,* 293**,** 67-75.

YVAN-CHARVET, L., RANALLETTA, M., WANG, N., HAN, S., TERASAKA, N., LI, R., WELCH, C. & TALL, A. R. 2007. Combined deficiency of ABCA1 and ABCG1 promotes foam cell accumulation and accelerates atherosclerosis in mice. *The Journal of clinical investigation,* 117**,** 3900-3908.

ZHANG, C., LI, D., ZHANG, J., CHEN, X., HUANG, M., ARCHACKI, S., TIAN, Y., REN, W., MEI, A. & ZHANG, Q. 2013. Mutations in ABCB6 cause dyschromatosis universalis hereditaria. *Journal of Investigative Dermatology,* 133**,** 2221-2228.

ZHANG, L., ALEKSANDROV, L. A., RIORDAN, J. R. & FORD, R. C. 2011. Do main location within the cystic fibrosis transmembrane conductance regulator protein investigated by electron microscopy and gold labelling. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1808**,** 399-404.

ZHANG, L., ALEKSANDROV, L. A., ZHAO, Z., BIRTLEY, J. R., RIORDAN, J. R. & FORD, R. C. 2009a. Architecture of the cystic fibrosis transmembrane conductance regulator protein and structural changes associated with phosphorylation and nucleotide binding. *Journal of structural biology,* 167**,** 242-251.

ZHANG, L., CHEN, G., TANG, H., CHENG, Q. & WANG, S. 2009b. Preparation and characterization of composite membranes of polysulfone and microcrystalline cellulose. *Journal of applied polymer science,* 112**,** 550-556.

ZHANG, Y., NASSER, V., PISANTY, O., OMARY, M., WULFF, N., DI DONATO, M., TAL, I., HAUSER, F., HAO, P. & ROTH, O. 2018. A transportome-scale amiRNA-based screen identifies redundant roles of Arabidopsis ABCB6 and ABCB20 in auxin transport. *Nature communications,* 9**,** 1-12.

ZHAO, F.-Q. & CRAIG, R. 2003. Capturing time-resolved changes in molecular structure by negative staining. *Journal of structural biology,* 141**,** 43-52.

ZHAO, X., LI, G. & LIANG, S. 2013. Several affinity tags commonly used in chromatographic purification. *Journal of analytical methods in chemistry,* 2013.

ZHOU, H.-X. & CROSS, T. A. 2013. Influences of membrane mimetic environments on membrane protein structures. *Annual review of biophysics,* 42**,** 361.

ZOU, P., BORTOLUS, M. & MCHAOURAB, H. S. 2009. Conformational cycle of the ABC transporter MsbA in liposomes: detailed analysis using double electron–electron resonance spectroscopy. *Journal of molecular biology,* 393**,** 586-597.

ZOU, P. & MCHAOURAB, H. S. 2009. Alternating access of the putative substrate-binding chamber in the ABC transporter MsbA. *Journal of molecular biology,* 393**,** 574-585.

Appendix- Supplementary data

1)

> **ABCB5 sequence** (Uniprot ID: B5X0E34)

Optimization of **ABCB5** for *S. cerevisiae*.

BamHI-Kozak(yeast optimized)-Start codon-ABCB5-XmaI

ggatccAAAACAATG
ANSERTNGLQETNQRYGPLQEQVPKVGNQAVGPIEIFRFADNLDIVLMTLGILASM
INGATVPLMSLVLGEISDHLINGCLVQTNRTKYQNCSQTQEKLNEDIIVLTLYYIGI
GAAALIFGYVQISFWVITAARQTTRIRKQFFHSILAQDISWFDGSDICELNTRMTGDI
NKLCDGIGDKIPLMFQNISGFSIGLVISLIKSWKLSLVVLSTSPLIMASSALCSRMIIS
LTSKELDAYSKAGAVAEEALSSIQTVTAFGAQEKEIQRYTQHLKDAKDAGIKRAT
ASKLSLGAVYFFMNGAYGLAFWYGTSLIFGGEPGYTIGTILAVFFSVIHSSYCIGSV
APHLETFTVARGAAFNIFQVIDKKPNIDNFSTAGFVPECIEGNIEFKNVSFSYPSRPS
AKVLKGLNLKIKAGETVALVGPSGSGKSTTVQLLQRLYDPEDGCITVDENDIRAQ
NVRHYREQIGVVRQEPVLFGTTIGNNIKFGREGVGEKEMEQAAREANAYDFIMAF
PKKFNTLVGEKGAQMSGGQKQRIAIARALVRNPKILILDEATSALDTESESLVQTA
LEKASKGRTTIVVAHRLSTIRGADLIVTMKDGMVVEKGTHAELMAKQGLYYSLA
MAQDIKKVDEQMESRTCSTAGNASYGSLCDVNSAKAPCTDQLEEAVHHQKTSLP
EVSLLKIFKLSKSEWPFVVLGTLASALNGSVHPVFSIIFGKLVTMFEDKNKATLKQ
DAELYSMMLVVLGIVALVTYLMQGLFYGRAEENLAMRLRHSAFKAMLYQDMA
WYDDKENNTGALTTTLAVDVAQIQGAATSRLGIVTQDVSNMSLSILISFIYGWEM
TLLILSFAPVLAVTGMIQTAAMAGFANRDKQALKRAGKIATEAVENIRTVVSLTRE
RAFEQMYEETLQTQHRNALKRAHITGCCYAVSHAFVHFAHAAGFRFGAYLIQAG
RMMPEGMFIVFTAIAYGAMAIGETLVWAPEYSKAKAGASHLFALLKNKPTINSCS
QSGEKPDTCEGNLEFREVSFVYPCRPEVPVLQNMSLSIEKGKTVAFVGSSGCGKST
CVQLLQRFYDPMKGQVLLDGVDVKELNVQWLRSQTAIVSQEPVLFNCSIAENIAY
GDNSRMVPLEEIKEVADAANIHSFIEGLPRKYNTLVGLRGVQLSGGQKQRLAIARA
LLRKPKILLLDEATSALDNESEKVVQQALDKARRGKTCLVVAHRLSTIQNADMIV
VLQNGSIKEQGTHQELLRNGDTYFKLVAAH
cccggg

2)

> **ABCB6 sequence** (Uniprot ID: A0A7N5P644)

Optimization of **ABCB6** for *S. cerevisiae*.

BamHI-Kozak(yeast optimized)-Start codon-ABCB6-XmaI

ggatccAAAACAATG
MSFVVGNYCEAEGPLGPAWVKGGLSPCFFFTLMPSILMALGALAVVLALPCKRRE
RPAGARELSWSAGPLVASYVLQLLLAILQVALPLATLVGRVGTAGGAPLPGYLLP
ASILGTLAGACGLGLLVVEHNQAWKKLAMGIWIEFRHSSGLLLLWTVAFTAENLA
LVSWNNPLWWWARADLGQQVQFSLWVLRYVVSGGLFILGLWAPGLRPQSYALK
VNEEDQDVERIEVQSTEAPPPRSTWRDLGRKLRLLSGYLWPRGSPVLQLVVLLCL

GLMGLERGLNVLVPIFYRDIVNLLTEKAPWSSLAWTVTTFVILKFLQGGGTGSTGF
VSNMRTFLWIRVQQFTSRQVELRLFSHLHELSLRWHLGRRTGEVLRIVDRGTSSVT
GLLSYLVFNVLPTLADITIGIIYFSMFFNAWFGLIVFLCMSLYLIVTIVVTEWRTKFR
RAMNTQENATRARAVDSLLNFETVKYYNAEGYEVDRYREAILKYQDLEWKSTAS
LVLLNQTQNLVIGLGLLAGSLLCAYFVSEQKLQVGDFVLFGTYIIQLYMPLNWFGT
YYRMIQTNFIDMENMFDLLKEKPEVKDLPGAGPLRFQRGQIEFENVHFSYTSGRET
LQDVSFTVMPGQTLALVGPSGAGKSTVLRLLFRFYDISSGCIRIDGQDISQVTQISL
RSHIGVVPQDTVLFNDTIANNIRYGCITAGDEEVMAAAQAAGIHEAILTFPEGYET
QVGERGLKLSGGEKQRVAIARTILKAPDIILLDEATSALDTSNERAIQASLAKVCAN
RTTIVVAHRLSTVVDADQILVFKDGCIVERGRHEALLSRGGMYADMWQLQQRGQ
EEVSEDAKPPD cccggg

3)

>**ABCG1** **(Uniprot ID: G1NSJ6)**

Optimization of **ABCG1** for *S. cerevisiae*.

BamHI-**Kozak**-**Start codon**-**8His**-**GFP**-**QF**-**TEV**-**ABCG1**-**Stop**-**XmaI**

ggatcc**AAAACA****ATG**CACCACCACCATCATCATCATCAT
TCTAAAGGTGAAGAATTATTCACTGGCGTTGTCCCAATTTTGGTTGAATTA
GATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGG
TGACGCTACTTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAA
ATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTCGGTTATGGTGTTCA
ATGTTTTGCGAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTC
TGCCATGCCAGAAGGTTATGTTCAAGAAGAACTATTTTTTTCAAAGATGA
CGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAG
TTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTT
TAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTACATCATGG
CTGACAAACAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACACAAC
ATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCA
ATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAA
TCTAAGCTTTCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTT
AGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTACAA
ACAATTTGAAAATTTATATTTTCAAGGT
LSGLFWIFLLGTIMNTSSYSAVMTELKSVCVSVDEVVSSNTELCEKDMLNGHLKK
VDNNLTEAQRLSSLPRRAAINIEFKDLSYSVPEGPWWKKKGYKTLLKGISGKFNSG
QLVAIMGPSGAGKSTLMNILAGYRETGMKGTILTNGKARDLRCFRKVSCYIMQDH
MLMPHLTVQEAMMVSAHLKLKEKDEGRKEMVKEILTALGLLSCATTRTGSLSGG
QRKRLAIGLELVNNPPVMFFDEPTSGLDSSSCFQVVSLMKGLAQGGRSIICTIHQPS
AKLFEMFDQLYVLSQGQCVYRGKVSNLVPYLRDLGLNCPTYHNPADFVMEVASG
EYGDQNGRLVRAVGEGLCDADCRKEPPGADGEVNPFLQHQPSEEVKQSKGWKG
RKKNSTSTESCHSVSASCLTQFFILFKRALLSTVRDAVLMHLRATSHIAIGLLIGLLY
LGIGNEAKKVMSNSGFLFFSMLFLMFAALMPTVLTFPLEMGVFLREHLNYWYSLK
AYYMAKTLADVPFQIVFPVAYCSIVYWMTSQPSDAVRFVLFAALGTMTSLVAQSL
GLLIGAAATSLQVAIFVGPMTTIPILLFSGFFVSLGTIPTYLQWMSYISYVRYGFEGV
MLSIYGLDREDLHCDVDETCHFQKSEAILRELDVENAKLYLDFIVLGIFFILLRIIAY
FVLRYKIRAER-stop-cccggg

4)

>ABCG4 (Uniprot ID: Q91WA9)

Optimization of **ABCG4** for *S. cerevisiae*.

BamHI-Kozak-Start codon-8His-GFP-QF-TEV-ABCG1-Stop-XmaI

ggatccAAAACAATGCACCACCACCATCATCATCATCAT
TCTAAAGGTGAAGAATTATTCACTGGCGTTGTCCCAATTTTGGTTGAATTA
GATGGTGATGTTAATGGTCACAATTTTCTGTCTCCGGTGAAGGTGAAGG
TGACGCTACTTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAA
ATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTCGGTTATGGTGTTCA
ATGTTTTGCGAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTC
TGCCATGCCAGAAGGTTATGTTCAAGAAGAACTATTTTTTTCAAAGATGA
CGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAG
TTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTT
TAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTACATCATGG
CTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACACAAC
ATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAATACTCCA
ATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAA
TCTAAGCTTTCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTT
AGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTACAA
ACAATTTGAAAATTTATATTTTCAAGGT
AEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITEAQRFSHLPKR
SAVDIEFVELSYSVREGPCWRKRGYKTLLKCLSGKFCRRELIGIMGPSGAGKSTFM
NILAGYRESGMKGQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMMVSA
NLKLSEKQEVKKELVTEILTALGLMSCSHTRTALLSGGQRKRLAIALELVNNPPVM
FFDEPTSGLDSASCFQVVSLMKSLAHGGRTVICTIHQPSAKLFEMFDKLYILSQGQC
IFKGVVTNLIPYLKGLGLHCPTYHNPADFIIEVASGEYGDLNPMLFRAVQNGLCTM
AEKKSSPGKNELPAHCPTCPPELDPIESHTFATSTLTQFCILFRRTFLSILRDTVLTHL
RFMSHVLIGVLIGLLYLHIGDDASKVFNNTGFLFFSMLFLMFAALMPTVLTFPLEM
AVFMREHLNYWYTLKAYYLAKTMADVPFQVVCPVVYCSIVYWMTGQPAETSRF
LLFSALAIATALVAQSLGLLIGAASTSLQVATFVGPVTAIPVLLFSGFFVSFKTIPTY
LQWSSYLSYVRYGFEGLILTIYGMERGHLTCLDEQCPFRGPTIILRELDVEEAKLY
MDFLVLGIFFLALRLLAYLVLRYRVKSER-stop-cccggg

Appendix Table 3.1: List of orthologs of whole ABCA family selected for TMCrys screening. Protein sequences were obtained from the UniProt database https://www.uniprot.org/

ABCA Family					
Organisms	**Solubilisation**	**Purification**	**Crystallization**	**Whole Process**	**Reliability of prediction**
ABCA1 Subfamily					
Human-O95477	0.681	0.963	0.097	0.58	0.88
Mouse-P41233	0.636	0.965	0.098	0.567	0.883
Rat	0.688	0.955	0.139	0.594	0.877

Dog	0.706	0.962	0.128	0.599	0.876
Bovine	0.669	0.962	0.132	0.588	0.879
Horse	0.679	0.968	0.107	0.585	0.879
Chimpanzee	0.65	0.961	0.11	0.573	0.882
Chick	0.706	0.961	0.092	0.586	0.879
Duck bill platypus	0.685	0.963	0.089	0.579	0.88
Cat	0.681	0.958	0.111	0.583	0.879
Opossum	0.667	0.967	0.093	0.576	0.881
American Chameleon	0.63	0.97	0.104	0.568	0.883
Mallard	0.681	0.963	0.127	0.59	0.878
Rabbit	0.688	0.967	0.147	0.6	0.876
Mustelo furo	0.681	0.96	0.125	0.589	0.878
Ground Squirrel	0.688	0.969	0.106	0.588	0.879
Sheep	0.662	0.961	0.134	0.586	0.879
Tasmanian Devil	0.644	0.955	0.097	0.565	0.883
Otolemur Garnetti	0.657	0.965	0.119	0.58	0.88
ABCA2 Subfamily					
Human	0.625	0.956	0.174	0.585	0.879
Rat	0.659	0.969	0.128	0.585	0.879
Mouse (P41234)	0.685	0.971	0.085	0.58	0.88
Bovine	0.579	0.948	0.816	0.781	0.861
Horse	0.64	0.956	0.169	0.588	0.878
Mouse (A2AJ26)	0.679	0.961	0.134	0.591	0.878
Chicken	0.7	0.953	0.098	0.584	0.879
Chimpanzee	0.635	0.951	0.294	0.627	0.871
Pongo abelii	0.645	0.959	0.126	0.576	0.881
Tasmanian Devil	0.709	0.976	0.114	0.599	0.876
Ground Squirrel-I3MLT5	0.675	0.947	0.693	0.772	0.859
Small eared galogo-H0XTW4	0.662	0.95	0.706	0.773	0.86
Mallard (A0A493TVP8)	0.681	0.965	0.088	0.578	0.881
Zebra finch-H0YW03	0.696	0.94	0.672	0.769	0.857
Zebra fish (A2BI69)	0.724	0.951	0.57	0.748	0.855
Opossum	0.709	0.971	0.083	0.588	0.879

ABCA3 Subfamily				
Human	0.679	0.937	0.254	0.623	0.871
Mouse	0.71	0.971	0.14	0.607	0.875
Rat	0.717	0.964	0.232	0.637	0.868
Brown Bat	0.695	0.953	0.249	0.633	0.869
Zebra fish	0.688	0.961	0.181	0.61	0.874
Chimpanzee	0.653	0.955	0.32	0.643	0.867
Opossum	0.74	0.982	0.14	0.62	0.872
Ground Squirrel	0.706	0.957	0.241	0.634	0.869
Western Clawed Frog	0.717	0.919	0.129	0.588	0.878
Chicken	0.725	0.979	0.103	0.602	0.876
Horse	0.712	0.961	0.201	0.625	0.871
ABCA4 Subfamily				
Human	0.705	0.967	0.175	0.616	0.873
Mouse	0.733	0.961	0.15	0.615	0.873
Rat	0.705	0.955	0.093	0.584	0.879
Dog	0.705	0.948	0.168	0.607	0.875
Bovine	0.705	0.948	0.168	0.607	0.875
Horse	0.691	0.969	0.148	0.603	0.875
Western Clawed Frog	0.626	0.928	0.161	0.572	0.882
Cat	0.709	0.965	0.205	0.626	0.871
Opossum	0.705	0.962	0.189	0.618	0.872
Mouse Ear Cress	0.645	0.982	0.095	0.574	0.881
Giant Panda	0.705	0.957	0.173	0.612	0.874
Rabbit	0.691	0.942	0.167	0.6	0.876
Bat	0.691	0.967	0.221	0.626	0.871
Chimpanzee	0.705	0.955	0.184	0.615	0.873
ABCA5 Subfamily				
Human	0.73	0.96	0.1	0.596	0.877
Mouse	0.717	0.964	0.146	0.609	0.874
Rat	0.748	0.975	0.1	0.608	0.874
Dog	0.759	0.971	0.146	0.625	0.871
Chimpanzee	0.73	0.958	0.077	0.588	0.878
Bovine	0.762	0.956	0.081	0.6	0.876
Horse	0.708	0.929	0.046	0.561	0.884
Chicken	0.686	0.926	0.131	0.581	0.88
Opossum	0.705	0.949	0.067	0.574	0.881
Brown Bat	0.748	0.974	0.088	0.603	0.875
Gorilla	0.73	0.964	0.068	0.587	0.879
Zebra Fish	0.729	0.971	0.071	0.59	0.878

Mallard	0.737	0.966	0.119	0.607	0.875
Duck bill platypus (F7G9P8)	0.761	0.929	0.639	0.776	0.86
ABCA6 Subfamily					
Human	0.752	0.927	0.042	0.574	0.881
Mouse	0.71	0.912	0.077	0.566	0.883
Rat	0.756	0.948	0.196	0.633	0.869
Mouse Ear Cress	0.667	0.986	0.129	0.594	0.877
Chimpanzee	0.752	0.935	0.049	0.579	0.88
Bovine	0.727	0.971	0.068	0.589	0.878
Horse	0.743	0.916	0.029	0.563	0.884
Cat	0.704	0.953	0.14	0.599	0.876
Sheep	0.7	0.965	0.085	0.583	0.879
Brown Bat	0.718	0.945	0.071	0.578	0.881
Rabbit	0.767	0.974	0.082	0.607	0.875
Giant Panda	0.714	0.946	0.157	0.606	0.875
ABCA7 Subfamily					
Human	0.568	0.832	0.171	0.523	0.892
Mouse	0.63	0.765	0.085	0.494	0.898
Rat	0.595	0.891	0.089	0.525	0.892
Mouse Ear Cress	0.686	0.96	0.111	0.586	0.879
Bovine	0.55	0.755	0.116	0.474	0.902
Dog	0.599	0.886	0.12	0.535	0.889
Opossum	0.66	0.868	0.073	0.534	0.89
Brown Bat	0.613	0.872	0.192	0.559	0.885
ABCA8 Subfamily					
Human	0.761	0.949	0.038	0.583	0.88
Mouse(a)	0.728	0.909	0.123	0.587	0.879
Mouse (b)	0.74	0.963	0.147	0.617	0.873
Horse	0.762	0.934	0.052	0.583	0.88
Dog	0.738	0.932	0.047	0.572	0.882
Rat	0.706	0.905	0.133	0.581	0.88
Rat (a)	0.743	0.923	0.088	0.585	0.879
ABCA9 Subfamily					
Human	0.748	0.907	0.175	0.61	0.874
Mouse	0.781	0.944	0.087	0.604	0.875
Rat	0.734	0.941	0.082	0.585	0.879
Bovine	0.679	0.943	0.086	0.57	0.882
Chimpanzee	0.737	0.902	0.072	0.57	0.882
Horse	0.749	0.943	0.033	0.572	0.882

Chicken	0.703	0.953	0.109	0.589	0.878
Cat	0.684	0.928	0.055	0.556	0.885
ABCA10 Subfamily					
Human	0.742	0.894	0.126	0.588	0.879
Bovine-F1MRL3	0.587	0.811	0.111	0.503	0.896
Horse	0.729	0.933	0.085	0.582	0.88
Cat	0.712	0.967	0.075	0.585	0.879
Duck bill platypus-F7G9R6	0.739	0.852	0.6	0.73	0.851
Zebrafinch-H0Z0V3	0.728	0.897	0.717	0.781	0.861
Sheep	0.711	0.933	0.049	0.564	0.883
ABCA12 Subfamily					
Human	0.689	0.972	0.031	0.564	0.883
Mouse	0.674	0.985	0.041	0.567	0.883
Rat	0.674	0.973	0.036	0.561	0.884
Chimpanzee	0.689	0.973	0.031	0.565	0.883
Bovine	0.687	0.98	0.049	0.572	0.882
Dog	0.689	0.979	0.044	0.571	0.882
Horse	0.687	0.97	0.028	0.562	0.884
Duckbill platypus	0.682	0.97	0.044	0.566	0.883
Chicken	0.687	0.97	0.061	0.573	0.882
ABCA13 Subfamily					
Human	0.704	0.902	0.012	0.539	0.889
Mouse	0.704	0.831	0.012	0.516	0.893
Rat	0.678	0.794	0.018	0.497	0.897
Dog	0.697	0.901	0.048	0.549	0.887
Chimpanzee	0.682	0.91	0.012	0.535	0.89
Horse	0.704	0.897	0.017	0.539	0.889
Bovine	0.706	0.865	0.118	0.563	0.884
Pongo abelii	0.688	0.796	0.129	0.538	0.889

Appendix Table 3.2: List of orthologs of whole ABCB family selected for TMCrys screening. Protein sequences were obtained from the UniProt database (https://www.uniprot.org/)

ABCB Family					
Organisms	**Solubilisation**	**Purification**	**Crystallization**	**Whole Process**	**Reliability of prediction**
ABCB1 Subfamily					

Human	0.634	0.985	0.567	0.729	0.851
Rat 1a	0.644	0.989	0.555	0.729	0.851
Rat1b	0.653	0.984	0.137	0.591	0.878
Dog	0.634	0.986	0.53	0.717	0.852
Horse	0.679	0.987	0.605	0.757	0.856
Chimpanzee	0.629	0.985	0.573	0.729	0.851
Giant Panda	0.634	0.989	0.646	0.756	0.856
Bat	0.629	0.983	0.515	0.709	0.853
Ground Squirrel	0.67	0.989	0.683	0.781	0.861
Opossum	0.635	0.975	0.286	0.632	0.869
Mouse 1a	0.644	0.987	0.753	0.794	0.864
Mouse 1b	0.658	0.983	0.448	0.696	0.856
ABCB2/TAP1 Subfamily					
Human	0.621	0.87	0.074	0.522	0.892
Mouse	0.625	0.905	0.225	0.585	0.879
Rat	0.627	0.821	0.289	0.579	0.88
Zebrafish	0.682	0.946	0.075	0.568	0.883
Chimpanzee	0.602	0.872	0.08	0.518	0.893
Horse	0.656	0.881	0.157	0.564	0.883
Bovine	0.642	0.761	0.202	0.535	0.89
Dog-Q5W414	0.663	0.893	0.249	0.602	0.876
Rabbit-G1SGW1	0.642	0.853	0.21	0.568	0.883
Giant Panda	0.568	0.682	0.488	0.585	0.879
Zebrafish	0.682	0.946	0.075	0.568	0.883
Chicken-B5BSK4	0.713	0.683	0.546	0.647	0.866
ABCB3/TAP2 Subfamily					
Human	0.618	0.937	0.33	0.626	0.871
Mouse	0.633	0.876	0.228	0.579	0.88
Rat	0.688	0.902	0.191	0.594	0.877
Bovine	0.668	0.751	0.261	0.56	0.884
Dog	0.598	0.82	0.316	0.578	0.881
Chimpanzee	0.669	0.959	0.258	0.629	0.87
Horse	0.656	0.886	0.232	0.591	0.871
Chicken-A5HUM1	0.703	0.929	0.67	0.768	0.859
Zebrafish	0.643	0.984	0.183	0.603	0.875
Rabbit-G1TNZ8	0.661	0.886	0.271	0.606	0.875
ABCB4 Subfamily					
Human-P21439	0.696	0.991	0.567	0.751	0.855
Mouse	0.652	0.988	0.36	0.667	0.862
Rat-Q08201	0.654	0.984	0.603	0.747	0.854

Bovine	0.67	0.981	0.365	0.672	0.861
Dog	0.64	0.99	0.283	0.637	0.868
Rabbit	0.67	0.989	0.489	0.716	0.852
Sheep-W5PWJ6	0.605	0.942	0.913	0.83	0.869
Giant Panda	0.653	0.989	0.272	0.638	0.868
ABCB5 Subfamily					
Human	0.672	0.961	0.473	0.702	0.855
Mouse	0.698	0.973	0.591	0.754	0.856
Rat	0.723	0.972	0.406	0.7	0.855
Chimpanzee	0.672	0.966	0.398	0.679	0.86
Dog	0.627	0.979	0.32	0.642	0.867
Bovine	0.676	0.984	0.11	0.59	0.878
Horse	0.681	0.973	0.192	0.615	0.873
Zebrafish	0.627	0.982	0.301	0.637	0.868
Bat	0.675	0.975	0.265	0.638	0.868
Sheep	0.676	0.979	0.103	0.586	0.879
Cat	0.662	0.984	0.113	0.586	0.879
Rabbit	0.662	0.978	0.497	0.712	0.853
ABCB6 Subfamily					
Human	0.579	0.966	0.474	0.673	0.861
Mouse	0.59	0.958	0.25	0.599	0.876
Rat	0.697	0.952	0.229	0.626	0.871
Bovine	0.62	0.974	0.297	0.63	0.87
Dog	0.607	0.898	0.206	0.57	0.882
Horse	0.61	0.962	0.346	0.639	0.868
Cat	0.635	0.955	0.284	0.625	0.871
Chimpanzee	0.563	0.953	0.262	0.593	0.877
Opossum	0.591	0.95	0.09	0.543	0.888
Rabbit	0.617	0.97	0.151	0.579	0.88
Giant Panda	0.613	0.925	0.847	0.795	0.864
ABCB7 Subfamily					
Human	0.593	0.978	0.36	0.644	0.867
Mouse	0.593	0.971	0.202	0.589	0.878
Rat	0.593	0.979	0.268	0.613	0.873
Zebrafish	0.639	0.955	0.132	0.576	0.881
Dog	0.581	0.952	0.324	0.619	0.872
Bovine	0.593	0.975	0.346	0.638	0.868
Horse	0.593	0.955	0.483	0.677	0.86
ABCB8 Subfamily					
Human-Q9NUT2	0.682	0.926	0.136	0.582	0.88
Mouse	0.685	0.947	0.058	0.563	0.884
ABCB9 Subfamily					

Organism	Solubilisation	Purification	Crystallization	Whole Process	Reliability
Human	0.636	0.934	0.487	0.686	0.858
Rat	0.693	0.939	0.117	0.583	0.88
ABCB10 Subfamily					
Human	0.563	0.763	0.661	0.574	0.859
Mouse	0.711	0.886	0.126	0.574	0.881
ABCB11 Subfamily					
Human	0.67	0.975	0.186	0.611	0.874
Rat	0.63	0.986	0.4	0.667	0.862

Appendix Table 3.3: List of orthologs of whole ABCC family selected for TMCrys screening. Protein sequences were obtained from the UniProt database (https://www.uniprot.org/)

ABCC Family					
Organisms	Solubilisation	Purification	Crystallization	Whole Process	Reliability of prediction
ABCC1/MRP1 Subfamily					
Human	0.665	0.961	0.035	0.554	0.886
Mouse	0.707	0.944	0.033	0.561	0.884
Rat	0.643	0.963	0.045	0.551	0.886
Bovine-Q8HXQ5	0.645	0.977	0.046	0.556	0.885
Dog	0.672	0.958	0.06	0.563	0.884
Chimpanzee	0.67	0.967	0.253	0.63	0.87
Cat	0.669	0.962	0.043	0.558	0.885
Chicken	0.651	0.98	0.049	0.56	0.884
ABCC2 Subfamily					
Human	0.699	0.971	0.051	0.574	0.881
Mouse	0.653	0.968	0.084	0.568	0.883
Rat	0.637	0.971	0.055	0.554	0.886
Rabbit	0.623	0.964	0.042	0.543	0.888
Bovine	0.651	0.983	0.034	0.556	0.885
Chimpanzee	0.699	0.973	0.054	0.575	0.881
Chicken	0.66	0.974	0.086	0.574	0.881
Dog	0.701	0.97	0.033	0.568	0.883
Opossum	0.709	0.971	0.082	0.587	0.879
Horse	0.682	0.97	0.029	0.56	0.884
ABCC3/MRP2 Subfamily					
Human	0.604	0.904	0.042	0.516	0.893
Rat	0.686	0.96	0.079	0.575	0.881

Mouse	0.681	0.961	0.047	0.563	0.884
Bovine	0.898	0.965	0.041	0.568	0.883
Dog-F1PRY2	0.643	0.95	0.065	0.553	0.886
Chimpanzee	0.618	0.911	0.64	0.723	0.851
Horse-F7CU08	0.614	0.948	0.079	0.547	0.887
ABCC4 Subfamily					
Human	0.604	0.973	0.121	0.566	0.883
Mouse	0.598	0.983	0.115	0.565	0.883
Rat	0.624	0.98	0.084	0.563	0.884
Cat	0.622	0.971	0.147	0.58	0.88
Chicken	0.659	0.967	0.117	0.581	0.88
Opossum	0.62	0.976	0.205	0.6	0.876
Duck bill platypus	0.67	0.972	0.156	0.599	0.876
Rabbit	0.639	0.977	0.091	0.569	0.882
Giant Panda	0.577	0.959	0.612	0.716	0.852
ABCC5 Subfamily					
Human	0.721	0.975	0.028	0.575	0.881
Rat	0.732	0.971	0.036	0.58	0.88
Mouse	0.722	0.968	0.042	0.577	0.881
Bovine	0.717	0.985	0.107	0.603	0.875
Dog	0.721	0.964	0.038	0.574	0.881
Horse	0.714	0.98	0.044	0.58	0.88
Cat	0.721	0.972	0.068	0.587	0.879
Chicken	0.71	0.971	0.068	0.583	0.88
Chimpanzee	0.721	0.973	0.03	0.574	0.881
Rabbit	0.698	0.963	0.049	0.57	0.882
Giant Panda	0.714	0.974	0.046	0.578	0.881
Opossum	0.706	0.955	0.132	0.598	0.877
Sheep	0.721	0.96	0.031	0.571	0.882
Duck bill platypus	0.734	0.967	0.174	0.625	0.871
Little Brown Bat-G1PGU4	0.687	0.918	0.493	0.699	0.856
ABCC6 Subfamily					
Human	0.698	0.926	0.071	0.565	0.883
Rat	0.595	0.893	0.117	0.535	0.89
Mouse	0.57	0.901	0.126	0.532	0.89
Bovine	0.639	0.935	0.126	0.567	0.883
Dog	0.651	0.922	0.101	0.558	0.885
Horse-F7B795(With	0.637	0.897	0.147	0.56	0.884

M)					
Horse-F7B795(Del M)	0.637	0.816	0.737	0.73	0.851
Cat	0.615	0.91	0.169	0.565	0.883
Chimpanzee	0.63	0.895	0.109	0.544	0.888
Rabbit	0.647	0.912	0.135	0.565	0.883
Sheep	0.622	0.909	0.106	0.546	0.887
Mallard-U3INJ9 Deleted M	0.634	0.941	0.844	0.806	0.867
Mallard-U3INJ9 With M	0.65	0.935	0.303	0.629	0.87
ABCC7 Subfamily					
Human	0.688	0.898	0.052	0.546	0.887
Mouse	0.645	0.964	0.066	0.558	0.888
Rat	0.666	0.958	0.065	0.563	0.884
Horse	0.663	0.934	0.032	0.543	0.888
Chimpanzee	0.685	0.883	0.055	0.541	0.888
Bovine	0.631	0.875	0.084	0.53	0.89
Duck bill platypus	0.68	0.917	0.041	0.546	0.887
Sheep	0.631	0.883	0.051	0.521	0.892
Rabbit	0.673	0.91	0.056	0.546	0.887
Opossum	0.643	0.912	0.059	0.538	0.889
ABCC8 Subfamily					
Human	0.665	0.969	0.098	0.577	0.881
Rat	0.657	0.971	0.081	0.57	0.882
Mouse	0.657	0.955	0.088	0.567	0.883
Chimpanzee	0.686	0.967	0.091	0.581	0.88
Bovine	0.678	0.95	0.09	0.573	0.882
Dog	0.67	0.949	0.091	0.57	0.882
Horse-F6S5D0	0.676	0.959	0.087	0.574	0.881
Cat	0.674	0.966	0.054	0.565	0.883
Zebrafish	0.658	0.917	0.041	0.539	0.889
Duck bill platypus	0.615	0.94	0.073	0.543	0.888
Opossum	0.673	0.927	0.069	0.557	0.85
Chicken	0.675	0.9	0.04	0.538	0.889
Giant Panda-G1L8T3	0.601	0.902	0.689	0.731	0.851
Little Brown Bat	0.641	0.95	0.087	0.56	0.884
ABCC9 Subfamily					

Human	0.653	0.949	0.042	0.548	0.887
Mouse	0.682	0.952	0.071	0.569	0.883
Rat	0.644	0.963	0.065	0.558	0.885
Rabbit	0.652	0.937	0.038	0.542	0.888
Bovine	0.653	0.971	0.054	0.559	0.885
Dog	0.652	0.941	0.039	0.544	0.888
Chicken	0.609	0.931	0.056	0.532	0.89
Cat	0.653	0.947	0.04	0.547	0.887
Opossum	0.637	0.917	0.074	0.543	0.888
Horse	0.682	0.97	0.066	0.573	0.882
Chimpanzee	0.653	0.954	0.036	0.547	0.887
Zebrafish	0.589	0.976	0.06	0.542	0.888
ABCC10 Subfamily					
Human	0.612	0.816	0.114	0.514	0.894
Mouse	0.584	0.889	0.118	0.53	0.89
Rat	0.576	0.796	0.117	0.496	0.897
Chimpanzee	0.6	0.823	0.113	0.512	0.894
Dog	0.569	0.88	0.091	0.514	0.894
Bovine	0.592	0.829	0.136	0.519	0.893
Zebrafish	0.691	0.934	0.153	0.593	0.878
Horse	0.623	0.83	0.13	0.527	0.891
Opossum	0.659	0.883	0.082	0.541	0.888
Cat	0.662	0.821	0.062	0.515	0.894
Rabbit	0.58	0.852	0.092	0.508	0.895
Sheep	0.594	0.849	0.095	0.512	0.894
ABCC11 Subfamily					
Human	0.684	0.947	0.09	0.573	0.882
Bovine	0.672	0.949	0.074	0.565	0.883
Dog	0.688	0.882	0.113	0.561	0.884
Horse	0.694	0.87	0.083	0.549	0.887
Duck bill platypus	0.699	0.942	0.247	0.629	0.87
Chimpanzee	0.713	0.953	0.121	0.596	0.877
Cat	0.694	0.968	0.08	0.581	0.88
Sheep	0.704	0.923	0.184	0.604	0.875
Rabbit	0.662	0.945	0.126	0.578	0.881
ABCC12 Subfamily					
Human	0.674	0.976	0.104	0.585	0.879
Mouse	0.71	0.983	0.061	0.585	0.879
Rat	0.654	0.98	0.054	0.563	0.884

Zebrafish	0.684	0.98	0.088	0.584	0.879
Bovine	0.69	0.98	0.05	0.573	0.882
Dog	0.658	0.968	0.054	0.56	0.884
Horse	0.698	0.958	0.085	0.58	0.88
Chimpanzee	0.68	0.977	0.11	0.589	0.878
Duck bill platypus	0.732	0.976	0.067	0.592	0.878
Opossum	0.664	0.966	0.154	0.595	0.877
Sheep	0.661	0.98	0.064	0.568	0.883

Appendix Table 3.4: List of orthologs of whole ABCD family selected for TMCrys screening. Protein sequences were obtained from the UniProt database (https://www.uniprot.org/)

ABCD Family					
Organisms	**Solubilisation**	**Purification**	**Crystallization**	**Whole Process**	**Reliability of prediction**
ABCD1 Subfamily					
Human-P33897	0.663	0.919	0.16	0.581	0.88
Mouse-P48410	0.699	0.909	0.131	0.58	0.88
Rat-D3ZHR2	0.706	0.96	0.345	0.67	0.862
Zebrafish-F1RBC8	0.718	0.892	0.102	0.571	0.882
Bovine-Q2KJ57	0.71	0.941	0.186	0.612	0.874
Dog-E2R4U7	0.721	0.892	0.236	0.616	0.873
Horse-A0A5F5PK56	0.675	0.929	0.067	0.557	0.885
Chimpanzee-K6ZKQ4	0.641	0.901	0.176	0.573	0.882
Cat-M3W1G3	0.682	0.912	0.185	0.593	0.877
Little Brown Bat-G1P2E0	0.678	0.902	0.135	0.572	0.882
ABCD2 Subfamily					
Human	0.573	0.678	0.148	0.466	0.904
Mouse	0.597	0.656	0.159	0.471	0.903
Rat	0.649	0.63	0.181	0.487	0.899
Chimpanzee	0.573	0.729	0.139	0.48	0.901
Bovine	0.494	0.757	0.186	0.479	0.901
Sheep	0.558	0.762	0.147	0.489	0.899
Rabbit	0.605	0.392	0.453	0.483	0.9
Ground Squirrel	0.584	0.659	0.195	0.479	0.901
Chicken	0.6	0.897	0.164	0.554	0.886
Little Brown Bat	0.657	0.809	0.114	0.527	0.891

Mallard	0.694	0.72	0.324	0.579	0.88
Duck bill platypus	0.714	0.473	0.652	0.613	0.873
Cat	0.643	0.895	0.073	0.537	0.889
ABCD3 Subfamily					
Human	0.689	0.941	0.229	0.62	0.872
Mouse	0.671	0.866	0.277	0.604	0.875
Rat	0.677	0.93	0.231	0.612	0.873
Bovine	0.709	0.936	0.191	0.612	0.874
Chimpanzee	0.709	0.94	0.206	0.618	0.872
Dog	0.73	0.922	0.149	0.6	0.876
Horse	0.67	0.966	0.249	0.628	0.87
Chicken	0.658	0.923	0.174	0.585	0.879
Rabbit	0.709	0.895	0.183	0.595	0.877
Giant Panda	0.714	0.918	0.162	0.598	0.876
Little Brown Bat	0.711	0.921	0.156	0.596	0.877
Zebra finch-H0Z5M6	0.688	0.704	0.627	0.673	0.861
Sheep-W5Q740	0.689	0.935	0.716	0.78	0.861
Mallard	0.671	0.905	0.247	0.608	0.874
Cat	0.7	0.961	0.196	0.619	0.872
Duck bill platypus-F65BNS (Deleted M)	0.713	0.892	0.859	0.821	0.87
Duck bill platypus-F65BNS (With M)	0.713	0.952	0.333	0.666	0.862
Ground Squirrel	0.693	0.909	0.181	0.594	0.877
Opossum	0.705	0.96	0.384	0.683	0.859
ABCD4 Subfamily					
Human	0.634	0.946	0.066	0.549	0.887
Mouse	0.655	0.977	0.191	0.608	0.874
Rat	0.567	0.847	0.341	0.585	0.879
Chimpanzee	0.624	0.97	0.246	0.613	0.873
Dog	0.637	0.961	0.084	0.561	0.884
Bovine	0.628	0.842	0.141	0.537	0.889
Horse	0.642	0.977	0.26	0.626	0.871
Zebrafish	0.681	0.981	0.079	0.58	0.88
Chicken	0.675	0.98	0.12	0.592	0.878
Rabbit	0.625	0.977	0.468	0.69	0.857
Sheep	0.64	0.966	0.126	0.577	0.881
Giant Panda-D2GZW5	0.639	0.922	0.729	0.763	0.858
Zebrafinch	0.678	0.906	0.448	0.677	0.86
Little Brown Bat	0.674	0.945	0.093	0.571	0.882

Ground Squirrel-I3MEC2	0.754	0.972	0.712	0.813	0.868
Opossum	0.7	0.975	0.195	0.623	0.871
African Elephant-G3SRA7	0.709	0.787	0.352	0.616	0.873

Appendix Table3.5: List of orthologs of whole ABCG family selected for TMCrys screening. Protein sequences were obtained from the UniProt database (https://www.uniprot.org/)

ABCG Family					
Organisms	**Solubilisation**	**Purification**	**Crystallization**	**Whole Process**	**Reliability of prediction**
ABCG1 Subfamily					
Human-P45844	0.717	0.958	0.188	0.621	0.872
Mouse-Q64343	0.692	0.949	0.221	0.621	0.872
Rat-G3V642	0.716	0.951	0.198	0.622	0.872
Bovine-A0A3Q1M833	0.749	0.961	0.185	0.631	0.87
Dog-J9P4X7	0.76	0.964	0.208	0.644	0.867
Cat-M3WHF1	0.716	0.956	0.209	0.627	0.871
Chicken-E1BYW4	0.7	0.966	0.182	0.616	0.873
Chicken-E1BYW4(M deleted)	0.7	0.957	0.853	0.836	0.873
Horse-A0A3Q2H070 (With M)	0.688	0.961	0.309	0.653	0.865
Horse-A0A3Q2H070 (Deleted M)	0.702	0.922	0.883	0.836	0.873
Duck bill platypus-F6VMA6 (Deleted M)	0.737	0.927	0.877	0.847	0.875
Duck bill platypus-F6VMA6 (With M)	0.737	0.959	0.229	0.642	0.867
Mallard-U3IMT4 (WITH M)	0.722	0.965	0.201	0.629	0.87
Mallard-U3IMT4 (Deleted M)	0.722	0.953	0.87	0.848	0.875
Little Brown Bat-G1NSJ6-Deleted M	0.724	0.975	0.897	0.861	0.878
Little Brown Bat-G1NSJ6-With M	0.717	0.981	0.265	0.654	0.865
Giant Panda-	0.738	0.959	0.178	0.625	0.871

G1L5M6					
Sheep-W5PLT3	0.737	0.954	0.146	0.612	0.873
African Elephant-G3T382	0.701	0.95	0.227	0.626	0.871
Ground Squirrel-I3MWJ9	0.708	0.951	0.189	0.616	0.873
ABCG2 Subfamily					
Human-Q9UNQ0 With M	0.677	0.958	0.09	0.575	0.879
Human-Q9UNQ0 Deleted M	0.68	0.929	0.723	0.777	0.861
Mouse	0.657	0.975	0.107	0.58	0.88
Rat	0.669	0.964	0.088	0.574	0.881
Pig	0.657	0.972	0.094	0.574	0.881
Bovine	0.699	0.938	0.101	0.579	0.88
Chicken	0.637	0.981	0.16	0.593	0.878
Dog-B3RFJ1	0.676	0.975	0.09	0.58	0.88
Chimpanzee	0.699	0.95	0.106	0.575	0.881
Horse-F6Q6E8	0.686	0.954	0.093	0.578	0.881
Opossum	0.621	0.979	0.092	0.564	0.884
Goat	0.72	0.95	0.109	0.593	0.877
Cat	0.701	0.975	0.075	0.584	0.879
Zebrafinch	0.629	0.936	0.119	0.561	0.884
Mallard	0.681	0.934	0.674	0.763	0.858
ABCG4 Subfamily					
Human	0.73	0.939	0.227	0.632	0.869
Mouse	0.72	0.93	0.92	0.69	0.857
Rat	0.716	0.922	0.197	0.612	0.874
Bovine	0.714	0.944	0.252	0.636	0.869
Dog	0.714	0.944	0.252	0.636	0.869
Horse	0.714	0.958	0.249	0.64	0.868
Chimpanzee	0.73	0.939	0.227	0.632	0.869
Chicken	0.713	0.932	0.213	0.62	0.872
Zebrafish	0.701	0.929	0.211	0.613	0.873
Opossum	0.745	0.919	0.226	0.63	0.87
Cat	0.714	0.925	0.166	0.601	0.876
Ground Squirrel	0.708	0.93	0.291	0.643	0.867
Giant Panda	0.716	0.944	0.25	0.637	0.868
ABCG5 Subfamily					
Human	0.685	0.98	0.483	0.716	0.852
Mouse	0.654	0.966	0.299	0.64	0.868

Rat	0.718	0.97	0.193	0.627	0.87
Chimpanzee	0.685	0.974	0.374	0.678	0.86
Bovine	0.713	0.967	0.171	0.617	0.873
Dog	0.602	0.975	0.245	0.607	0.875
Chicken	0.654	0.964	0.132	0.584	0.879
Zebrafish	0.662	0.977	0.169	0.602	0.876
Opossum-F6X4Y4	0.707	0.973	0.127	0.602	0.876
Ground Squirrel	0.657	0.973	0.177	0.603	0.875
Cat	0.705	0.931	0.081	0.572	0.882
Little Brown Bat	0.623	0.964	0.215	0.601	0.876
Sheep	0.695	0.964	0.238	0.632	0.869
Rabbit	0.615	0.96	0.275	0.617	0.873
Zebrafinch-H0YZT4	0.644	0.975	0.753	0.79	0.863
Giant Panda	0.636	0.97	0.253	0.62	0.872
ABCG8 Subfamily					
Human	0.621	0.915	0.095	0.544	0.888
Mouse	0.695	0.942	0.085	0.574	0.881
Rat	0.717	0.935	0.067	0.573	0.882
Chimpanzee	0.621	0.913	0.073	0.536	0.889
Dog	0.657	0.96	0.117	0.578	0.881
Bovine	0.594	0.909	0.107	0.537	0.889
Horse	0.592	0.937	0.741	0.757	0.856
Chicken	0.625	0.948	0.022	0.531	0.89
Cat	0.605	0.92	0.148	0.558	0.885
Opossum	0.669	0.906	0.488	0.687	0.858
Duck bill platypus-F7BLQ8	0.663	0.897	0.685	0.748	0.855
Rabbit	0.628	0.897	0.069	0.531	0.89
Sheep	0.596	0.922	0.107	0.541	0.888
Zebrafinch-H0YZX7	0.653	0.915	0.593	0.72	0.851
Giant Panda	0.646	0.933	0.228	0.603	0.875
Little Brown Bat	0.617	0.945	0.104	0.555	0.885
Mallard	0.635	0.912	0.578	0.708	0.854
Human ABCG5+ ABCG8					
	0.661	0.899	0.15	0.57	0.882

A

```
D-GFP    TCCACTCAATCTGCCTTATCCAAAGATCCAAACGAAA
S-GFP    TCCACTCAATCTGCCTTATCCAAAGATCCAAACGAAA
         ***********************************
```

B)

```
sp|P42212|GFP_AEQVI    HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFV
D-GFP                 HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFV
MS-GFP               HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFV
                      **********************    ******************
```
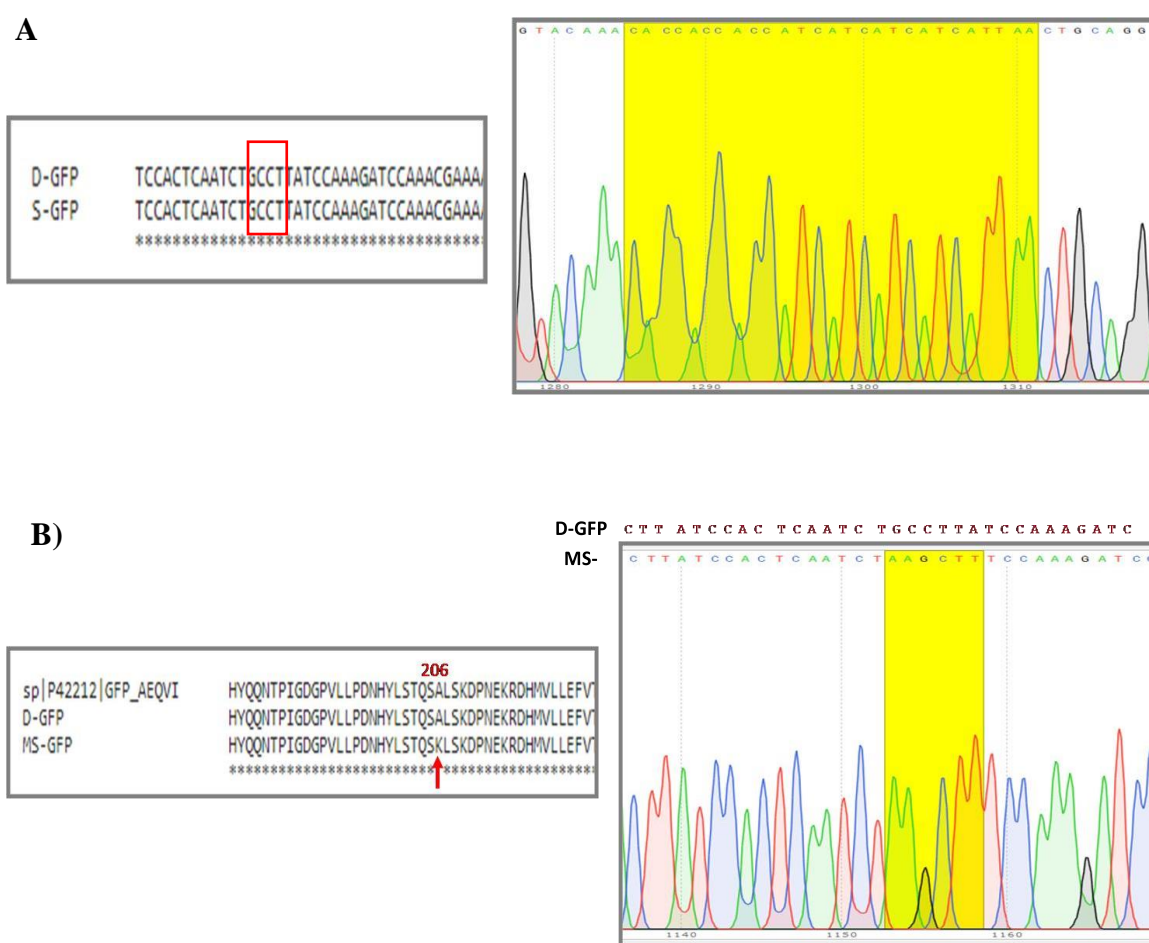
Appendix Figure 4.1: Sequencing results of plasmid DNA (p424GAL1). A) D-GFP: Sequence provided by David Drew. **S-GFP**: Sequencing results of miniprep plasmid DNA. Both sequences are compared by Clustal Omega. DNA sequencing confirmed the absence of A206K mutation in GFP tag. 8 His-tag confirmed by sequencing results. The highlighted sequence corresponds to 8 His tag. After 8 His tag, stop codon TAA is present. **B) sp|P422212|GFP-AEQVI**: GFP sequence from UniProt. **D-GFP**: Sequence provided by David Drew, **MS-GFP**: GFP sequence after A206K mutation. All sequences were compared using Clustal Omega. Results confirmed the lysine at 206 instead of alanine in MS- GFP.

Appendix Table 81.: Structures of ABC family on protein data bank (80 so far)

Organism	PDB Code	Method	Resolution	Authors	Doi
ABCA1					
Human (2017)	**5XJY**	cryo-EM	4.1 Å	Qian, H.W., Yan, N., Gong, X.	10.1016/j.cell.2017.05.020

ABCB1					
Mouse (2018)	**6GDI**	cryo-EM	7.9 Å	Nopnithi, Richard. Collins, Alessandro, Talha Shafi, Alistair Siebert and Robert C. Ford	10.1186/s12900-018-0098-z
Mouse (2018)	**6Q81** **Ligand:** ADP	cryo-EM	7.9 Å	Same as above	Same as above
ABCB1HM (Human mouse chimeric) (2018)	**6FN1** (Zosoquidar and UIC2 bound)	cryo-EM	3.58 Å	Alam, A., Locher, K.P.	10.1073/pnas.1717044115
ABCB1HM (Apo form) (2018)	**6FN4**	cryo-EM	4.14 Å	Same as above	Same as above
Human (2019)	**6QEX**	cryo-EM	3.6 Å	Alam, A., Kowal, J., Broude, E., Roninson, I., Locher, K.P.	10.1126/science.aav7102
(ABCB1HM)-EQ (2019)	**6QEE**	cryo-EM	3.9 Å	See above	See above
Human (2018)	**6C0V**	cryo-EM	3.4 Å	Kim, Y.J., Chen, J.	10.1126/science.aar7389
Mouse (2013)	**4KSB**	X-ray Diffraction	3.8 Å	Chang, G.	10.1073/pnas.1309275110
Mouse (2013)	**4KSC**	X-ray Diffraction	4.0 Å	See above	See above
Mouse (2013)	**4LSG**	X-ray Diffraction	3.8 Å	Chang, G., Szewczyk, P.	To be published
Mouse (2017)	**5KO2** (34 linker deleted mutant Hg derivative)	X-ray Diffraction	3.3 Å	Xia, D., Esser, L., Zhou, F.	10.1074/jbc.M116.755884
Mouse	**5KOY** (34 linker deleted bound with ATP)	X-ray Diffraction	3.85 Å	See above	See above
Mouse (2017)	**5KPD** (34 linker deleted double EQ mutant)	X-ray Diffraction	3.35 Å	See above	See above
Mouse native PGP	**5KPI**	X-ray Diffraction	4.01 Å	See above	See above
Mouse pgp methylated protein	**5KPJ**	X-ray Diffraction	3.5 Å	See above	See above
Corrected Structure of	**4M1M** (2014)	X-ray Diffraction	3.8 Å	Li, J., Jaimes, K.F., Aller, S.G.	10.1002/pro.2387

Mouse P-glycoprotein					
Mouse	**4Q9H** (2015)	X-ray Diffraction	3.4 Å	McGrath, A.P., Szewczyk, P., Chang, G.	10.1107/S1399004715000 978
P-gp cocrystallise d with QZ-Ala	**4Q9I** (2015)	X-ray Diffraction	3.78 Å	See above	See above
Pgp cocrystallise d with QZ-Val	**4Q9J** (2015)	X-ray Diffraction	3.6 Å	See above	See above
Pgp cocrystallise d with QZ-Leu	**4Q9K** (2015)	X-ray Diffraction	3.8 Å	See above	See above
Pgp cocrystallise d with QZ-Phe	**4Q9L** (2015)	X-ray Diffraction	3.8 Å	See above	See above
Pgp co-crystallized with BDE-100	**4XWK** (2015)	X-ray diffraction	3.5 Å	See above	See above
ABCB2/ TAP1					
Rat (2014)	**4K8O** Atpase domain of TAP1 with ATP (D645N, D651A)	X-ray diffraction	2.65 Å	Grossmann, N., Vakkasoglu, A.S., Hulpke, S., Abele, R., Gaudet, R., Tampe, R.	10.1038/ncomms6419
Human, Human herpesvirus 1	**5U1D** (2017)	cryo-EM	3.97 Å	Oldham, M.L., Chen, J., Grigorieff, N.	10.7554/eLife.21829
ABCB4					
Human (2019)	**6S7P**	cryo-EM	3.42 Å	Olsen, J.A., Alam, A., Kowal, J., Stieger, B., Locher, K.P.	10.1038/s41594-019-0354-3

				ABCB6	

Human (2010)	**3NH6** (Apo)	X-ray Diffraction	2.0 Å	Haffke, M., Menzel, A., Carius, Y., Jahn, D., Heinz, D.W.	[10.1107/S0907444910028593](#)
Human	**3NH9** (ATP bound)	X-ray Diffraction	2.1 Å	See above	See above
Human	**3NHA** (ATP & Mg)	X-ray Diffraction	2.1 Å	See above	See above
Human	**3NHB** (ADP)	X-ray Diffraction	2.15 Å	See above	See above

				ABCB8	

Human (2018)	**5OCH**	X-ray Diffraction	3.4 Å	Faust, B., Pike, A.C.W., Shintre,	To be published

				ABCB10	

Human (2013)	**4AYT**	X-ray Diffraction	2.85 Å	Pike, A.C.W., Shintre, C.A., Li, Q	[10.1073/pnas.1217042110](#)
Human	**3ZDQ** (Nu free)	X-ray Diffraction	2.85 Å	Pike, A.C.W., Shintre, C.A., Li, Q	[10.1073/pnas.1217042110](#)
Human	**4AYW** (Plate-form)	X-ray Diffraction	3.3 Å	See above	**See above**
Human	**4AYX** (Rod form B)	X-ray Diffraction	2.9 Å	See above	**See above**

				ABCC1	

Human (2014)	**4C3Z** NBD1 (Nu free)	X-ray Diffraction	2.1 Å	Chaptal, V., Gueguen-Chaignon,	To be published
Bovine (2017)	**5UJA** (Leukotriene C4 bound)	cryo-EM	3.34 Å	Johnson, Z.L., Chen, J.	[10.1016/j.cell.2017.01.041](#)
Bovine	**5UJ9**	cryo-EM	3.49 Å	See above	**See above**
Bovine	**6BHU** (ATP bound)	cryo-EM	3.14 Å	See above	**See above**
Human	**2CBZ**	X-ray Diffraction	1.5 Å	Ramaen, O., Leulliot, N.,	[10.1016/j.jmb.2006.04.005](#)

				ABCC6	

Human (2018)	**6BZS** (NBD1 apo form)	X-ray Diffraction	2.3 Å	Ran, Y., Zheng, A., Thibodeau, P.H.	[10.1074/jbc.RA118.004806](#)
Human (2018)	**6BZR** (NBD2 ADP bound)	X-ray Diffraction	2.8 Å	Zheng, A., Thibodeau, P.H.	To be published
Human (2018)	**6NLO** (NBD1 H812A Apo form)	X-ray Diffraction	2.85 Å	See above	See above

ABCC7/ CFTR					
Human (2018)	**6MSM** (ATP bound, phosphorylated)	cryo-EM	3.2 Å	Zhang, Z., Liu, F., Chen, J.	10.1073/pnas.1815287115
Mouse (2012)	**3SI7** (NBD1 delta F508 mutant)	X-ray Diffraction	2.25 Å	Brautigam, C.A., Caspa, E., Thomas, P.J.	10.1016/j.cell.2011.11.023
Human (2018)	**5TFI** (NBD1 bound to dGTP)	X-ray Diffraction	1.89 Å	Wang, C., Aleksandrov, A.A., Yang, Z.,	To be published
Human (2017)	**5UAK** (dephosphorylated, ATP free)	cryo-EM	3.87 Å	Liu, F., Zhang, Z., Chen, J.	10.1016/j.cell.2017.02.024
Mouse (2004)	**1Q3H** (NBD1 with AMP.PNP)	X-ray Diffraction	2.5 Å	Lewis, H.A., Buchanan, S.G., Burley, S.K.,	10.1038/sj.emboj.7600040
Human (2015)	**4WZ6** (aa 389-678 NBD1, delta F508 with 3 stabilizing mutations, ATP bound)	X-ray Diffraction	2.05 Å	Byrnes, L.J., Hall, J.	10.1002/pro.2821
Human (2018)	**5TF7** (NBD1 with ATP)	X-ray Diffraction	1.93 Å	Wang, C., Aleksandrov, A.A., Yang, Z.,	To be published
Human (2018)	**5TF8** (NBD1 with dTTP)	X-ray Diffraction	1.86 Å	See above	See above
Human (2018)	**5TFA** (NBD1 with dUTP)	X-ray Diffraction	1.87 Å	See above	See above
Human (2018)	**5TFB** (NBD1 with 7-methyl GTP)	X-ray Diffraction	1.87 Å	See above	See above
Human (2018)	**5TFC** (NBD1 with GTP)	X-ray Diffraction	1.92 Å	See above	See above
Human (2018)	**5TFD** (NBD1 with CTP)	X-ray Diffraction	1.89 Å	See above	See above
Human (2018)	**5TFF** (NBD1 with UTP)	X-ray Diffraction	1.89 Å	See above	See above
Human (2018)	**5TFG** (NBD1 with 5-methyl UTP)	X-ray Diffraction	1.91 Å	See above	See above
Human (2018)	**5TFJ** (NBD1 with dCTP)	X-ray Diffraction	1.85 Å	See above	See above
Human (2018)	**5TGK** (NBD1 with dATP)	X-ray Diffraction	1.912 Å	Wang, C., Aleksandrov, A.A., Yang, Z.,	10.1074/jbc.RA117.000819
Human (2007)	**2PZE** (NBD as head to tail dimer)	X-ray Diffraction	1.7 Å	Atwell, S., Conners, K., Emtage, S.	10.1093/protein/gzq004
Human (2007)	**2PZF** (NBD as head to tail dimer with delta F508 mutation)	X-ray Diffraction	2.0 Å	See above	See above
Human (2007)	**2PZG** (NBD as monomer)	X-ray Diffraction	1.8 Å	See above	See above
Mouse (2003)	**1R0W** (NBD1 apo)	X-ray Diffraction	2.2 Å	Lewis, H.A., Buchanan, S.G.,	10.1038/sj.emboj.7600040
Mouse (2003)	**1R0X** (NBD1 with ATP)	X-ray Diffraction	2.2 Å	See above	See above

Mouse (2003)	**1R0Y** (NBD1 with ADP)	X-ray Diffraction	2.55 Å	See above	See above
Mouse (2003)	**1R0Z** (NBD1 with ATP)	X-ray Diffraction	2.35 Å	See above	See above
Mouse (2003)	**1R10** (NBD1 with ATP, I4122 space grp)	X-ray Diffraction	3 Å	See above	See above
Mouse (2004)	**1XF9** (NBD1 F508S mutant)	X-ray Diffraction	2.7 Å	Thibodeau, P.H., Brautigam, C.A.,	10.1038/nsmb881

ABCD4

Human	**6JBJ**	cryo-EM	3.6 Å	Xu, D., Feng, Z., Hou, W.T.,	10.1038/s41422-019-0222-z

ABCG1

Human (2021)	**7OZ1**	cryo-EM	4 Å	Locher, K.P lab	10.1016/j.jmb.2021.167218
Human (2022)	**7FDV**	cryo-EM	3.26 Å	Xu, D., Feng, Z.,	10.1016/j.celrep.2022.110298
Human (2021)	**7R8E**	cryo-EM	3.7 Å	Sun, Y., Wang, J., Long, T., Qi, X	10.1073/pnas.2110483118
Human (2021)	**7R8D**	cryo-EM	3.2 Å	Same as above	Same as above
Human (2021)	**7R8C**	cryo-EM	3.7 Å	Same as above	Same as above

ABCG2

Human (2018)	**6ETI** (Inhibitor bound)	cryo-EM	3.1 Å	Jackson, S.M., Manolaridis, I., Locher, K.P.	10.1038/s41594-018-0049-1
Human (2018)	**6FEQ**	cryo-EM	3.6 Å	See above	See above
Human (2018)	**6HBU** (E211Q mutant with ATP& Mg)	cryo-EM	3.09 Å	Locher, K.P.	10.1038/s41586-018-0680-3
Immune system	**5NIV** (5D3 Fab)	X-ray diffraction	1.498 Å	Manolaridis, I., Locher, K.P.	10.1038/nature22345
Human (2017)	**5NJ3**	cryo-EM	3.78 Å	Locher, K.P.	10.1038/nature22345
Human (2017)	**5NJG**	cryo-EM	3.78 Å	Locher, K.P.	10.1038/nature22345
Human (2018)	**6HCO** (E211Q bound to estrone 3 sulfate& 5D3-Fab)	cryo-EM	3.58 Å	Locher, K.P.	10.1038/s41586-018-0680-3
Human (2018)	**6HIJ** (ABCG2-MZ29-Fab with cholesterol& PE lipids)	cryo-EM	3.56 Å	Locher, K.P.	10.1038/s41594-018-0049-1
Human (2018)	**6HZM** (E211Q mutant bound to Mg& ATP)	cryo-EM	3.09 Å	Locher, K.P.	10.1038/s41586-018-0680-3

Human (2018)	**6FFC** (Inhibitor bound)	cryo-EM	3.56 Å	Locher, K.P.	10.1038/s41594-018-0049-1
ABCG5/8					
Human (2016)	**5DO7**	X-ray Diffraction	3.93 Å	Lee, J.-Y., Kinch, L.N., Borek, D.M.,	10.1038/nature17666

Publication
Book Chapter 2

ATP-Binding Cassette Transporters: Snap-on Complexes?

Iqra Younus, Sofia Kochkina, Cheri C. Choi, Wenjuan Sun, and Robert C. Ford

Abstract

ATP-binding cassette (ABC) transporters are one of the largest families of membrane proteins in prokaryotic organisms. Much is now understood about the structure of these transporters and many reviews have been written on that subject. In contrast, less has been written on the assembly of ABC transporter complexes and this will be a major focus of this book chapter. The complexes are formed from two cytoplasmic subunits that are highly conserved (in terms of their primary and three dimensional structures) across the whole family. These ATP-binding subunits give rise to the name of the family. They must assemble with two transmembrane subunits that will typically form the permease component of the transporter. The transmembrane subunits have been found to be surprisingly diverse in structure when the whole family is examined, with seven distinct folds identified so far. Hence nucleotide-binding subunits appear to have been bolted on to a variety of transmembrane platforms during evolution, leading to a greater variety in function. Furthermore, many importers within the family utilise a further external substrate-binding component to trap scarce substrates and deliver them to the correct permease components. In this chapter, we will discuss whether assembly of the various ABC transporter subunits occurs with high fidelity within the crowded cellular environment and whether promiscuity in assembly of transmembrane and cytoplasmic

components can occur. We also discuss the new AlphaFold protein structure prediction tool which predicts a new type of transmembrane domain fold within the ABC transporters that is associated with cation exporters of bacteria and plants.