Bioinformatic analysis of the origins of ABCA and ABCG families and development of the thermostability assay for ABCG2

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

2021

Abilasha Balakrishnan

School of Biological science

Table of Contents

Final word count: 42094

List of figures

List of Tables

List of Abbreviations

ABC	ATP-binding cassette
ABCA1	ATP-binding cassette super-family A member 1
ABCA1 TMD1	ABCA1 transmembrane domain 1
ABCA1 TMD2	ABCA1 transmembrane domain 2
ABCA1 IH1 TM 1- 4	ABCA1 transmembrane 1 to 4 including IH1
ABCA1 IH3 TM 7-10	ABCA1 transmembrane 7 to 10 including IH3
ABCA1 NBD1	ABCA1 nucleotide binding domain 1
ABCA1 NBD2	ABCA1 nucleotide binding domain 2
ABCA4	ATP-binding cassette super-family A member 4
ABCC6	ATP-binding cassette super-family C member 6
ABCC7	ATP-binding cassette super-family C member 7
ABCG2	ATP-binding cassette super-family G member 2
ABCG2 TMD	ABCG2 transmembrane domain
ABCG2 TM1-4	ABCG2 transmembrane 1 to 4
ABCG2 NBD	ABCG2 nucleotide binding domain
ABCG5/G8	ATP-binding cassette super-family G member 5 and 8
ABCG5/G8 TMD	ABCG5G8 transmembrane domain
ABCG5 TM1-4	ABCG5 transmembrane 1 to 4
ABCG8 TM1-4	ABCG8 transmembrane 1 to 4
ABCG5 NBD	ABCG5 nucleotide binding domain
ABCG8 NBD	ABCG8 nucleotide binding domain
AEBSF	Aminoethylbenzenesulfonyl fluoride
ATP	Adenosine triphosphate
BBB	Blood brain barrier
BSA	Bovine serum albumin protein
CRD	C-terminal regulatory domain
CETSA	Cellular thermal shift assay
CF	Cystic Fibrosis
CFTR	the Cystic fibrosis transmembrane conductance regulator
CMC	Critical micelle concentration
Cryo-EM	Cryogenic electron microscopy
DDM	n-dodecyl-β- D- maltoside
DMSO	Dimethylsulfoxide

DTT	Dithiothreitol
2D	2-dimentional
3D	3-dimentional
E-64	Epoxysuccinyl-Leucylamido-Butane
E.coli	*Escherichia coli*
ECF	Energy coupling factor
ECL	Extracellular loop
ECD1	Extracellular domain 1
ECD2	Extracellular domain 2
EDTA	Ethylenediaminetetraacetic acid
EH	Extracellular helix
ER	Endoplasmic reticulum
EL3	External cellular loop
GFP	Green fluorescent protein
HCL	Hydrochloric acid
HDL	High density lipoprotein
His	Histidine
ICL	Intracellular
IM	Inner membrane
K_2HPO_4	di-Potassium hydrogen orthophosphate 3-hydrate
KH_2PO_4	Potassium dihydrogen orthophosphate
MP	Membrane proximal
MQ	Milli-Q
MWCO	Molecular weight cut off
NaCL	Sodium chloride
NBD	Nucleotide binding domain
NBD1	Nucleotide binding domain 1 of ABCA1
NBD2	Nucleotide binding domain
OD	Optical density
OM	Outer membrane
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
P-gp	P-glycoprotein
PIs	Protease Inhibitor

PMSF	Phenylmethylsulfonyl fluoride
PSC	Periplasmic subdomain N
PSN	Periplasmic subdomain C
Q score	Quality score
RCSB	Research Collaboratory for Structural Bioinformatics
RMSD /rmsd	Root mean square deviation
SBD	Substrate binding domain
SDM	Structural Distance Measure
TGS	Tris Glycine SDS
TMD	Transmembrane domain
UniProt	Universal protein Resources
v/v	Volume/Volume
w/w	weight/volume
YET	Yeast Extract Trypton
YNB	Yeast Nitrogen Base

Abstract

The ATP-binding cassette super-family G member 2(ABCG2); also known as the breast cancer resistant protein) mapped on chromosome 4q22 is known to have important physiological roles in many tissues such as mammary gland, blood brain barrier and gastrointestinal tract. Furthermore, ABCG2 is involved in the transportation of chemically diverse substrates such as steroids and other organic ions. Most importantly, ABCG2 has been suggested to be involved in multidrug resistance in chemotherapy. Understanding the thermostability of the ABCG2 protein, will allow a better understanding of how drugs and inhibitor interact with the protein. The first aim in this thesis was to express, characterise and purify ABCG2 using *Pichia pastoris*, hence establish whether the thermostability cellular thermal shift assay (CETSA) showed proof of principle on ABCG2. Results concluded that *Pichia pastoris* is an excellent expression system which can be used to express and characterise ABCG2. The thermostability CETSA assay performed on ABCG2 showed that the melting temperature was 55°C, thus, proving that the CETSA assay could be used in future to examine the thermostability of ABCG2 in the presence of a ligand. Understanding the structure and mechanism of ABCG2 starts by looking at the evolutionary relationship with a common ancestor. The second aim of this thesis was to find potential evolutionary relationship between ABCA and ABCG families to bacterial ABC transporter such as: mechanotransducer MacB, WzmWzt, MlaE and MlaF. Results showed MacB transmembrane domain (TMD), and nucleotide binding domain (NBD) had significant alignment to TMD and NBD of ATP-binding cassette super-family A member 1 (ABCA1), ABCG2 and ATP-binding cassette super-family G member 5 and 8 (ABCG5/G8). Henceforth, indicating ABCA and ABCG families may have originated from MacB transporters. Similarly, the TMD and NBD of WzmWzt, MlaE and MlaF are seen to align significantly well to the TMD and NBD of ABCA and ABCG families, this suggest that convergence evolution may be taking place between the two families.

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 institute 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 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 commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=2442 0), 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

Acknowledgement

I would like to firstly convey my greatest appreciation to my supervisor Prof Robert Ford, who has kindly guided me in my laboratory work and theoretical studies. He has offered me a very interesting project and has always been patient and helpful. Secondly, I would like to thank all the members in Ford Lab, especially Talha Shafi and Iqra Younus for providing advice and always being supportive. I would like to thank my friends Megan Eadsforth and Sonia Hussain for making my time at the University of Manchester unforgettable. Finally, I would like to thank my parents for the continued love and support.

Chapter 1: Introduction

1.1 ATP binding Cassette family

In all living organisms the cell membrane has a crucial role to play, such that it needs to maintain the cell's integrity and the equilibrium conditions (Wilkens, 2015; Hulme *et al*., 2010). Maintenance of the steady state conditions within the cell is assisted through the presence of membrane protein transporters (Robey *et al*., 2018). The ATP binding cassette (ABC) family is the largest protein transporter superfamily, which is present in all organisms. As ABC transporters are found in all living organisms this proves how understanding the structure and mechanism can help in understanding and treating various diseases associated to disfunction of the ABC transporter (El-Awady,2017; Becker *et al*., 2007).

ABC transporters can be found in both prokaryotic and eukaryotic organisms; hence they are known to serve many functions. Unlike the eukaryotic transporter, the prokaryotic transporter has the ability to influx nutrients into the cell, efflux out cell waste and toxins of the cell (Sheps *et al*., 2004). With a few exceptions, the eukaryotic ABC transporters also known as 'efflux transporter' have the feature of only expelling out toxins, and unwanted cell product out of the cell (Nasim *et al*., 2020). The question still remains to this day as to why the eukaryotic transporter has most probably lost the ability to influx substrates into the cell? (El-Awady,2017; Moitra *et al*, 2011; Rice *et a*l., 2014). Some of the answers to this question can be found by looking at the evolution of the protein through bioinformatic analysis and comparing the structure to its nearest evolutionary ancestors such as bacteria, to determine an understanding of the mechanism of transport which could be beneficial in developing treatments for diseases.

There are only four fundamentally different classes of membrane-bound protein transporters: ion channels, transporters, aquaporins and ATP powered pumps. The mechanism used to transport substrates by the ABC transporters is via ATP powered pumps, these membrane protein transporters are involved in either primary active transport or secondary active transport (Vasiliou *et al*.,2009; Forrest *et al*., 2011). Primary active transport is involved in the mechanism of using energy in the form of ATP to transport substrates against a concentration gradient. Secondary active transport uses energy in other forms hence, this energy is generated through electrochemical gradient which is created through pumping ions out of the cell (Yang *et al.,* 2015; Mishra *et al.*, 2014). Primary transport mechanism is

assisted by the primary transporters which includes: the rotary motor ATPases (F-, A- and V-ATPase), P-type ATPase and ABC transporters (Wilkens, 2015; Hollenstein *et al*., 2007). To understand how ABC transporters functioned, the first step was to determine which category of transporters that it belonged to (Forrest *et al*., 2011).

It is known that the ATPase transporters use the hydrolysis of ATP to pump substrates into or out of the cells (Hollenstein *et al*., 2007), however, the rotary motor and p-type ATPases are only known to import and export metal ions and protons (Tomaszowski *et al*., 2015; Forrest *et al*., 2011). The ABC transporters are known to transport a large amount of chemically diverse molecules across the cell membrane (Robey *et al*., 2018; Jasinski *et al*., 2003). Therefore, this allows the ABC transporters to regulate the cellular homeostasis such as the levels of hormones, lipid molecules, saccharides, metal ions, oligonucleotides, xenobiotics and other molecules which are required for various physiological roles (Robey *et a*l., 2018; Wilkens,2015; Tarling *et al*., 2013; Jasinski *et al*., 2003).

ABC transporters are involved in intracellular compartmental transport, designed to regulate the levels of intracellular organelles such as mitochondrion, lysosomes, endoplasmic reticulum and Golgi apparatus (Hinz.,2012). ABC transporters are viewed as essential proteins for the cell, as loss of these proteins leads to a large number of genetic diseases (Tarling *et al*., 2013; Domenichini *et al*., 2019). The importance of the presence of the ABC transporters are seen through germline mutation and the associated diseases, such as: a mutation on ATP-binding cassette super-family C member 7 (ABCC7) also known as CFTR (cystic fibrosis transmembrane conductance regulator) results in cystic fibrosis, mutation in (ABCA1) results in Tangier disease; Stargardt Macular degeneration which is associated with mutation in which is associated with mutation in ATP-binding cassette super-family A member 4(ABCA4); mutations among two of the ABCG transporters, ABCG5 and ABCG8, have been associated with Sitosterolaemia, mutations on ATP-binding cassette super-family C member 6 (ABCC6), is associated with Psuedoxanthoma, and the mutations on ATP-binding cassette super-family A member 12 (ABCA12) have been associated with Harlequin Ichthyosis (Robey *et al*., 2018; Domenichini *et al*., 2019). There have been many research studies which have been carried out to manage these disorders such as gene therapy, stabilization of the messenger mRNA, and correction of trafficking defects to name a few (Robey *et a*l., 2018; Wilkens, 2015; Hollenstein *et al*., 2007). However, there has been many biological and clinical challenges which has led to the failure of these treatments due to the

lack of understanding on the mechanism of transport for each transporter (Aguiar *et al*., 2019). Understanding protein structure and mechanism enables one to understand how specific substrates are transported across the cell membrane. However, obtaining high resolution protein structures in action of transport is one of the most difficult tasks faced by structural biologists.

1.2 The general architecture and structure of ABC transporters

The first step to understand the functional links and properties of the ABC transporter is by looking at the fundamental layout of the ABC transporter. General architecture of ABC transporters are composed of four functional units; two transmembrane domains: transmembrane domain 1 (TMD1), transmembrane domain 2 (TMD2) embedded in the membrane bilayer. Furthermore, the two nucleotide binding domains: nucleotide binding domain 1(NBD1) , nucleotide binding domain 2 (NBD2) located in the cytoplasm, as seen in figure 1 (Rees *et al*., 2009; Linton,2007).

The combination of the domains are encoded by a single polypeptide in prokaryotes, or two polypeptides in eukaryotes, with the latter being half transporters (with either homodimeric or heterodimeric halves) like in the ABCG family (Lewinson and Livnat-Levanon., 2017). The four domains in bacteria may be found in a combination of individual pairs, pairwise identical subunits, or a combination of fused NBDs and/or TMDs (Wilkens., 2015). Although in bacteria there may be a varied combination of NBD and TMD, between eukaryotic and bacteria there are high sequence similarities in the NBD, thus many similar conserved regions as shown in the example of figure 1.

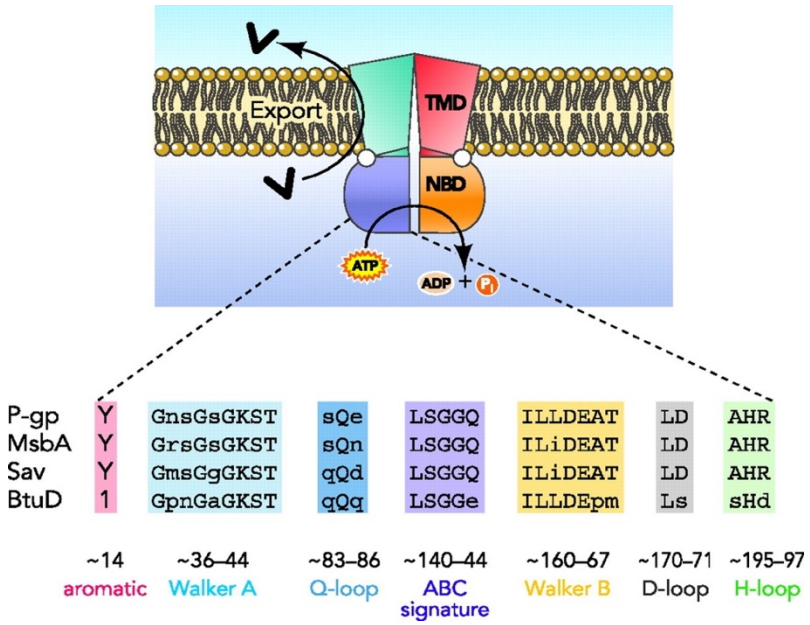


Figure 1: The fundamental features of ABC transporter and conserved regions in NBD

This figure is known as the 'minimal ABC transporter', the diagram represents the presence of two transmembrane domains which are embedded in the membrane bilayer; exporting out ligands is one of its functions whilst transport is driven by the ATP binding and hydrolysis by the two nucleotide binding domains. The sequence alignment shows there are seven highly conserved regions of the NBD throughout different families: P-gp(human ABC transporter), MsbA (ABC transporter which function as a homodimer and is found in Gram-negative bacteria), Sav (Bacterial multidrug ABC transporter), BtuD (ABC transporter found in Escherichia coli) – important in the process of uptake of Vitamin B12) (Linton,2007).

1.3 ABC Importers and Exporters

The general architecture of the ABC transporter superfamily provides the first features for detecting whether the membrane protein belongs in the family, when positively identified the membrane transporters are grouped into either exporters or importers depending on the basis of transport reaction, architecture and mechanism of transport (Hollenstein *et al*., 2007). Thus, the ABC importers are further divided into two classes (I and II) depending again on the small details of differences in the mechanism and architecture and a third class (III) known as energy coupling factor (ECF) transporters as shown in figure 2, however this transporter will not be discussed any further. The mechanism in which ABC importers transport their substrate is by requiring a binding protein that delivers captured substrates to the external face of the transporter, whereas ABC exporters recruits their substrate from their cytoplasm or from the inner leaflet of the bilayer. (Wilkens., 2015; Hollenstein *et al*., 2007; Lewinson and Livnat-Levanon., 2017). The presence of ABC importers and exporters are seen in bacteria, however, eukaryotes (with a few exceptions) are known to only have ABC exporters. ABCA4 is one of the few novel importers present in eukaryotes; it is found in

retinal photoreceptor cells and has association with Stargardt macular degeneration. One of the main questions related to ABCA4 is why is it one of the few importers found in eukaryotes? Furthermore, why are there only a few importers found in eukaryotes?

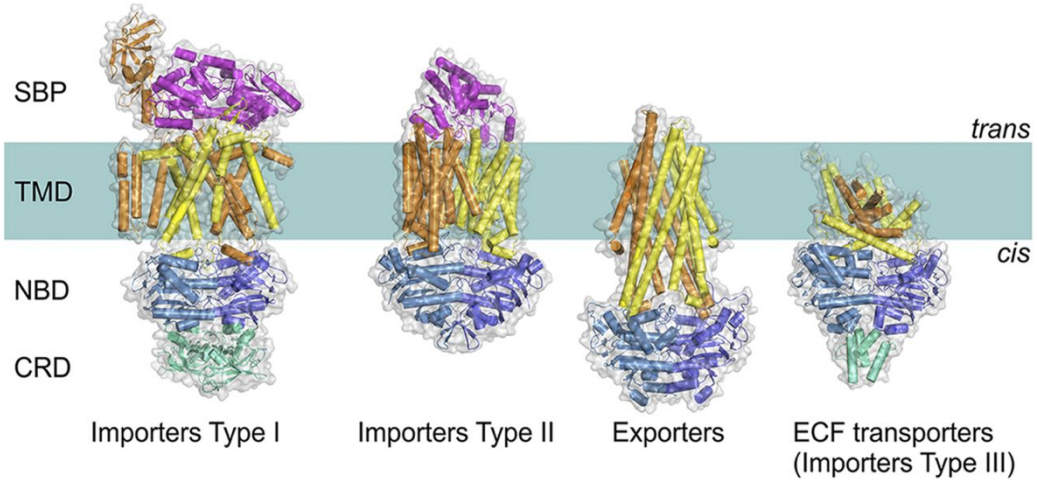


Figure 2: The difference in features of importers: types I, II, III and exporters

The diagram shows four distinct fold of ABC transporters, they all share a general structure with two NBD (shown as a blue and sky blue) and two TMD (orange and yellow). There are a few differences which can be seen in the diagram of which some of the transporters have extra domains (represented green). There has been association of this extra domain to have regulatory function – C-terminal regulatory domain (CRD). There are specific differences noticed between the importers and exporters; ECF as the energy coupling factor (which will not be discussed any further). Type I and II importers both have an substrate binding domain (SBD) represented in magenta, present in the periplasm for Gram negative bacteria or external space for Gram positive and Archaea; whereas the exporter does not have any SBD as it is only exporting compounds out of the cell and not importing it in the cell (ter Beek et al.,2014).

1.4 Structure and properties of Nucleotide binding domains (NBDs)

Determining whether an integral transmembrane protein belongs to the superfamily of ABC transporters is based on a few characteristics which are known as the hallmarkers. In both eukaryotes and bacteria, the sequence similarity of the NBD could be of high percentage between 30-50%, thus, through evaluation it shows there to be a similar three-dimensional fold and a conserved mechanism of energy coupling (Wilkens., 2015). In all ABC transporters the crucial role of the nucleotide binding domains is to act as the motor providing the driving force to facilitate the movement of the substrate (Chen *et al*.,2017). Regardless of the physiological role of the transporters in different classes, the three-dimensional organisation of the NBD remains highly conserved (Lewinson and Livnat-Levanon., 2017). In addition to the three-dimensional similarity, the NBD contains approximately 90-110 amino acids and can be identified at the sequence level with seven highly conserved motifs as seen in figure 3:

- **The A loop** – Positioning of the ATP for hydrolysis is an important role, thus the A loop is seen to contain a conserved aromatic residue which usually contains a tyrosine; hence this helps the positioning of the ATP through stacking of the adenine ring

- **Walker A motif or P loop - this** is the catalytic core domain hence the phosphate-binding loop, containing a highly conserved lysine residue. The lysine residue is crucial as the amide nitrogen and the ε- amino group is responsible for the interaction of the β- and γ- phosphate of the ATP.

- **The Walker B motif -** This motif is involved in coordinating the movement of the magnesium ion via a conserved aspartate residue. Another role of walker B motif is that the acidic residue at the end of walker B motif is likely to be the general base that polarizes the attacking water molecule. Furthermore, the presence and the role of the glutamate residue is still unclear and under debate, however recent crystal studies of the maltose transporter MaLEFGK2 from *Escherichia coli* (*E.coli*)suggests that glutamate is associated with functioning as the general base.

- **The D loop (motif SALD) –** as seen from figure 3, the D loop is seen to follow the walker B motif, from bioinformatic studies it is observed that the D loop from the two monomers run alongside each other and any conformational changes affect the D loop geometry of the catalytic site, hence assisting and forming the ATP hydrolysis site.

- **The H loop –** The H loop is also known as the switch region; this loop contains a highly conserved histidine which is responsible for forming the hinge between the β strand and the α-helix near the C-terminus of the NBD. This histidine is important as it interacts with the conserved aspartate from various conserved regions such as D loop, the glutamate residue on walker B motif and the γ- phosphate of the ATP.

- **The Q loop –** As seen from figure 3, the Q loop is located in the interface between RecA- like subdomain and the α-helix subdomain. It is observed that when there is a conformational change during ATP hydrolysis cycle the Q loop allows the movement of the glutamine in and out of the active site; it forms the active site when the Mg-ATP is bound and is disrupted when the ATP hydrolysis is complete**.** The Q loop is

also seen to be a crucial site for interactions with the TMDs, which suggests its role in ligand movement.

- **The ABC signature motif -** This motif is a significant and characteristic feature of the ABC superfamily, it is located on the N terminal end of the long helix, thus, assists a positive change of the helical dipole towards the γ- phosphate of the ATP during the ATP hydrolysis.

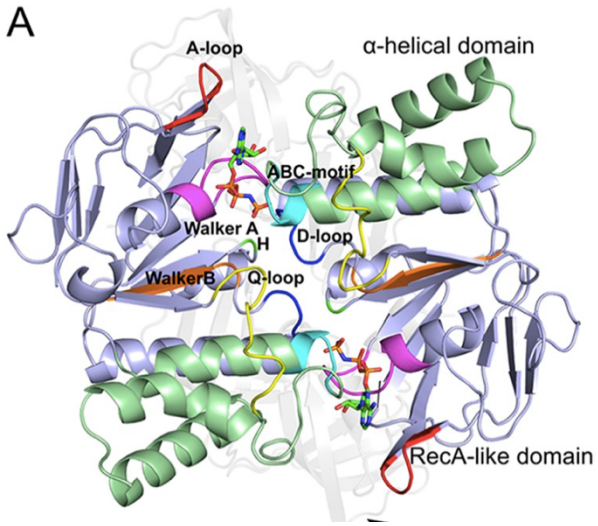


Figure 3: The conserved regions of NBD

In the catalytic core domain Walker A (P-loop) and Walker B (the residues in this motif is mainly hydrophobic); signature motifs (also known as the 'C' loop or the ABC signature motif); Q, H, D loop (Chen et al.,2017). Each of the motifs are believed to play an important role for binding and hydrolysing ATP to ADP+P and energy coupling, forming the ATP binding sites as well as the coupling of ATP hydrolysis to substrate translocation by the TMDs. Interaction between the motifs of separate NBDs is also crucial and shown when the P-loop binds to the signature motif which forms contacts with the nucleotide in the opposing NBD in the ATP- bound state (Hollenstein et al., 2007). Furthermore, the glutamate which is present in Walker B acts as the catalytic base for ATP hydrolysis; specifically, it is involved in the nucleophilic attack from H_2O on ATP (Hollenstein et al., 2007; Chen et al., 2017). The glutamate is seen to be a major contributor in hydrolysing ATP as it is seen to be conserved in both prokaryotic and eukaryotic ABC transporters. The histidine side chain on the H loop (also known as the 'switch motif') is thought to contribute to the catalytic reactions; the A loop contributes by providing an aromatic side chain (Lewinson and Livnat-Levanon., 2017; Hollenstein et al., 2007). More importantly, throughout all ABC transporters the NBD dimers are conserved and are important to form a head to tail dimer sandwich in order to hydrolyse the ATP, thus, this conformational change is dependent on the binding of the ATP (ter Beek et al., 2014).

1.5 Properties of the transmembrane domains (TMDs)

In contrast to the nucleotide binding domains, the TMDs lack the primary structure conservation. The transmembrane domains have been less well conserved throughout evolution and therefore they are highly heterogenous, hence producing a large diversity of transporters with varying function. The membrane spanning TMDs vary in primary sequence, length, architecture and the number of transmembrane helices (Hollenstein *et al*.,2007). The transmembrane domains consist of two highly hydrophobic domains which come together to form a channel, through which the substrate by-passes during translocation (Biemans-Oldehinkel, Doeven and Poolman., 2005). Typically, the transmembrane domain contains approximately 6-11 membrane-spanning **α** helices (with most exporters having 6), when the bundle of alpha helices gathers together, they form into a zigzag fashion (Stefkova *et al*., 2003). Furthermore, most commonly the ABC transporters form into a 'six -times-six' alpha helix structure, which yields approximately of 12 transmembrane segments per-functional units (Sridharan *et al*., 2019). However, the transmembrane segments may vary from 5-11 for each individual transmembrane domain (Biemans-Oldehinkel, Doeven and Poolman., 2005; Hollenstein *et al*.,2007).

1.6 Eukaryotic ABC transporters

To our current understanding, we know so far that ABC transporters are present in all living organisms, and the various classes of importers and exporters had probably evolved quite early in the development of life (Kos and Ford, 2009). It is possible that at some point in evolution eukaryotes had lost their importer, in Ward *et al* study the eukaryotic ABC exporter still retained a high resemblance to the bacterial ABC exporters: Sav1866 and MsbA (Ward *et al*., 2007; Arumugam *et al*., 2016). The understanding as to why the eukaryotes evolved to not have ABC importers still remains a mystery until this day, as many components of the eukaryotic cell still contains relics of prokaryotic cell such as the mitochondria.

The eukaryotic ABC transporters contain seven sub-families (ABCA, ABCB, ABCC, ABCD, ABCE, ABCF, ABCG) and 48 ABC systems (Dean *et al*., 2001). Although theABCE and ABCF families are classified as transporters, they do not function as transporters as they only contain twin nucleotide binding domains, of which the twin nucleotide binding domain function is still not known (Kerr *et al*., 2004; Karbanova *et al*., 2019). Furthermore, these two families are known to be involved in cellular processes such as ribonuclease inhibition and

translation control, but not transportation of substrates, since they are not transporters they will not be discussed further in this project (Xiong *et al*., 2015; Sun *et al., 2016*).

All the other eukaryotic ABC transporters are characterised with both nucleotide binding domains and transmembrane domains. The eukaryotic ABC genes are organised as full transporters, containing two nucleotide binding domains and two transmembrane domains, or as half transporters (Dean *et al*., 2001; Kerr., 2004). Through sequence analysis conducted in the Kerr *et al* study, it was concluded that the ABC transporters were a probable result from the gene fusion events, leading to at least two of the domains fused together and in most cases four of the domains are fused together (Kerr., 2004).

1.6.1 ABCA family

The ABCA family contains 12 other subfamily members, where they are associated with the involvement in lipid trafficking across various tissues (Xavier *et al*., 2019). Out of the ABC transporters the ABCA subfamily is known to contain the largest members, such that ATP-binding cassette super-family A member 13 (ABCA13) contains 5058 residues, making it the largest known protein in the ABC transporter family. Any mutation in the ABCA genes have resulted in genetic disorders related to lipid transportation, such as Tangier disease, high density lipoproteins (HDL) deficiency and age-related muscular degeneration (Vasiliou *et al*., 2008). Furthermore, one of the many studies conducted by Xiong *et al* have shown that ABCA1 have the highest sequence similarity to ABCG2; understanding the structure and function of ABCA1 will help to make connection to ABCG2 and a possible common ancestor (Xiong *et al*., 2016; Liu *et al*., 2019).

1.6.2 ABCA1 and its expression in tissue

ABCA1 is mapped on chromosome 9q31.1 (Albrecht *et al*., 2006); ABCA1 is predominantly found to be expressed in all human tissues but of a high abundance in the liver and lungs (He *et al*.,2019). The tissue expression of ABCA1 is seen to be the second highest in the lungs after the liver; suggesting its critical role in lipid homeostasis and transport in pulmonary function (Sonett *et al*., 2018). Alongside involvement in cellular lipid homeostasis, ABCA1 has been found to have a major role in anti-inflammatory function (Chai *et al*., 2017).

1.6.3 The structure of ABCA1 family

ABCA1 is part of the ABCA family, studies have suggested ABCA1 to be involved in mediating cellular efflux of the phospholipids and cholesterol to the extracellular acceptor apolipoprotein A-1, thus allowing the generation of nascent high -density lipoprotein. Any harmful mutations to ABCA1 are known to lead to diseases such as tangier disease and familial HDL (Qian *et al*., 2017; Kopecka *et al*., 2020). The cryogenic electron microscopy (cryo-EM) structure of the human ABCA1 mapped out by the Qian *et al* group, as shown in figure 4, shows the roles of each TMD and NBD in the lateral access mechanism proposed for the export of lipid.

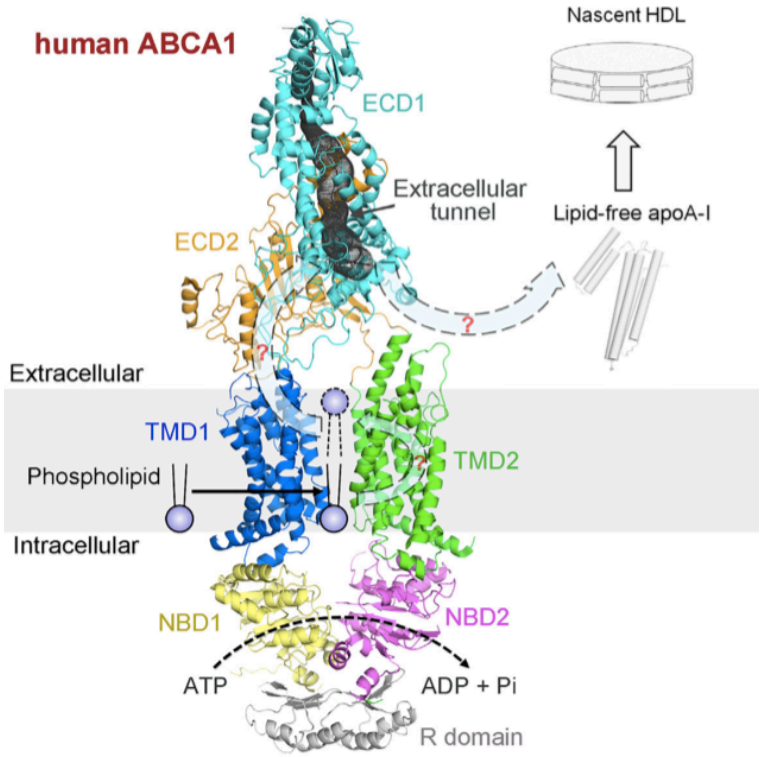


Figure 4: The human lipid exporter ABCA1

The human lipid exporter ABCA1 represented in this diagram has been resolved using cryo-EM at a resolution of 4.1A. The structure is an outward facing conformation in the nucleotide free state, there is an elongated hydrophobic tunnel which is formed by the extracellular domain. The structure contains 2 TMDs and each TMD contains 6 TMs, this is followed by the presence of 2 NBD and regulatory (R) domain; the NBD and the R domain are both components which are significantly conserved throughout evolution. Another unique property of ABCA1 is that it is the only member of ABCA subfamily which contains extracellular domain 1(ECD) between TM1 and TM2; extracellular domain 2 (ECD2) between TM7 and TM8 (Qian et al., 20 17).

1.6.4 ABCG family

The human ABCG subfamily contains five members (ABCG1, ABCG2, ABCG4, ABCG5 and ABCG8), the members of this family are represented as half transporters. ABCG1 is known to be involved in cholesterol efflux, whilst ABCG2 is known to be involved in multidrug resistance in chemotherapy and transportation of chemically diverse substrates such as steroids and organic ions (Lopez-Marques *et al,2015*; Rosenberg *et al*., 2015). The expression of ABCG5 and ABCG8 is predominantly seen in the liver, colon and intestines, and mutations to either of these genes has been linked sitosterolaemia (Vasiliou *et al*., 2008).

1.6.5 ABCG2 and its expression in tissue

ABCG2 is mapped on chromosome 4q22; initially ABCG2 was known as the breast cancer resistant protein (BCRP). It was found highly expressed in the placental tissue (Woodward *et al*.,2011; Taylor *et al*, 2017; Domanitskya *et al*.,2014). ABCG2 was discovered as a xenobiotic multidrug transporter, with important physiological roles in many tissues such as the placenta and mammary gland, maternal-foetal barriers, testis, blood brain barrier (BBB), liver, gastrointestinal tract and kidneys (Gil-Marins *et al*., 2020). Research has shown that ABCG2 has a minor role in the transportation of Uric acid. However, the major role associated with ABCG2 is the urate excretion in the kidney as the expression of ABCG2 was seen on the apical membrane of proximal tubules. A dysfunction in the transport of Uric acid results in several diseases such as hyperuricemia (one of the four stages of gout formation), kidney disease, and hypertension (Fetsch *et al*., 2006; Huls *et al*.,2008; Xu et al., 2017). A significant amount of ABCG2 expression is seen in the luminal surface of the brain microvessel endothelium, this suggests the involvement of ABCG2 and its importance to the blood brain barrier (Harwood *et al*., 2016). ABCG2 has been found to be a marker for a number of diseases, including cancer.

1.6.6 The structure of ABCG2

Recently accomplishments in understanding the architecture of ABCG2 have been made. As a member of the ABCG family, ABCG2 is a half transporter and contains one nucleotide binding domain and one transmembrane binding domain which are fused into a single polypeptide chain (Ford *et al*, 2020; Taylor *et al*., 2017). A few important features of the ABCG2 structures were found. For example, the ABC signature motif (ALSGGQ) was found

on the N-terminal nucleotide binding domain. Furthermore, six hydrophobic segments and an extracellular loop, which is located between transmembrane 5 (TM5) and transmembrane 6 (TM6), was found at the C terminal transmembrane domain as shown in figure 5 (Basseville *et al*., 2016; Manolaridis *et al*., 2018). ABCG2 existing as a homodimer is an important feature allowing it to work as an efflux pump. Furthermore, through crystallisation studies it has been observed that the homodimers form a disulphide bridge linkage at the cysteine 603 which is located on the extracellular loop, as shown in figure 5, and the molecular mass of the homodimer has been approximated to be 144kDa.

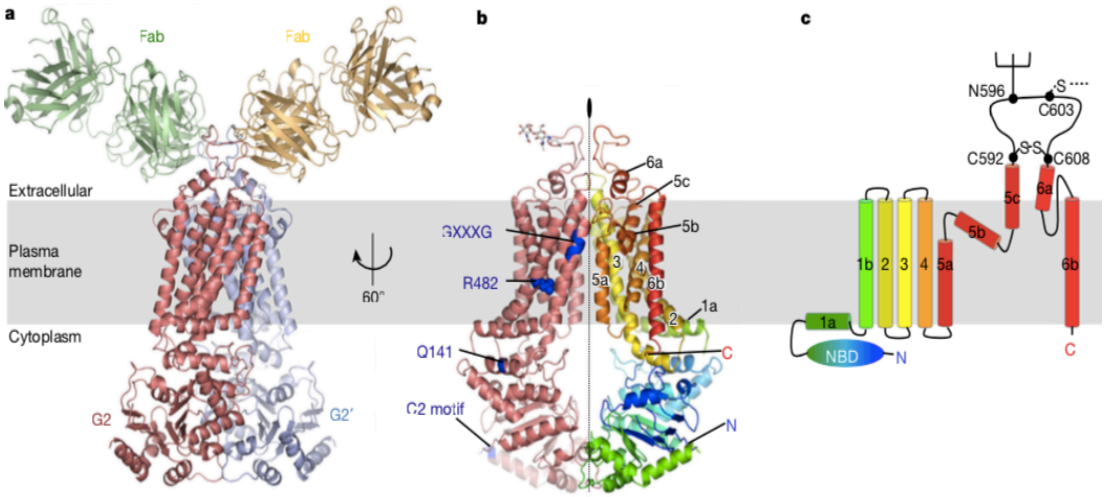


Figure 5: The structure of ABCG2

This figure shows three representations of ABCG2 ***a)*** *the ribbon representation of ABCG2-5D3 Fab complex.* ***b)*** *ABCG2 structure without the Fab complex being present, hence one of the monomers is coloured in salmon, the important motifs loops and transmembrane domains are coloured as a rainbow spectrum as in the third diagram. Furthermore, this figure shows the TMD to be made up of TM1a, TM1b, TM2, TM3, TM4, TM5a, TM5b, TM5c, TM6a, TM6b* ***c)*** *The topology is coloured in the rainbow spectrum and focuses on the disulphide bonds formed with the cysteines (C592, C603 and C608) and shows the presence of the NBD near the N terminal (Taylor et al., 2017).*

Similar to ABCG2, other members of the ABCG family are half transporters. Compared to the other eukaryotic ABC transporters the ABCG family members have a back to front arrangement, as a result this imposed difficulties in building a homolog model of ABCG2 and the ABCG members (Bickers *et al*., 2020). Although difficulties were encountered, structural studies were still conducted at a low resolution and were interpreted in terms of the bacterial multidrug homolog Sav1866-model, however this interpretation turned out to be wrong (Rosenberg.,2010; Ford *et al*., 2019). Structural studies on ABCG2 revealed that there were differences in the folds compared to the Sav1866-model and distinct differences compared to the ABCB subfamily. Results showed a smaller distance between the nucleotide binding

domain and the membrane, therefore the transmembrane helices and intracellular loop in ABCG2 are shorter compared to that of ABCB subfamily (Taylor *et al*., 2017; Ford *et al* 2019; Canet *et al*., 2015). As aforementioned, ABCG2 and ABCG5/G8 have a large structural similarity, as they belong from the same family, however there are some differences. The first difference noticed is the inward facing cavity which is formed by the shift in TM2 and TM5a present in ABCG2 and not in ABCG5/G8. The second difference noticed is the distinct structure of the external loop 3 (EL3) associated to each structure, this may be a result from the presence of cavity 2 in ABCG2 and not in ABCG5/ABCG8. However, in general it was noticed that there was a low structural and sequence similarity of EL3 between the ABCG subfamily members.

The first step in understanding how ABCG2 are involved in specific diseases starts by understanding the mechanism of substrate recognition and ATP driven transport. In the recent studies conducted by Taylor *et al*, the formation of an ABCG2-5D3 (Fab) complex revealed the intramolecular disulphide bonds were crucial for stabilising the structure of EL3. As EL3 contains the three cysteines (C592, C603 and C608) an absence or destabilisation of the structure will interfere with the disulphide bonds made by the three cysteines; this disruption interferes with C592 and C608 which are crucial for ABCG2 maturation and activity. Another residue important to the maturation of ABCG2 is N596, it was observed that N-glycosylation of N596 was essential, and a mutation of this residue resulted in protein destabilisation and an increased ubiquitin-mediated degradation (Taylor *et al*., 2017; Nakagawa *et al*., 2009; Wakabayashi *et al*., 2006).

In previous studies it was proposed that only one Fab fragment was required to inhibit the transportation activity of the asymmetric ABCB1 transporter. Therefore, as ABCG2 has a twofold symmetry the question was whether ABCG2 required two Fab fragments to inhibit its ATPase activity. It was observed in the Taylor *et al* study that only one fragment was required to clamp the ABCG2 monomers, thus prevent it from forming an outward facing conformation and preventing the transport of substrate.

1.6.7 The translocation pathway in ABCG2

The ability to transport chemically diverse substrates requires specific structural features enabling the protein to function efficiently. Analysis from the ABCG2-Fab complex revealed a deep slit like curve or cavity (cavity 1) when the protein was in an inward-open

conformation, where cavity 1 is shown to open to the cytoplasm and the inner part of the bilayer during this conformation. Furthermore, as cavity 1 has a large surface area and hydrophobic nature within ABCG2, this designated area is known as the multidrug binding pocket. There is also the presence of cavity 2 which is separated by the leucine residues (L554 and L554') which forms a plug between these two cavities (cavity 1 and cavity 2) as shown in figure 5 (Taylor *et al*.,2017). So far little is known about cavity 2 as it is inaccessible, however due to its less hydrophobic nature it is suggested to be involved in the transportation of substrates out of the cell.

1.6.8 The transport mechanism in ABCG2

The alternative access model is a mechanism initially proposed for ABC exporters and is the most widely accepted model for ABC transporters to date (Beis., 2015; Ford *et al*., 2019). The alternative access mechanism involves the transporter alternating from an inward and outward facing state to expose the ligand binding site to the substrate and translocate it from inside the cell to the outside or vice versa, with the help of ATP binding and hydrolysis. However, as more structural studies are being investigated at higher resolution it has come to light that there may be other mechanisms that the ABC transporters adopt, such as the transient opening, and the outward only mechanism (Bountra *et al*., 2017; Ford *et al*., 2019).

In general, most studies to date on ABCG2 have proposed that the transporter adopts an ATP driven alternating access mechanism. Once the substrate is bound, the ATP conformation changes from an outward facing state to an inward facing state and there is a closure of the nucleotide binding domain interface (Petravicius *et al*., 2019). The inward-open conformation of ATP to an outward-open conformation allows the movement of the substrate to cavity 2 (Taylor *et al*.,2017). As cavity 2 has a less hydrophobic nature, it has been suggested that it facilitates the release of the substrate outside of the cell. In a recent study conducted by Manolaris *et al*, the high resolution Cryo-EM structure of ABCG2 revealed that during an ATP driven conformational change they noticed a ridged body shift in the transmembrane domain and a change in the orientation of the nucleotide binding domain (Manolaris *et al*., 2018). Once the substrate is released and no ATP is bound to ABCG2 and the hydrophobic substrate binding site (cavity 1) is vacant, it starts recruiting substrates from the cytoplasm (cavity 2 is empty during this stage). Although there is a current in-depth understanding of how the alternating access mechanism works in ABCG2, finding similarities in the TMDs

WzmWzt and MlaE/MlaF bacterial ABC transporters, the knowledge of mechanism of transport may change to some degree.

1.6.9 ABCG5/G8 and its structure

In all recent studies conducted it has been reported that ABC transporters have an important role in maintaining sterol balances in higher eukaryotes (Lee *et al*., 2016). ABCG5/G8 has been reported to have an important role in mediating excretion of neutral sterols in both the liver and intestines (Lee *et al*., 2016: *Zein et al*., 2019). Furthermore, mutations found in ABCG5/G8 were associated to Sitosterolaemia, this disorder is characterised by sterol accumulation and premature atherosclerosis (Lee *et al*.,2016). The epic structural determination of the crystal structure of ABCG5/G8 was achieved by Lee *et al,* ABCG5/G8 was revealed to function as a heterodimer and 6 transmembrane helices were found in the TMD of ABCG5 and ABCG8. ABCG5/G8 was found to contain an extracellular domain (ECD) which has relatively close contact to TMD as seen in figure 6. It was observed, even in the absence of ATP, the NBD had a tightly closed dimer. At the NBD site, the ABC cassette and TMD were found to have close proximity to the triple helix bundle as seen in figure 6.

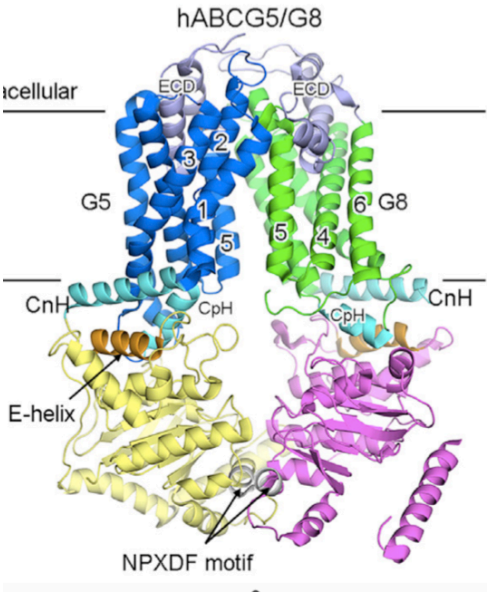


Figure 6: The structure of ABCG5/G8

ABCG5 is represented as the blue ribbons, ABCG8 is represented as the green ribbon, ABCG5 NBD is represented as the yellow ribbon, ABCG8 NBD represented in the pink ribbon. This figure illustrates ABCG5 TMD and ABCG8 TMD to be composed of 6 transmembrane helices. ECDs and triple bundle helix is seen in closed contact to the TMDs (Qian et al., 2017)

Alongside publishing the crystal structure of ABCG5/G8, it was also postulated that the overall structure of ABCG5/G8 had high resemblance to bacterial importers: ModBC, MalFGK2 and MetNI (Lee *et al*., 2016). However, in Ford *et al* review this postulation was disagreed based on several features which were found to not align with the theory (Ford *et al., 2019*). In the TMD of the bacterial importers the loops connecting them were very short and there was no domain swapped cytoplasmic loop, when compared with ABCG5/G8. It was concluded in the review that although the overall shape had similarity, the detailed arrangement of the six transmembrane α helices of ABCG5/G8 and ABCG2 were very different compared to ModBC. In general, it has been extremely hard to find the origin of the ABCG family.

1.7 Bacterial ABC transporters

In the Ford *et al* review, alongside disagreeing with the theory of ABCG family having high resemblance to ModBC as mentioned earlier, as an alternative bacterial mechanotransducers MacB was postulated to have very similar arrangement to ABCG and ABCA family. Although the theory was suggested, the results of the detailed arrangement of the TMD and NBD of ABCA family and ABCG family to the mechanotransducers MacB was still to be made. If there was significant structural alignment of the ABCG and ABCA family to MacB it can potentially suggest that ABC transporters adopt a more classical transmembrane pumping function.

1.7.1 MacAB-TolC

MacAB – ToLC (MacA - MacB – ToLC) assembly is an ABC- type tripartite multidrug efflux pump which is found mainly in *E.coli* and other Gram-negative bacteria. However, it can be found in *Streptococcus pneumonia* which is a gram- positive bacteria; the *spr0693-spr0694-spr0695* operon of *Streptococcus pneumonia* encodes for the ATP- Binding cassette type efflux pump (Yang *et al*., 2018). In general, all MacAB-TolC efflux pumps and other efflux pumps with the similar characteristics spans in the cell envelope, hence it is known as the transmembrane machine with the function of extruding a range of substrates including antibiotics such as macrolide, polypeptide virulence factor and enterotoxin STII. Each part of the MacAB-Tol C efflux pump has a crucial role in the efficient transport of substrates.

1.7.2 Mac A

MacA is also known as the periplasmic adaptor protein; the specific construction in this hexamer adaptor protein is thought to act as a gate and form a nanotube channel with a gating ring which helps to regulate the flow of substrates through the MacAB-TolC efflux pump as shown in figure 7 (Fitzptrick *et al*., 2017).

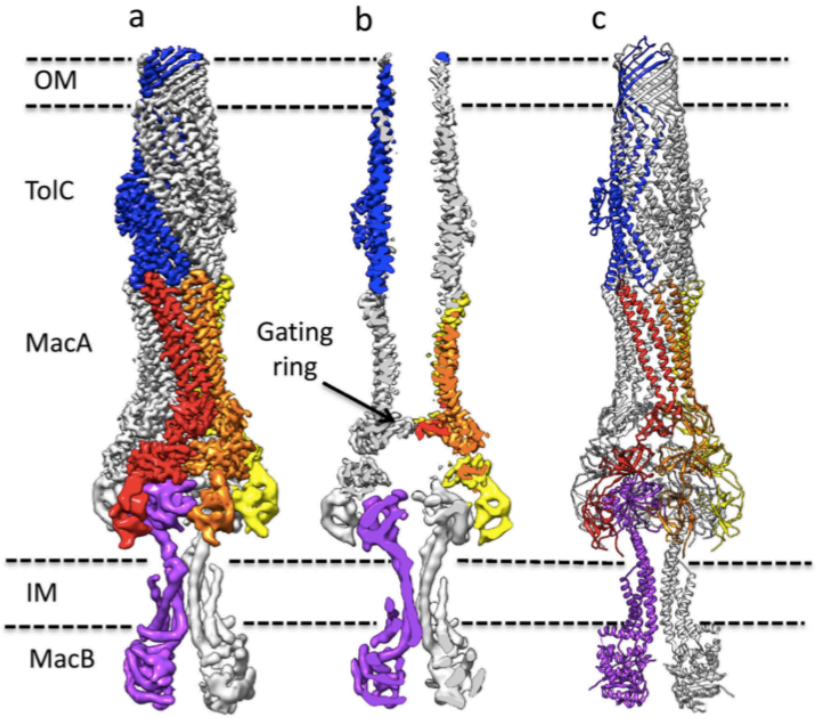


Figure 7: MacAB-TolC efflux pump

This figure represents the Pseudo-atomic model for the MacAB-TolC pump. IM – inner membrane, OM – outer membrane. ***A)*** *The density map for the pump which was created by Fitzpatrick et al; there are five different colours which are represented on the diagram showing the five different protomers with ToLC - blue, MacA – red, orange and yellow, MacB – purple.* ***B)*** *The arrangement in 7b is slightly different from 7A as it shows the sliced view showing presence of the gating ring. The gating ring supposedly functions as a one-way valve to substrate transport.* ***C)*** *This figure shows the presence of β- barrel and helical sheets; specifically, the interactions between MacA and the periplasmic domain of MacB is shown (Fitzpatrick et al., 2017).*

1.7.3 MacB

Evaluating both crystal and cryo-EM studies, MacB is the motor domain of the MacAB-TolC efflux pump which is responsible for the transportation of substrates such as antibiotics and enterotoxins STII out of the bacterial cell (Crow *et al*., 2017). In Crow *et al* study, it was found that MacB lacks a central cavity which allows the substrate to pass through, this is a similar feature to what is found in ABCA1 suggesting that there may be a link to ABCA1 evolving from the same or similar family. As it was suggested in Ford *et al* review, ABCA1

and ABCG2, obtaining the detailed structural and sequence alignments of the TMD and NBD of ABCG2 and ABCA1 to MacB may potentially show an evolutionary link (Ford *et al*., 2019).

MacB contains four transmembrane helices as seen in figure 8 with the NBD being the N terminal. There is the extensive periplasmic domain which is located between transmembrane 1 (TM1) and transmembrane 2 (TM2); it is specifically seen the periplasmic domain is elevated above the cytoplasmic membrane by a four-helix bundle; this periplasmic domain is the extensions of TM1 and TM2 where both these monomers are seen to interact along the dimer's symmetry axis (Crow *et al*., 2017). In comparison to TM1 and TM2, transmembrane 3 (TM3) and transmembrane 4 (TM4) are shorter and not involved in dimerization outside the four-helix bundle. Another observation that can be made between TM3 and TM4 is the presence of an 8-residue loop on the periplasmic side which essentially connects TM3 and TM4 (Crow *et al*., 2017). On the cytoplasmic side of TM1 there is an 18 amino acid amphipathic which is connected to the NBD by a 41-residue skirting loop. The 41-residue skirting loop is highly flexible, however it has been found to lack a secondary structure, making it hard to determine its role. Understanding the role of each fold, loops and helices gives a significant understanding of each topology and an insight into the involvement into the pathway in which substrates, especially drugs, are transported across the transmembrane. The detailed structural alignment of the MacB to ABCG and ABCA family may reveal how mechanism used by MacB may be more suitable for ABC proteins (Ford *et al*., 2019).

The analysis of the transport of substrates through MacB would be to start by looking at the general ABC transport of substrate through the transmembrane. The general pathway of transportation of the substrate through the transmembrane shows the presence of a vestibule of cavity which is located between the two halves of the dimer (Greene *et al*., 2018; Crow *et al*., 201). However, significant analysis has revealed that there is no transmembrane pathway for the substrate in MacB in either its ATP or nucleotide free state. Furthermore, through studies it was concluded that MacB does not transport substrates across the inner membrane, however, it gets in contact with the periplasm and becomes in contact with MacA, then the substrate is ejected through the outer membrane via the Tol-C exit duct (Greene et al., 2018). This type of mechanism that MacB has adapted is known as mechanotransmission mechanism. In this proposed type of mechanism when the substrate needs to be transported, as usual ATP binds to NBD and hydrolysis takes place, this causes a significant change in the

transmembrane in the MacB structure. As mentioned before, MacB does not transport substrates through the inner membrane, but has the ability to coordinate the reversible dimerization of the NBDs in the periplasmic conformational change (Greene *et al*., 2018). If there is a detailed structural alignment of ABCA and ABCG family to MacB, this can suggest that ABC proteins adopts more mechanotransmission mechanism rather than the alternating access mechanism.

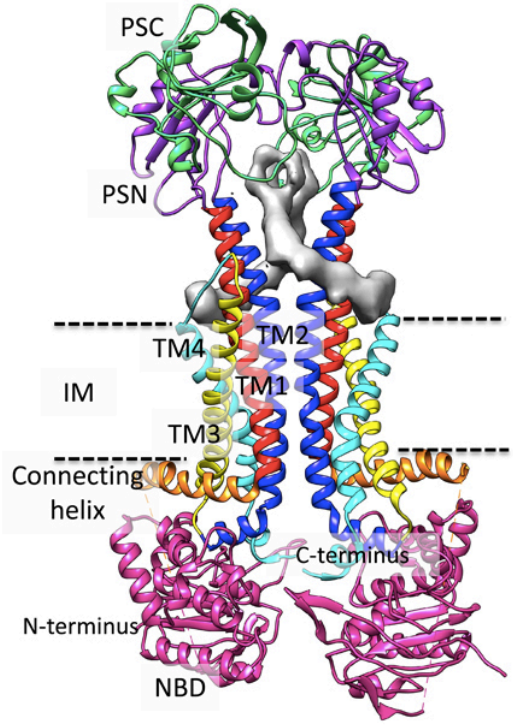


Figure 8: The structure of MacB

This figure represents MacB, hence the periplasmic domain (PLD) bears the N terminal and C terminal subdomains (PSN and PSC respectively). Furthermore, the figure shows MacB to be a homodimer; the TMD to be composed of 4 transmembrane helices (Fitzpatrick et al., 2017).PSN: periplasmic subdomain N; PSC: periplasmic subdomain C; IM: Inner membrane.

1.7.4 Lipopolysaccharide ABC transporter

The alternating access mechanism is the most widely associated mechanism to most ABC proteins to date. However, through structural and biochemical determination more bacterial ABC transporters have emerged to not obey this mechanism and are seen to adopt another. $LptB_2FG$ is bacterial ABC transporter which belongs to a group of gram-negative bacteria which comprises of Lipopolysaccharide (LPS). LPS is composed of lipid A, core oligosaccharide and O-antigen polysaccharide (Caffalette *et al*., 2019; Ford *et al*., 2019).
In Caffalette *et al* study, it was found that in LPS there is a lipid gating mechanism for the channel forming O-antigen ABC transporter. Within this lipid gating mechanism, the O-

antigen was transported by the inner membrane WzmWzt ABC transporter; whilst translocation took place large structural changes within the TMD and NBD was noticed, such that the hydrophobic residue was present at the entry site towards the periplasm (Caffalette *et al*., 2019). In Ford *et al* study the difference in mechanism in WzmWzt was noticed but no structural similarities were looked at in detail. Looking at the structure of WzmWzt, Wzm represents the TMD and is comprised of 6 transmembrane helices, hence Wzt represents the NBD as seen in figure 9. Comparing the structures of ABCA1, ABCG2 and ABCG5/G8 it can be postulated that the TMD and NBD may have similar topology and fold, if a significant structural alignment is found between ABCA1, ABCG2 and ABCG5/G8 to WzmWzt an evolutionary link may suggest convergence or divergence evolution is taking place.

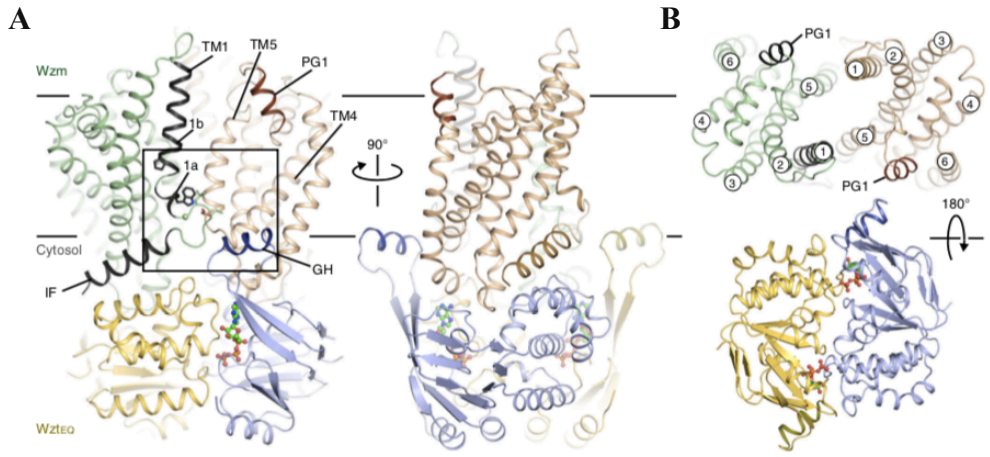


Figure 9: The structure of AaWzmWztEQ bound to ATP

A) *The ortholog view of AaWzmWztEQ with ATP bound. The Wzm protomer is represented in green ribbons, the N terminal interface helix (IF) and TM 1 helix is highlighted in black.* ***B)*** *The periplasmic and cytoplasmic view of AaWzmWztEQ, thus each Wzm protomer contains 6 transmembrane helices. The loop present between TM5 and TM6 is the periplasmic gate helix (PGI) (Caffalette et al., 2019)*

1.7.5 IM MlaFEDB

Understanding the mechanism of phospholipid translocation has been a major challenge faced when trying to understand bacterial phospholipid homeostasis. The IM MlaFEDB is a complex which functions as an ABC transporter to translocate phospholipid between the IM and periplasm (Chi *et al*., 2020). MlaFEDB is an example of another Gram-negative bacteria which does not obey the alternative access mechanism and instead has an extrusion mechanism as suggested by Chi *et al.* However, as this mechanism was determined from two low resolution structures of MlaFEDB, it still remains largely unclear to how translocation

takes place. The MlaFEDB structure is comprised of MlaD, MlaE, MlaF and MlaB as seen in figure 10. The structure is seen to be assembled in a stoichiometry of 2:2:6:6 (MlaF: MlaE: MlaD: MlaB), MlaE and MlaF are found to represent TMD and NBD respectively (Chi *et al*., 2020). Furthermore, MlaE and MlaF are homodimers which are associated with unique auxiliary proteins MlaB and MlaD (Chi *et al*., 2020). MlaE is found to contain 5 transmembrane, one extracellular helix and one periplasmic helix, and the MlaF is shown to contain the helical and Rec A like subdomain which makes up the NBD as seen in figure 11.

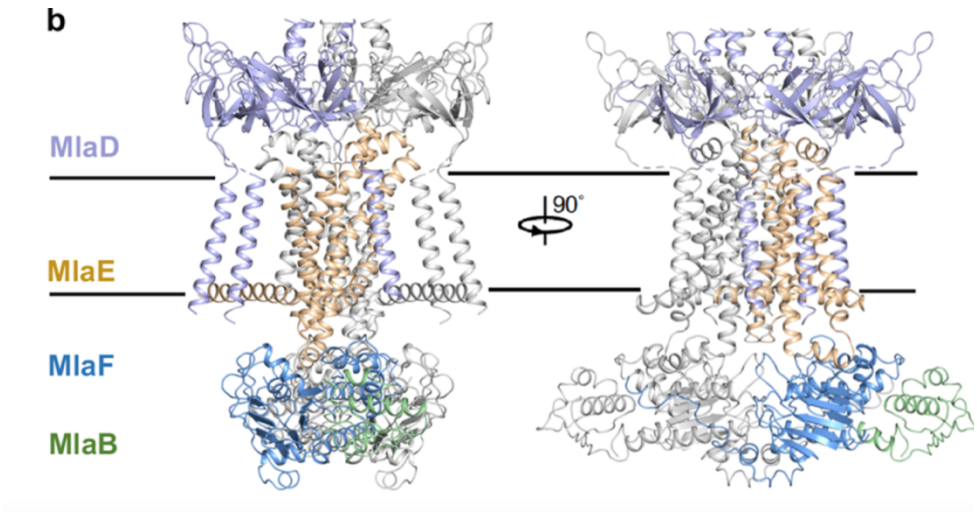


Figure 10: The structure of MlaFEDB complex

The overall structure of MlaFEDB complex, the following is represented in each colour; MlaD = light blue, MlaE = Wheat, MlaF= actinium and pale green (Chi et al., 2020)

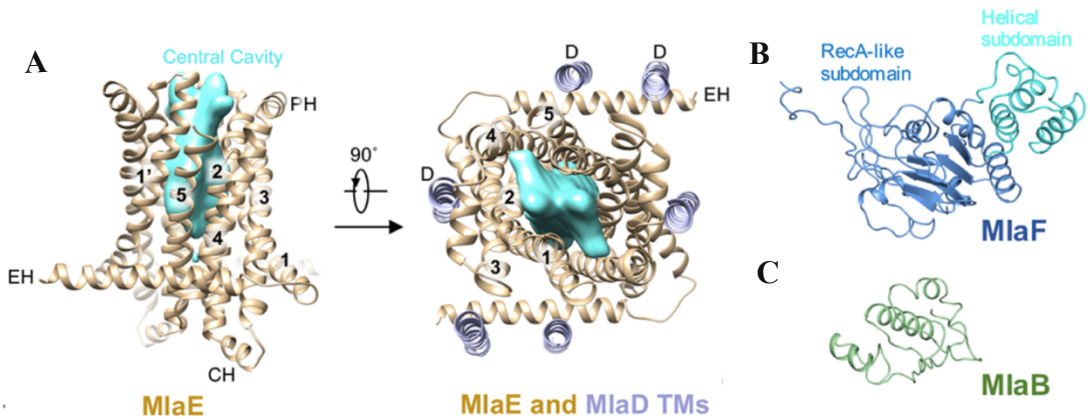


Figure 11: The separate structure of MlaE, MlaF and MlaB

A) *MlaE contains CH and 5 transmembrane helices as part of the TMD, and the TM view of MlaE and MlaD is seen when rotated 90°.* ***B)*** *Structure of MlaF composed of RecA like subdomain and helical subdomain* ***C)*** *Cartoon structure of MlaB (Chi et al., 2020).*

1.8 Evolution

The simplest word defining evolution is change. Understanding change is by finding evolutionary traits associated between ABCG and ABCA to bacterial ABC transporter. Naturally on earth, biological evolution is how species have originated and the process of evolution is a crucial and ongoing event in all living organisms (Sikosek *et al*., 2014). There are many factors which are associated with the mechanism of evolution such as natural selection, biased mutations, genetic drift and gene flow (Ashraf and Sarfraz., 2016). The main focus in proteins is looking at patterns of mutations in order to understand the cause for evolution. Mutations are generally neutral or harmful, however sometimes it can be beneficial if the environment is under state change. There are several types of mutations that take place such as: point mutation, frame shift, chromosome, deletion, duplication or insertion, inversion and translocation mutation. All of these types of mutation lead to adaptations, the adaptions have traits which helps in the survival and reproduction of the population.

Mutations in cancer cells result in a major limitation in the treatment of cancer. The main reason for acquired drug resistance to arise is changes in the host environmental factors alongside genetic or epigenetic alterations in the cancer cell due to excess or long-term exposure to the drug (Foo and Michor, 2014). The theory of evolution has been a major factor in understanding the dynamics of resistance mutation in a cancer cell population, the estimation risk of developing pre-existing resistance before the initiation therapy and calculating the optimum drug administration in order to reduce the risks of acquired drug resistance (Foo and Michor, 2014).

In the events of structural evolution of the eukaryotic ATP binding cassette transporter superfamily, Xiong *et al* carried -out research to investigate the structural evolution of the eukaryotic ABC transporter family. It was found that the ABCA & ABCG family have the highest sequence similarity of 22.4% (Xiong et al., 2015). In addition to the conservation of the Q loop, Walker B motif and the various amino acid sites of the NBD domain, the close relationship between the NBDs suggest a common ancestor for the NBD. Furthermore, in Xiong et al study they propose a theory of heterofusion and homofusion of the different half transporters families giving rise to full transporter families.

1.9 Membrane proteins and expression systems

Expressing, purifying and characterising of integral proteins starts by understanding its role in the biological system, then understanding its role in a certain disease. These embedded integral membrane proteins make approximately a quarter of the human genome and are responsible for various extremely complex physiological roles including transport, signal transducing, cell adhesion and response to physical and chemical stimuli (Danielczak *et al*., 2020). Therefore, any mutations or misfolding in the structure of these integral membrane proteins may result in a disease, thus making membrane proteins pivotal and attractive for therapeutic target. Approximately 50% of pharmaceutical drugs are designed to target integral membrane proteins such as G-protein coupled receptor (GCRPs), ion channels, solute carrier transporters and ABC transporters (Kwan *et al*., 2019; Frick *et al* 2019). Although many therapeutic drugs are designed to be targeted towards integral membrane proteins, most of the time there is significant reduced efficacy and side effects. This relates to the poor understanding of biochemical behaviour of the protein target and the protein interaction with the specific drug (drug mechanism of action). Henceforth, the understanding of structure and function relationship of membrane proteins is an absolute necessity and essential requirement in drug discovery (Lengger *et a*l., 2020). Understanding the structure - function relationship is achieved through structural characterisation; however, the study of membrane proteins encounters numerous technical difficulties during each step of crystallisation due to their natural embedding in hydrophobic environment made by lipids (Galkin *et al*.,2018; Kwan *et al*., 2019).

There are a few steps required in the process of structural determination of membrane proteins: expression, solubilisation and purification. Expression of the recombinant protein is carried out in heterologous hosts; thus an expression system is used to optimise and produce the membrane proteins of either bacteria, yeast based, mammalian or insect.

- **Bacterial protein expression system- *E.coli***

  Bacteria are very useful systems as they are rapid and simple in expressing their recombinant protein. Furthermore, creating the media to grow the cell culture is both cheap and scale-up bioproduction are straightforward as well (Rosano *et al*., 2014). Currently the most widely used host system is *E.Coli* as there is extensive research and knowledge upon the genetics, genome sequence and physiology. The advantages of using *E.Coli* are that genetic manipulation is relatively easy and it can grow in high

densities for large scale fermentations (Almenara *et al*., 2020; Magrone *et al*., 2020). The disadvantage of using *E. Coli*, is that the cell wall contains toxic pyrogens resulting in the expressed proteins to be extensively tested before being used in assays or any type of experiments (Rosano *et al*., 2014; Goh *et al*., 2020).

- **Yeast protein expression system – *Pichia pastoris***
  In general yeast systems are highly developed genetic systems; using yeast as an expression system to produce large quantities of recombinant protein has the benefits of being easy to use and has reduced time input. Furthermore, yeast can carry specifically designed plasmids which is highly valuable in recombinant expression system (Ahmed *et al*, 2014; Krainer *et al.*, 2012). *Pichia pastoris* and *Saccharomyces cerevisiae* are both very attractive organisms which are used to produce large quantities of recombinant protein.

- **Insect cell expression system – Sf9 and Sf21**
  Sf9 and Sf21 are the two types of insect cells which are used to produce recombinant proteins, these cells are derived from *Spodopetra frugiperda*. Both Sf9 and Sf21 are frequently used in recombinant expression system, with cells used from the lepidoptera family – baculovirus (lytic) and dsDNA (virus) are routinely amplified in cells belonging from these families (Geisler *et al*., 2017)

- **Mammalian cell expression system**
  Although using mammalian cell expression system is ideal, the main challenges facing expression of recombinant proteins is the reduced efficiency and the yield of proteins which are expressed is relatively insufficient for purification. However, over the past year HEK293 and CHO have been designed better to become efficient transient and stable expression systems (Hayat Khan *et al*., 2013)

1.9.1 Detergents and solubilisation

Detergents are key in the process of cell lysis and protein extraction. Detergents are both useful and unique as they have properties which enables them to disrupt or form hydrophobic interactions, and hence manipulate hydrophobic and hydrophilic interactions between biological molecules (Morrison *et al.*, 2016; Rajesh *et al.,* 2016). Understanding the chemical properties of detergents are useful, such that detergents are amphipathic molecules;

amphipathic molecules have both a nonpolar 'tail' and a polar head. The nonpolar tail features include aliphatic and aromatic residues whereas the polar head can be ionic, non-ionic or zwitterionic, hence the detergents are classed accordingly (Duseel *et al*., 2007; Robajac *et a*l., 2017) A crucial property of the detergents is the ability to form a micelle in aqueous environment due to the detergents having hydrophilic (head region facing outwards) and hydrophobic (tail facing the interior) properties (Arnold *et al*., 2008; Cho *et al*., 2016). The hydrophobic and hydrophilic properties help with the solubilisation of the targeted protein, the detergent helps by disintegrating the lipid bilayer while incorporating lipids and proteins in the detergent micelle. Although detergents are structurally diverse an important step before solubilisation is determining the critical micelle concentration (CMC) (Orwick – Rydmark *et al*., 2016; Swainsbury *et al*., 2017). Each detergent has its own properties such as pH, temperature, ionic strength and CMC; there is an associated certain concentration in aqueous environment as detergent molecules have the tendency of forming multimolecular micelle complexes (Arnold *et al*, 2008; Duseel *et al*, 2007; Orwick – Rydmark *et al*, 2016). Choosing the correct detergent and concentration (between 0.1-5% (v/v)) is important as the main aim is to dissociate the targeted protein, stabilise and preserve the protein's activity.

ABC transporters are known to be distributed throughout the biological cell membrane where they are integrated into the bilayer of the membrane, furthermore the presences of membrane lipids are known to affect the function of the ABC transporter (Neumann *et al*., 2016). In many studies the importance of the ABC transporter- lipid interaction is seen in biochemical assays such as the ATPase activity (Tarling *et al*., 2013; Aguiar *et al*., 2019). It has been evaluated, when ABC transporters have undergone solubilisation, the type of detergent and the concentration is extremely important as too harsh solubilisation condition may result in removal of important lipids. Thus, the function of the ABC transporter is known to be influenced by membrane lipids however it has been hard to distinguish its particular role and involvement in the activity (Neumann *et al*., 2017). Some of the question that are still raised and investigated by scientist today are: do lipids act as a structural stabiliser in ABC transporters? If so, does this imply that lipids have an integral role part of ABC transporters? If not, does the absence of the lipid causes the ABC transporter to be less stable? (Neumann *et al*., 2017; Tarling et al., 2013; Woebking *et al*., 2005).

1.9.2 Thermostability CETSA assay

Cellular thermal shift assays (CETSA) are based on the principle of how thermal shift assays work but on a cellular basis (Friman *et al*., 2020). Thermal shift assays are important in the study of stabilisation of proteins upon ligand binding. Many of the thermal shift assays have been used in drug discovery industry and academia to study purified proteins and drug interactions (Martinez *et al.,*2018; Guo *et al*., 2020). However, with CETSA it gives an insight into target engagement of drug candidates in a cellular context. Furthermore, when the cells are treated with drug of interest, the drug directly binds to the protein, thus the stabilisation between the protein and drug mainly relies on the principle of thermodynamics stabilisation (Lin *et al*., 2019; Jafari *et al,* 2014). This principle will give an insight into the estimation of binding free energies, hence other important thermodynamic properties in an isolated system of equilibrium.

CETSA assays works as follows:

- An initial baseline at which the cells denature at a certain temperature
- The cells are then treated with the drug at a certain concentration
- Heat is applied to denature the cell and precipitation of the protein takes place
- Cell lysis takes place
- The separation of the cell debris and the aggregation from the soluble fraction. Any unbound protein denatures and precipitates at the elevated temperature, ligand present and bound to the protein will remain in the solution (Jafari *et al*., 2014).

A comparison between the melting curve (known as the temperature induced aggregation curve) of the untreated protein and the ligand (drug) treated protein allows for calculation of the shift in thermal stability of the protein (ΔTm) (Martinez *et al*, 2018; Jafari *et al,* 2014). A positive ΔTm indicates protein stabilisation and ligand binding, using varying concentrations of a ligand the ΔTm can be used to calculate the binding affinity of ligand to the protein.

2.0 Bioinformatic tools and analysis

Bioinformatic analysis of protein structure and sequence is important for determining the function of that protein, hence looking at the structural evolutionary relationship between different proteins. Chimera is an extensive molecular modelling bioinformatic tool, it specifically helps in looking at the structural alignment between proteins hence conclude the

validity and similarity between the structures. The main parameters used in chimera to evaluate the structural alignment of the proteins are: root mean square deviation (RMSD), structural distance measure (SDM) and quality (Q) score.

- RMSD: The rmsd command evaluates essentially the root-mean-square-deviation between specified set of atoms; the data of the structural alignment interpreted as the smaller the RMSD value the better the alignment hence indicating high level of similarity in structure (Aier *et al*., 2016).
- SDM: The structural distance measure is a quantitative way of measuring the similarity of structure, with a value of 0 indicating identical structure and the value increases as the similarity between structures decreases. The SM value was found linearly related to the sequence derived distances and to yield of phylogenetic tree (Johnson *et al*., 1990; Krissinel *et al*., 2004)
- Q score: The quality score is generated pairwise with the SDM score, evaluating the quality in similarity when structures are superimposed. Unlike SDM score, when the Q score is 0 this indicates dissimilarity or unsuperimposed structures, thus 1 indicating identical structures superimposed (Krissinel *et al*., 2004).

2.0.1 Project aim:

The aim of this project is to:

- Express and purify ABCG2 using *Pichia pastoris*
- Conduct CETSA assay on ABCG2 and establish the proof of principle of whether this assay could be used to screen drugs and inhibitors in the future.
- Conduct Bioinformatic analysis to establish the structural evolutionary relationship between:
    1) ABCA and ABCG families
    2) ABCA and ABCG families to mechanotransducer MacB
    3) WzmWzt to ABCA and ABCG families
    4) MlaE and MlaF to ABCA and ABCG families

Chapter 2: Materials and Methods

2.1 Materials

The buffers were made using Mili-Q (MQ) water, sterilized through a 0.22 μm filter and stored at 4°C. All media were prepared using MQ water and autoclaved prior to use. Furthermore, the following can be found in the supplementary information: the overall reagents (appendix 1), protease inhibitors (appendix 2), buffers (appendix 3) and equipment (appendix 6).

2.1.1 Yeast Expression system – *Pichia pastoris*

The *Pichia pastoris* used in this project overexpresses ABCG2. Furthermore, *Pichia pastoris* strain SMD1163, specifically designed to have two mutations making it more protease deficient with the expression vector being pPICZA. DNA coding the full-length wild type human ABCG2 and the Kozak sequence were obtained from ThermoFisher Scientific. The FLAG, GFP and a Histidine tag consisting of 10 tandem histidines were introduced in order to facilitate future protein purification. The ABCG2 UniProt (Q9(UNQ0)) DNA sequence was used as the full BCRP DNA sequence. Thus, the plasmid map for GFP/His/FLAG-ABCG2 expressed using *Pichia pastoris* is yet to be published.

2.1.2 Yeast growth media:

- LB agar:
    - 35g/L LB Broth Lennox
    - 20g/L Agar
    - 1ml of 50μg/ml Zeocin
- BMGY Medium for 1L:
    - 700ml of YET (10g of yeast extract, 20g of Tryptone and MQ water)
    - 100ml of 13.4 % (w/v) yeast nitrogen base without amino acid
    - 100ml of 10% (v/v) glycerol
    - 1M of 100ml phosphate buffer at pH 6.0 (13.6g of KH_2PO_4: dissolved in 100ml 5.226g of K_2HPO_4 dissolved in 30 ml)
    - 1ml of 50μg/ml of zeocin added once BMGY medium had cooled
- BMMY medium for 1L: protocol for BMGY medium followed, however 100ml of 5% methanol used in place of glycerol.

2.2 Experimental methodology

2.2.1 Protein expression of *Pichia pastoris*

2.2.1.1 Inoculation of colonies of *Pichia pastoris* on Agar plate

- LB Broth Lennox and Agar reagents are weighed out and dissolved in 900ml of MQ water, the solution was then made up to 1L with MQ water. The medium was autoclaved for 2-3 hrs, once cooled (but not solidified) zeocin was added.

- Zeocin solution equivalent to the dilution of 50ug/ml was made up; 1ml of 50ug/ml zeocin was added to LB Broth Lennox and Agar solutions upon cooling. The Zeocin was stored in the fridge at 4 °C.

- Ethanol was used to disinfect the surface before *Pichia pastoris* cells were plated on the petri dish to avoid any cross contamination of the plates. The petri dishes were placed near a Bunsen burner before the agar solutions were plated to reduce contamination; the agar plates solidified after 30 minutes. Cells were streaked on the agar plates and left for 10 minutes before sealing and storing in an incubator at 30°C. After 1-2 days the colonies would start to grow, and the plates streaked again for larger growths and incubated for a total of 3 days.

2.2.1.2 Growth of *Pichia pastoris* cells

- Whilst the *Pichia pastoris* cells were left to grow over a 3-day period, the BMGY medium was prepared. Yeast extract, Tryptone, Nitrogen Base without amino acid and glycerol were added to 900 ml of MQ water. The medium was autoclaved, and Phosphate buffer was added to maintain the pH.

- As expression and purification of membrane proteins are difficult, a large-scale cell growth was required; in this project 5L of *Pichia pastoris* cells were grown.

- The *Pichia pastoris* cells grown on the petri dishes were inoculated into a starter culture of 100ml of BMGY medium and left in the orbital shaking incubator at 225rpm at 30°C for 24hrs. The starting culture OD600 = 4.38 (YET was used as a blank to get an OD600 = 0.00, then a 1/10 dilution of starting culture and blank were

used to measure the OD600 of the starting culture as the spectrometer does not measure above OD of 2.0).

- 5L of BMGY media was prepared, autoclaved and phosphate buffer.
- 20ml of starting culture was added to each 1L flask. Each flask was placed in the orbital shaking incubator for 24hrs. After 24hrs the OD600 = 5.63, the 1st methanol induction (8% of 100ml) was made.
- OD600 = 6.54 after the 1st induction of methanol, after 24hrs the 2nd methanol induction (8% of 100ml) was made.
- OD600 = 7.14 after the 2nd induction of methanol, after 24hrs the 3rd and final methanol induction (8% of 100ml) was made.
- Once the 3rd methanol induction was made the culture appeared cloudy with a fruity odour indicating the growth of *Pichia pastoris*. A variation in the fruity odour could indicate a sign of potential culture contamination. Finally, a reading of OD600 = 7.14 corresponded to approximately a total of 3.0 x 10^{11} cells.
- Flasks containing 1L of *Pichia pastoris* culture were weighed and transferred into centrifuge containers. Cells were harvested using the centrifuge, where each container was centrifuged for 10min at 3000xg, 4°C.
- Upon centrifugation the supernatant was discarded, and the pellet was weighed; a total of 105.21g of cells were weighed from 5L of *Pichia pastoris* culture.
- Each Pellet was suspended in 50ml of mPiB before being snap frozen using liquid nitrogen and stored in -80°C. Snap freezing is an important process before storing any protein in the freezer, it reduces the presence of ice crystals in the sample whilst the freezing process takes place. This slows the action of proteases, therefore preventing degradation of membrane proteins and other molecules. In this instance, mPiB buffer was specifically used due to its high protease inhibitor content, thus slowing the protease degradation process.

2.2.2 Cell Breakage, Microsome preparation and Microsome solubilisation

Cell breakage of *Pichia pastoris* cells took place using a BeadBeater. This involved the use of glass beads with alternating shaking and cooling phases; the cooling phase was created by the use of dry ice. It was crucial that cell breakage took place in the cold room at 4 °C.

- *Pichia pastoris* cells which were snap frozen and stored in mPiB at -80 °C were thawed at room temperature. The cells were placed in 50 ml falcon tubes and centrifuged. The pellet was collected and resuspended in fresh 100ml of mPiB buffer containing DTT and protease inhibitor.

- The *Pichia pastoris* cells were resuspended (into a barrel specifically designed for the BeadBeater machine) with 250g of dry glass beads with an approximate diameter of 426-600um.

- Cell breakage involved cycles of 1 min shaking and 2 min ice cooling phase which involved taking the barrel and placing it in a box of dry ice (leaving the samples in dry ice for more than 2 minutes risks the samples freezing up really quickly).

- To confirm cell breakage, OD600 of the sample was taken at various time intervals. Firstly, the sample was centrifuged for 5 mins at 12000rpm at 4°C:
    - Before starting breakage OD_{600} = 0.006
    - Cycle 11 OD_{600} = 0.019
    - Cycle 15 OD_{600} = 0.019
- The OD did not change between cycle 11 and 15, therefore 11 cycles were adequate for complete cell breakage. The sample was decanted from the beads in the barrel, the bead and barrel were washed with ethanol and water several times after use.

- The sample was centrifuged at 11000rpm for 10 mins at 4°C using Beckman Avanti J26XP with JLA25.50 rotor

- Once the sample had been centrifuged, the supernatant contained the membranes expressing ABCG2, and the pellet contained unbroken cells, nuclei and other organelles which were discarded.

- The supernatant was centrifuged for a further 90mins at 40000 rpm (200000 X g) in the Ti45 rotor at 4°C. Once complete, the supernatant was discarded, and the pellet was collected and resuspended in 50ml of High salt Buffer (a paint brush was used to aid the mixing).
- Samples were then centrifuged at 30000 rpm (100000 x g) for 60 mins at 4°C. The supernatant was discarded, and the pellet was resuspended in 10ml of Buffer A.

- Next, the Bradford assay was used to determine the protein concentration of the microsomes:

    - The microsomes were diluted to 1/10 with buffer A to achieve a total volume of 20 ul.
    - Prior to measuring OD_{595} of the microsomes, a blank was measured in order to reduce the background. This consisted of 800ul of MQ water and 200ul of Bradford reagent.

    - To measure OD_{595} of the diluted microsomes, 10ul of the diluted microsomes were added to the blank solution. The OD_{600} measured at 0.198

    - The total protein concentration was determined via the following equation:
        - Protein concentration (mg/ml) = OD_{595} X 15 = 0.198 X 15 = 2.97mg/ml

Once the protein concentration was calculated, the Fluorimeter was used in order to determine the GFP signal. YET was placed in the cuvette and was used as a blank to reduce any background noise, N_2 was used to dry the cuvette before placing the microsomes for detection of GFP signal. A peak was present at 512nm which is associated with the GFP tag, thus positive expression of ABCG2. The microsomes were snap frozen and stored in an -80°C freezer for solubilisation.

2.2.3 Solubilisation

- The best solubilisation took place at 2% and 4% DDM, thus solubilisation at 2% of DDM was shown to be the most effective. The following protocol was used for solubilisation of the microsomes:

    - 1ml of 20% DDM was added to the microsomes which were stored in 10ml of Buffer A, giving a final concentration of 2% DDM
    - Once the detergent was added to the microsomes, the falcon tube containing the sample was transferred to the shaker for 2hrs in the cold room (4°C); the sample was centrifuged after 2hrs on the shaker at 10000 x g at 4°C (On the Ti45 rotor it is 30000rpm).
    - The supernatant collected after centrifugation contained the soluble proteins and the pellet contained the insoluble fraction.
    - To determine whether solubilisation has been effective the soluble and insoluble fractions were loaded on the 10% SDS PAGE, however the bands on the gel were very faint and therefore unclear. Despite this, purification was continued in order to determine whether significant pure protein was present.

2.2.4 SDS–PAGE

- Sodium Dodecyl Sulphate Polyacrylamide gel electrophoresis is a method used to separate proteins based on their molecular weight, the thickness of the gels can vary between 0.75-1.5mm; the gel thickness used in this project was 1.5mm. The gels prepared in this project were either 8% or 10% SDS–PAGE following the protocol provided by the lab, with Tris-Glycine-SDS (TGS) as the running buffer. The following protocol explains how to run an SDS-PAGE:
    - Samples were first mixed with an equal volume of 2 X loading dye and left on the bench for 1hr; the loading dye contains SDS and DTT, both these reagents are prime components required for the migration of the protein:

        - DTT is a reagent which prevents oxidation of the thiol group and reduces formation of intramolecular and intermolecular disulphide

bonds which form between the cysteine residues present on the protein. This helps with the migration of the protein along the gel.

- SDS essentially denatures the protein by binding to the hydrocarbon chain, the amount of SDS bound to the protein is proportional to the length of the hydrocarbon chain

- After 1hr, the samples were spun in a microcentrifuge for 2 mins in order to obtain a homogenous solution. The samples were then loaded into the wells of the polyacrylamide gel.

- Electrophoresis was carried out at 120 V for approximately 2 hrs, the length of time required for the completion of the electrophoresis varies on the equipment used, size of protein to be separated and the thickness of the gels. Once the dye disappears, this indicates that the electrophoresis has finished.

- The gels are scanned using Biorad - Chemidoc equipment under Alexa488, which detects the presence of GFP fluorescence signals. Furthermore, in the toolbar a new protocol had been pre-set for the multichannel gel scan, where channel 1 is represented in red named as colorimetric (excitation wavelength: 650nm, emission wavelength: 670nm); channel 2 is green represented as Alexa 488 (excitation wavelength: 485nm, emission wavelength: 530nm) determining the GFP fluorescence, channel 3 represented in blue (excitation wavelength:440nm, emission wavelength:500nm). Once the SDS-PAGE had been scanned the multichannel presented bands of particular molecular weight, hence each corresponding to green, blue and red.

2.2.5 Determining the ABCG2 concentration using the fluorimeter

To calculate the concentration of the ABCG2 microsomes fluorescence spectroscopy was used; a fluorescence intensity at 512nm on the fluorimeter indicated the presence of a GFP signal, the concentration was later determined using the following equation: $y = 0.5 * x$

y = ABCG2 or other membrane protein concentration (ug/ml)

x = the intensity of the GFP fluorescence at 512nm (arbitrary units)

Before measuring the concentration of the sample, the quartz cuvette was washed and blown with N_2 gas in order to freeze any remaining water droplets, this was to ensure the measurement of the intensity of the sample was not disturbed. The fluorimeter was set up with an excitation wavelength at 485nm and emission range of 500-600nm.

2.2.6 Protein Purification of ABCG2

2.3.6.1 Purification using Ni NTA agarose beads

Initial purification of ABCG2 began with the use of Ni NTA agarose beads via a batch method.

- As Ni NTA agarose beads are prone to solidification having been stored at -4°C, their container is shaken lightly and left on the bench for approximately 15 minutes for de-solidification.
- 2ml of Ni NTA and 2ml of buffer A (BA) was mixed into the falcon tube and left on the bench for 1hr
- 4ml of Ni NTA suspended in buffer A were added to soluble proteins expressing ABCG2 and left on the roller in the cold room (4°C) for 24hrs, this allows the Ni NTA to bind to the His tag which is present on the sequence. During purification, however, the tag can be cleaved off which is why this particular method of purification may not always have a high success rate.
- The solution is transferred to an Eppendorf and centrifuged at 3000 x g 4°C for 10 minutes.
- Once completed, the supernatant (representing the unbound fraction) and the pellet (containing Ni NTA-bound membrane proteins) were collected. The supernatant was collected as the wash fraction and the pellet was suspended in the first elution buffer. The first elution buffer consisted of:
    - 8mM Imidazole
    - 0.1% DDM
- After suspension of the soluble proteins in the first elution buffer, the falcon tube was left on the roller for 1hr in the cold room.

- The sample was later loaded onto an 10% SDS PAGE in order to determine the presence of a molecular band of 90kDa (which corresponds to GFP/His/FLAG-ABCG2)
- After 1hr in the cold room on the roller, the sample was transferred to an Eppendorf and centrifuged at 3000 x g at 4°C. The supernatant was kept as the second wash fraction and the pellet was resuspended into the second elution buffer.

  The second elution buffer consisted of:

  - 80mM Imidazole
  - 0.1% DDM
- After suspension of the soluble proteins in the second elution buffer, the falcon tube was left on the roller for 1hr in the cold room.
- The sample was later loaded onto an 10% SDS PAGE in order to determine the presence of a molecular band of 90kDa (which corresponds to GFP/His/FLAG-ABCG2)
- After 1hr in the cold room on the roller, the sample was transferred to an Eppendorf and centrifuged at 3000 x g at 4°C. The supernatant was kept as the third wash fraction and the pellet was resuspended into the third elution buffer.

  The third elution buffer consisted of:

  - 100mM imidazole
  - 0.1%DDM
- After suspension of the soluble proteins in the third elution buffer, the falcon tube was left on the roller for 1hr in the cold room.
- The sample was later loaded onto an 10% SDS PAGE in order to determine the presence of a molecular band of 90kDa (which corresponds to GFP/His/FlAG- ABCG2)
- After 1hr in the cold room on the roller, the sample was transferred to an Eppendorf and centrifuged at 3000 x g at 4°C. The supernatant was kept as the fourth wash fraction and the pellet was resuspended into the fourth elution buffer.

  The Fourth elution buffer consisted of:

  - 400mM Imidazole
  - 0.1% DDM
- The sample was later loaded onto an 10% SDS PAGE in order to determine the presence of a molecular band of 90kDa (which corresponds to GFP/His/FlAG- ABCG2)

- After many attempts of purification using Ni NTA resins, no molecular bands were observed at 90kDa on the SDS–PAGE. This result was not dissimilar to previous attempts of purification performed by a former student who found cleavage of the His tag. A resolution to this involved the use of a GFP trap resin for purification in place of NI NTA resins.

2.2.6.2 Purification using GFP trap resin

Purification of ABCG2 in this project was also carried out using the batch method and a GFP trap resin.

- As GFP trap resin is stored at 4°C, the resins were rehydrated using buffer A and left on the bench at room temperature for an hour.
- 50µl of the resin was mixed with 5ml of solubilised protein supplemented with 100ul of protease inhibitor and the left on the shaker at 4°C overnight. This allows the protein to bind to the resin and consequently a more efficient purification.
- The sample was spun down by centrifugation 2500 x g at 4°C for 2 minutes. The supernatant collected was kept as the unbound fraction.
- The pelleted resin present in the falcon tube after the suspension of the supernatant is washed with buffer AA and spun down by centrifugation at 2500 x g at 4°C for 2 minutes. The supernatant collected was kept as the wash fraction.
- The sample resin was then spun down by centrifugation at 2500 x g at 4°C for 2 mins and the supernatant collected as the pure fraction.
- The pure, unbound and washed fractions were loaded onto an 8% SDS PAGE. If purified proteins were present then a molecular band at 90kDa would be observed, however, this was not the case. After many attempts and due to time constraints, further work could not be carried out in order to obtain purified ABCG2 proteins. ABCG2 microsomes, however, were used in my cellular thermal shift assay as part of my functional studies.

2.2.7 Cellular thermal shift assay (CETSA)

- The concentration of microsomes required for the CETSA assay was 5ug/20ml. Therefore 52.08ul of ABCG2 microsomes were added to 15.7ul of assay buffer and 0.628ul Mg-ATP and incubated in a PCR tube for 30 minutes at 37 °C

- After 30 minutes, 3-minute incubation was carried out at various temperatures (37°C, 40°C, 43°C, 46°C, 49°C, 52°C, 55°C, 58°C, 61°C)
- The sample was treated with ice-cold PBS (made up of 0.8% DDM)
- The sample was snap frozen in liquid nitrogen and placed in a thermometer at 25°C to initiate thawing before being transferred to ice until entire content were thawed
- Ultracentrifugation was carried out in order to precipitate the denatured protein at 100,000 x g for 20 minutes at 4 °C
- The soluble fraction was subjected to 10% SDS PAGE and Biorad Alexa 488 was used in order to determine the presence of GFP fluorescence signals, which ultimately indicated the presence of ABCG2.

2.3 Bioinformatic methodology

The bioinformatic tools used in this project are Chimera, Clustal omega and Uniprot. Chimera is a programme designed to interactively visualise molecular structures and related data (Pettersen *et al*., 2004; Sievers *et al*., 2011; UniProt: a worldwide hub of protein knowledge,2018; Morgat *et al*., 2020). This programme can provide an insight into the structural and sequence alignment of proteins.

2.3.1 The structural relationship between ABCA and ABCG families

- Chimera was the programme used to achieve the structural alignment and the structural sequence alignment in this project (Pettersen *et al*., 2004).
- In order to establish the structural relationship between ABCA and ABCG families, ABCA1, ABCG2 and ABCG5G8 were used in the structural alignment.
- The PDB of each structure was obtained from the RCSB protein data bank (Bank,R., 2019):
    - ABCA1 : 5XJY (Qian *et al*., 2017)
    - ABCG2 : 5NJ3 (Taylor *et al*., 2017)
    - ABCG5G8 : 5DO7 (Lee *et al*., 2016)

- Only the TMD and NBD were aligned in all the structural alignment; using fetch in chimera the PDB structure ABCA1 and ABCG2 were inserted.
- Using action then colour the chains were coloured accordingly.
- Using tools then 2D depiction each chain was labelled.

- Structural comparisons were made using Tools then Structure Comparison, within that section Matchmaker was selected (Petterson *et al*., 2004). The reference structure (ABCA1) was matched to ABCG2, hence the distance between the atoms were chosen based on the whether the structural alignment was significant.
- After the structural superposition alignment was achieved using Matchmaker, Match and align was used to generate the structural superimposed sequence alignment. This programme gave the conservation and consensus of each alignment, alongside the overall RMSD, SDM and Q score.
- To obtain the Clustal Omega sequence, Tools was selected first then Sequence; within Sequence, Align Chain Sequence was chosen. The Clustal Omega sequence was generated, and the percentage identity of each structural alignment were determined (Petterson *et al*., 2004).

2.3.2 The structural evolutionary relationship between ABCA and ABCG families to Mechanotransducer MacB

- To establish the structural evolutionary relationship between ABCA and ABCG to MacB, three structures of MacB were aligned with:
    - 5WS4: Crystal structure of tripartite-type ABC transporter MacB from *Acinetobacter baumannii (*Okada *et al*., 2017).
    - 5LIL: Structure of *Aggregatibacter actinomycetemcomitans* MacB bound to ATPyS (P21) (Crow *et al*., 2017)
    - 5NIK: Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump (Fitzpatrick *et al*., 2017)
- Each MacB structure as mentioned above were structurally aligned with ABCA1, ABCG2 and ABCG5G8. The structural alignment, structural sequence alignment, overall RMSD, SDM, Q score and percentage identity were obtained in the method above (method used when establishing the structural relationship between ABCA and ABCG families).

2.3.2 The structural evolutionary relationship between ABCA and ABCG families to WzmWzt

- The structure used to find the evolutionary relationship between WzmWzt to ABCG and ABCA families was the ATP-bound conformation WzmWzt O antigen ABC transporter structure determined by Caffalette *et al.* The PDB for this structure was 6m96 (Caffalette *et al*., 2019).
- The WzmWzt structure as mentioned above was structurally aligned with ABCA1, ABCG2 and ABCG5G8. The structural alignment, structural sequence alignment, overall RMSD, SDM, Q score and percentage identity were obtained in the method above (method used when establishing the structural relationship between ABCA and ABCG families).

2.4.3 The structural evolutionary relationship between ABCA and ABCG families to MlaE and MlaF components in MlaFEDB complex

- Only MlaE and MlaF components of MlaFEDB complex were used in the structural alignment to ABCA and ABCG families, as they represent the TMD and NBD respectively. The structure used to find the evolutionary relationship between ABCA and ABCG families was the overall structure of the nucleotide free MlaFEDB complex published (Chi *et al*., 2020).
- The MlaE and MlaF structure was structurally aligned with ABCA1, ABCG2 and ABCG5G8. The structural alignment, structural sequence alignment, overall RMSD, SDM, Q score and percentage identity were obtained in the method above (method used when establishing the structural relationship between ABCA and ABCG families).

Chapter 3 Experimental Results

3.1 Expression of ABCG2 using *Pichia pastoris*

The ABCG2 protein had a GFP-tag attached to the N-terminus and was expressed using *Pichia pastoris* cells. Finding the correct conditions to obtain the optimal expression of ABCG2 using *Pichia pastoris* is crucial. The results recorded after the 3-day methanol induction revealed that 0.8% of methanol, induced at 24hrs interval, gave the highest OD_{600} reading of 7.14, as seen in table 3. Diversely, the 3-day methanol induction at 22, 23 and 25hrs gave lower readings as seen in tables 1, 2 and 4, respectively. The lowest reading recorded after the 3-day methanol induction was at 22hrs, with an OD_{600} reading of 5.09. This indicates that methanol utilising pathway was optimised when methanol was induced at 24hrs interval over 3 days.

Literature analysis have reported from many studies that membrane proteins are hard to express. Since ABCG2 is a membrane protein, 5L of *Pichia pastoris* culture was made to ensure good yield of protein. Each 1L culture pellet weighed approximately total amount of 30.00g.

***Pichia pastoris* culture 1**	
Time of induction (hrs)	**OD_{600}**
0	4.38
22	4.60
22	5.00
22	5.09

Table 1: Pichia pastoris culture prepared with methanol induced at 22hrs over 3 days.

The table represents the first Pichia pastoris culture expressing ABCG2, prepared with 22hr interval of 0.8% methanol induction over 3 days and its corresponding OD600 reading.

Pichia pastoris culture 2	
Time of induction (hrs)	**OD_{600}**
0	4.38
23	5.02
23	5.36
23	6.23

Table 2: Pichia pastoris culture prepared with 0.8% methanol induced at 23hrs over 3 days

The table represents the second Pichia pastoris culture expressing ABCG2 prepared with 23hr interval of 0.8% methanol induction over 3 days and its corresponding OD600 reading.

Pichia pastoris culture 3	
Time of induction (hrs)	**OD_{600}**
0	4.38
24	5.63
24	6.54
24	7.14

Table 3:Pichia pastoris culture prepared with 0.8% methanol induced at 24hrs over 3 days

The table represents the second Pichia pastoris culture expressing ABCG2 prepared with 24hr interval of 0.8% methanol induction over 3 days and its corresponding OD600 reading.

Pichia pastoris culture preparation 4	
Time of induction (hrs)	**OD_{600}**
0	4.38
25	4.63
25	5.04
25	5.14

Table 4: Pichia pastoris culture prepared with 0.8% methanol induced at 25hr over 3 days

The table represents the second Pichia pastoris culture expressing ABCG2 prepared with 25hr interval of 0.8% methanol induction over 3 days and its corresponding OD600 reading.

3.2 ABCG2 Expression

After the pellet was collected, the cells were broken and the microsomes were isolated and obtained. Before proceeding further to solubilisation and purification the expression was checked. As previously mentioned, the ABCG2 has a GFP-tag at its N-terminus, the fluorimeter was therefore used to detect the presence of the GFP. Most proteins are not fluorescent, especially recombinant proteins cultured from *Pichia pastoris,* as the green fluorescent protein has a unique structure which is not common among other proteins. Hence, this allows GFP to be useful as a tag in detecting the gene expression of the particular protein of interest. Furthermore, the useful property of GFP is the fluorescence; any florescent molecule absorbs light at a certain wavelength and emits light at a longer wavelength, for GFP it has an exciting wavelength of 485 nm and emission wavelength of 530 nm. The microsomes were placed in a cuvette and subjected to the fluorescent spectrophotometer, figure 12 shows a small peak at 512 nm with an intensity of 52.59 - indicating the presence of the GFP tag, hence the presence of ABCG2. The results from figure 12 were input into the equation y=0.5*x, the estimation of the ABCG2 microsome concentration was determined as 29.795 ug/ml, as shown in table 5.

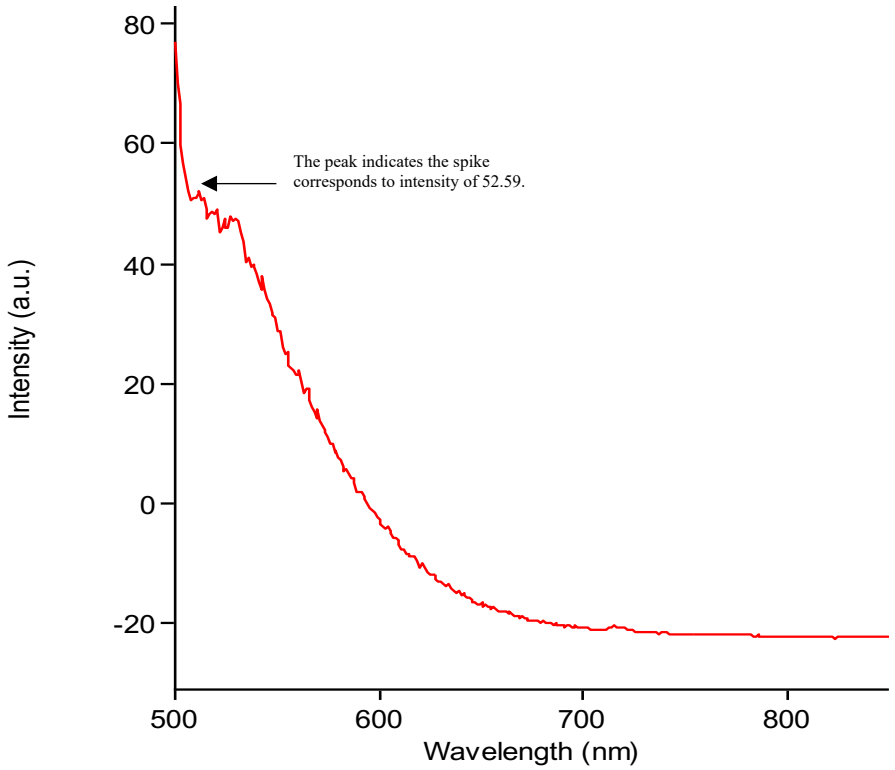


Figure 12: The fluorescent spectrophotometer graph indicating the presence of ABCG2-GFP

The fluorescent spectrophotometer generated this intensity against wavelength graph showing the presence of a small peak 512nm with an intensity of 52.59, thus around 512nm there are several peaks which can be seen hence they indicate the background noise.

Wavelength (nm)	Intensity (a.u)	ABCG2 concentration (ug/ml)
512.98	52.59	29.795

Table 5: ABCG2 concentration determined using the intensity wavelength graph

*This table represents the concentration of ABCG2 determined using the equation y=0.5*x, hence applying x as the intensity value determined at 512.98 by the fluorescence spectrophotometer.*

Although figure 12 indicates the presence of the N-terminal GFP-tag on ABCG2, the microsomes samples were subjected to a 10% SDS PAGE gel to further confirm the presence of ABCG2-GFP rather than just a GFP tag.

The gels were scanned using the Biorad - Chemidoc equipment and Alexa 488 was used to determine the presence of the fluorescence signal corresponding to the GFP. The GFP- Tag has a molecular weight of 27kDa and ABCG2 has a molecular weight of 72kDa, however through other studies such as Mao *et al* and Rosenberg *et al* it was found that when ABCG2 is expressed using *Pichia Pastoris* it is normally in the under-glycosylated form, which reduces the molecular weight by 10kDa. Therefore, the resultant molecular weight of the fused ABCG2-GFP tag expressed by *Pichia pastoris* is approximately around 90kDa.

In figure 13, there is a clear indication of a band intensity at 90kDa which indicates to ABCG2-GFP protein when scanned with Alexa488. In figure 14 the green fluorescence at 90kDa band intensity further confirms the presence of ABCG2-GFP, this concludes that the microsomes have a significant amount of ABCG2 expressed. There are also bands at 50kDa and approximately 30kDa present on figure 14, however within the microsomes other molecules transcribed in *Pichia pastoris* may be present. It is possible that the 50kDa band, represented in the blue fluorescence corresponds to the glycosylated secretomes (Burgard *et al*., 2019). The 30kDa band represented in red fluorescence corresponds to *Pichia pastoris* lipidome or proteomes present in the microsomes (Ivashov *et al*., 2013). Another way of determining positive expression of ABCG2-GFP is by doing an immunoblot using anti-ABCG2 or anti-GFP, however within the lab did not have any anti-ABCG2 or anti-GFP, thus these antibodies were not bought in as they were expensive. Therefore, in the future to further validate the detection for the expression of ABCG2-GFP, an immunoblot using anti-ABCG2 or anti-GFP should be used.

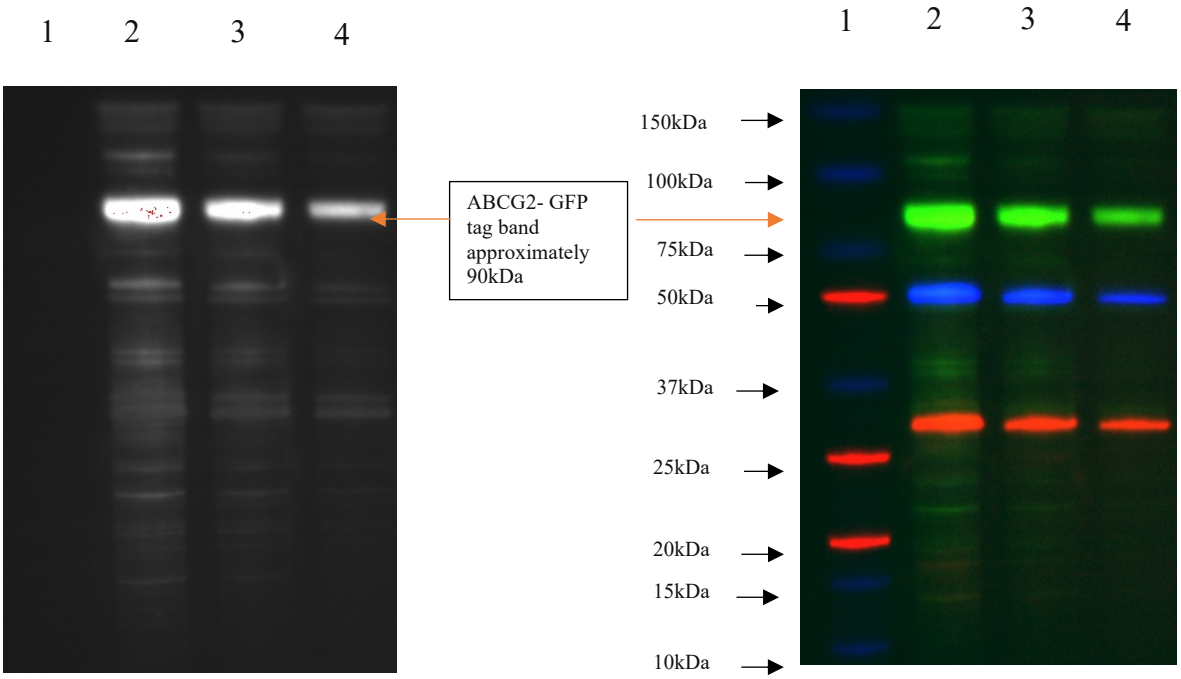


Figure 13: SDS-PAGE gel scanned using Alexa 488 showing the presence of ABCG2-GFP

Lane 1- protein biomarker. Lanes 2,3,4 - microsomes expressing ABCG2-GFP. Alexa 488 fluorescent band – (excitation wavelength:485nm, emission wavelength: 530nm)
This 10% SDS Page gel scanned using Alexa488 a strong band intensity present at 90kDa which corresponds to ABCG2-GFP

Figure 14: 10% SDS-PAGE gel scanned using multichannel showing the presence of ABCG2-GFP

Lane 1- protein biomarker. Lanes 2,3,4 - microsomes expressing ABCG2-GFP. Alexa 488 green, fluorescent band – (excitation wavelength:485nm, emission wavelength: 530nm). Blue fluorescent band – (excitation wavelength: 440nm, emission wavelength:500nm). Red fluorescent – (excitation wavelength: 650nm, emission wavelength: 670nm).
This 10% SDS PAGE gel scanned using the multichannel shows the presence of ABCG2-GFP at 90kDa, glycosylated secretomes at 50kDa and other lipidome or proteomes present at 30kDa.

3.2.1 Total protein concentration

The total protein concentration of the microsomes expressing ABCG2 were determined using the quick Bradford protein assay; 10 µL of 10-fold protein sample was diluted to 1ml. Initially, to measure the OD_{600} blank it was composed of 800 µL of MQ water which was mixed with 200 µL of the Bradford reagent, once the OD_{600} was measured the blank was set. 10 µL of the diluted samples were added to the blank solution, which gave $OD_{600} = 0.198$, thus the total protein concentration worked out as 2.97mg/ml, as shown in table 6.

Sample OD_{600}	Total Protein concentration (mg/ml) = OD_{600} x 15
0.198	2.97

Table 6: Total protein concentration of ABCG2

The table represents the OD600 from the sample expressing ABCG2, and the calculations required to find the corresponding protein concentration

3.3 Solubilisation and Purification

The microsomes expressing ABCG2 were solubilised with 2% DDM. The microsomes were stored in a total volume of 10 mL of buffer A; to obtain a final concentration of 2% DDM of the ABCG2 (microsomes) sample, 1 mL of 20% DDM was added to the ABCG2 (microsomes) sample. This sample was then left in the detergent on the shaker in the cold room (4°C) for 2hrs; after the completion of the 2hrs the sample was centrifuged at 30000 rpm at 4°C for 1hr. Both the supernatant and the insoluble fraction was collected as they were expected to contain the soluble and insoluble fraction, respectively. When both insoluble and soluble fractions were run on both 8% and 10% SDS PAGE no visible bands around 90kDa were present in the soluble fraction. Although no fractions were seen on the SDS PAGE gels from the process of solubilisation, the ABCG2 sample was subjected to purification.

In addition to GFP-tag, ABCG2 has a His-tag and a FLAG-tag. There were two attempts of purification made; the initial method was using Ni NTA agarose beads in a batch process and the second method was using GFP trap resin. Ni NTA agarose beads have high affinity to the his-tag present on ABCG2 therefore in theory the purification method should have worked, however after many attempts of purifying with Ni NTA there were no molecular bands found at 90kDa on the 8% or 10% SDS – PAGE, this may be as a result from the his-tag being cleaved during the purification process. After the second method of purification using GFP trap resin, there were still no molecular bands seen on the 8% or 10% SDS PAGE. Image J was used to scan the gels for very faint intensity bands which was not very visible to the naked eye.

The method of using the BSA calibration curve, as shown in figure 15, and the equation $y= 13988x + 12036$ *with* $R^2 = 0.9396$, the protein concentration of pure ABCG2 worked out to be 0.00143mg/ml. The BSA curve was generated by doing a Bradford protein assay, a blank

sample of buffer A was used to calibrate the spectrophotometer, however it gave a higher reading than expected. Therefore, in the future a different buffer could be used to calibrate the instrument to give all subsequent sample absorbance is reflected to a more accurate analyte absorbance. However, after many attempts it was determined that the concentration of the pure protein was significantly small, therefore only small changes may be seen if the buffer for calibration had been determined. Furthermore, as pure protein obtained through many attempts were significantly small, dilution of the pure protein sample reduced the detection of the pure protein present to almost 0 mg/ml. Therefore, no dilution of the pure protein samples took place. After many attempts of purifying with both Ni NTA and GFP trap resins no further steps were made to purify ABCG2 due to time constraint and the microsomes expressing ABCG2 were used in the thermostability CETSA assay.

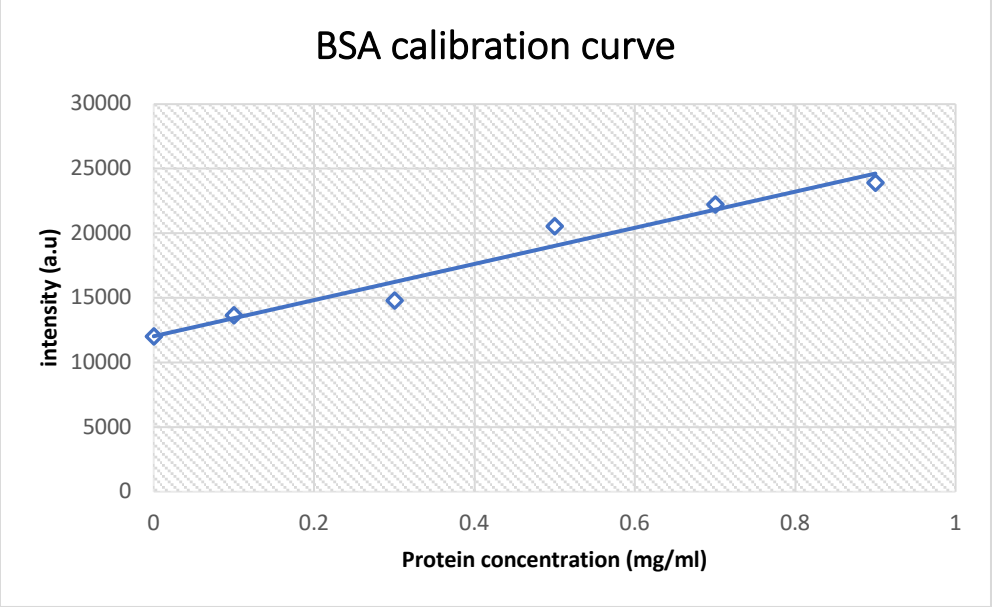


Figure 15: The BSA calibration graph used to determine the protein concentration

The BSA calibration standard graph generated using the protein concentration of BSA and its corresponding intensity from the SDS PAGE gel. The trend line is shown to have the equation of y= 13988x + 12036 with R^2 = 0.9396. In general, the BSA standard curves constructed are not very accurate hence the R^2 = 0.9396 therefore the actual protein concentration might be much less.

3.4 CETSA assay

As both purification methods did not give rise to sufficient pure ABCG2 proteins, ABCG2 microsomes were used in a functional study to determine the thermostability of ABCG2. The CETSA (cellular thermal shift) assay involves the study of the thermo-stabilisation of proteins. Proof of principle of using ABCG2 within this assay would then make it possible to screen small molecules as potential inhibitors and drug candidates. The initial baseline at which ABCG2 denatured was determined by following the protocol for the CETSA assay as mentioned in the methods. In figure 16, as the temperature increases from 37°C to 61°C (increment of 3°C) the intensity of the fluorescent band present on the 10% SDS PAGE gel, corresponding to the presence of the fused ABCG2 - GFP (90kDa) decreases (scanned using Alexa488). After 55°C no bands were seen, concluding that ABCG2-GFP denatures around 55°C. Furthermore, this shows that ABCG2 expressed from *Pichia pastoris* has a melting temperature of 55°C and shows proof of principle for using the CETSA assay to screen for potential inhibitors. Evaluation of the associated thermal shift in the denaturation temperature in the presence of a ligand of interest would allow for the identification of potential binding partners.

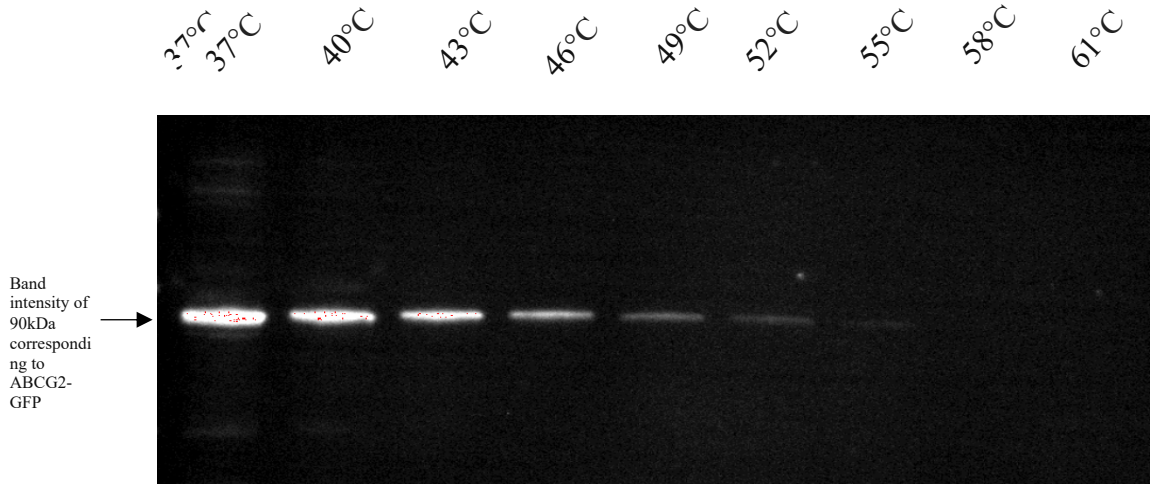


Figure 16: CETSA assay gel conducted on ABCG2 microsomes

On this 10% SDS PAGE gel microsome expressing ABCG2 denatured at around 55 °C, as the intensity of the band corresponding to ABCG2 (90kDa) decreases as the temperature increases up until 55°C then no bands are seen. Furthermore, the band intensity corresponds the green fluorescent detected by Alexa 488 (excitation wavelength:485nm, emission wavelength: 530nm).

Chapter 4: Bioinformatics results

4.1: The structural relationship between ABCA and ABCG family

Understanding the protein's structure and mechanism starts by looking at the structure in detail, this is achieved by using bioinformatic tools such as Chimera. Chimera has been used in this project to look at both structural and sequence alignment between ABCA1, ABCG2, ABCG5/G8, bacterial ABC transporter MacB, MlaE, MlaF WzmWzt; the structural alignment determined in this project between these ABC transporters will provide information on whether there are similarities and differences of the TMD and NBD; with the results, it can be postulated and determined whether convergence or divergence evolution taking place.

Observing the structural alignment has to come with significance to make the alignments valid; Match and Align is a programme used from chimera to look at the alignment from a structural superposition point of view, the programme shows the conserved and weak residues, overall RMSD across the sequence length, SDM and Q score. Looking at origin of evolution of any protein is a very difficult task especially as the common ancestor may not exist anymore, however the fully and weakly conserved residues in the structural alignment helps map a picture of any possible evolutionary links that may exist.

4.1.1 The structure of ABCA1 used in the alignments

Looking at the eukaryotic evolution of ABCG family, helps by looking at other ABC transporter families with similar structure to ABCG transporters. Ford *et al* has suggested the topology and folds of the transmembrane helices of TMDs of ABCA1 are similar to that of transmembrane helices of the ABCG2 TMDs (Ford *et al*., 2019). Before obtaining the structural sequence alignment between ABCA1and ABCG2, the extracellular domain (ECD) and R domain (composed of C-terminal extensions 1 and 2) were removed as shown in figure 17 to obtain a better understanding of the alignment between ABCA1 TMD NBD to ABCG2 TMD NBD. The structural alignments of TMD and NBD were made separately to obtain an even better understanding of the similarities and difference in alignment of the TM helices in TMD, and the α – helices and β- pleated sheets in the NBD.

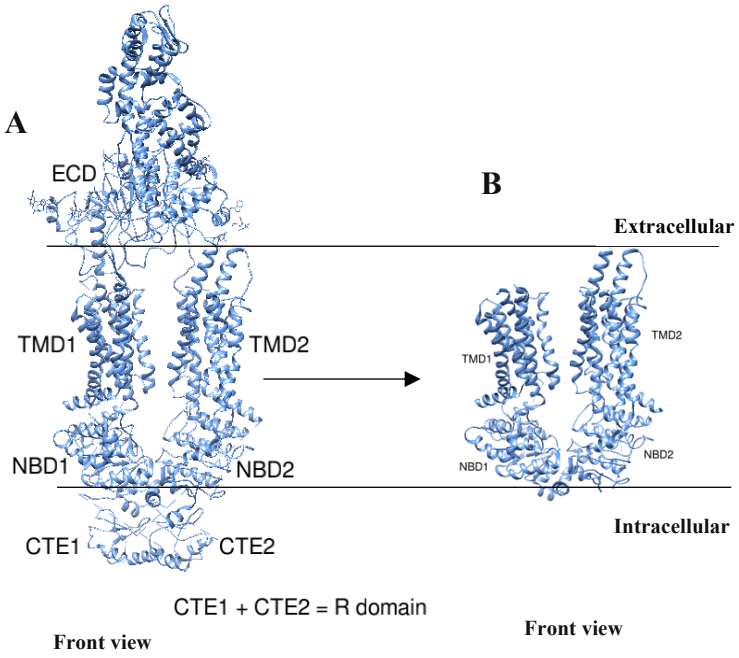


Figure 17:The original ABCA1 present in chimera verses the structure of ABCA1 used in the alignments

A) *Front view: Original structure with all the components present such as the ECD, TMD1,TMD2,NBD1,NBD2, CTE1and CTE (which represents the R domain).* ***B)*** *Front view: The structure of ABCA1 which only contains TMD1, TMD2,NBD1 and NBD2. Both structures are represented in cornflour blue ribbons.*

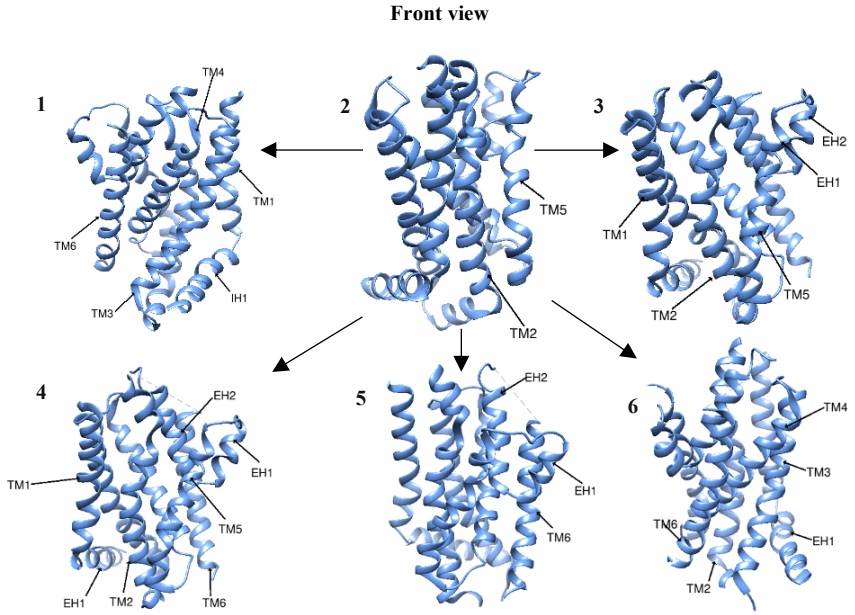


Figure 18:ABCA1 TMD1 used in the alignments to ABCG2 and ABCG5/G8

In this figure there are 6 different views of ABCA1 TMD1 which contains IH1, TM1, TM2, IH2, TM3, TM4, TM5, EH1, EH2 and TM6. The structure is represented in cornflour blue ribbons . All views have been rotated corresponding Front view (view 2); view 1: rotation by anticlockwise 90°, view 3: rotation by clockwise 90°, view 4: rotation by anticlockwise 180°, view 5: rotation by anticlockwise 200°, view 6: rotation by anticlockwise 300°.

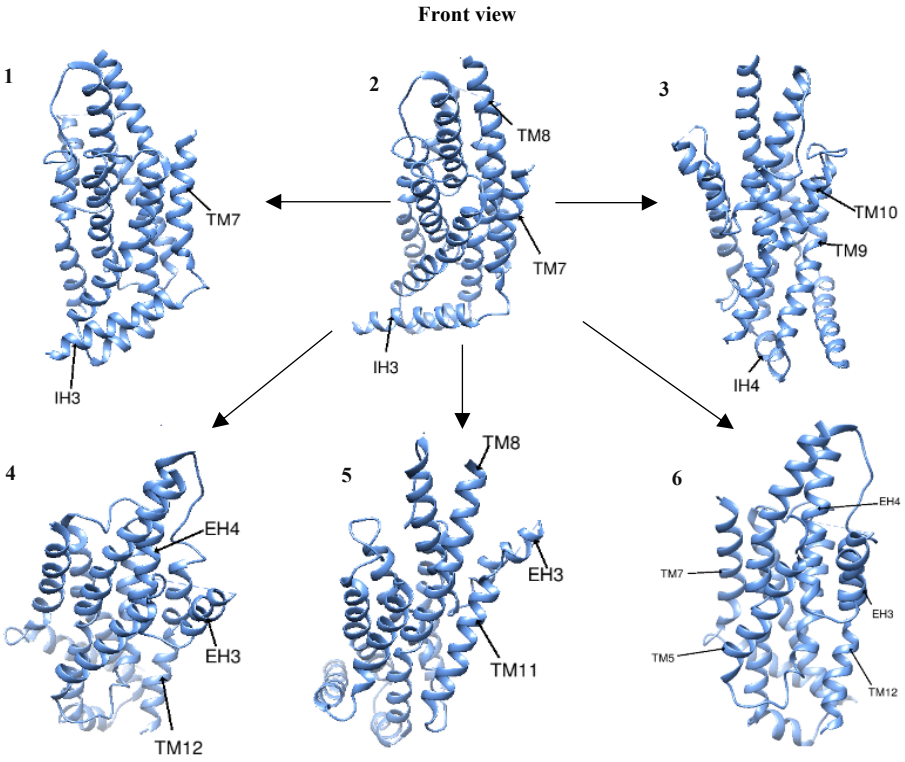


Figure 19:ABCA1 TMD2 used in the alignments of TMDs of ABCG2 and ABCG5/G8

In this figure there are 6 different views of ABCA1 TMD2 which contains IH3, TM7, TM8 ,IH4,TM9,TM19,TM11,EH3,EH4 and TM12. The structure of ABCA1 TMD2 is represented in cornflour blue ribbons. All views have been rotated corresponding Front view (view 2); view 1: rotation by anticlockwise 10°,°, view 3: rotation by clockwise 90°, view 4: rotation by anticlockwise 180°, view 5: rotation by anticlockwise 180°,view 6: rotation by anticlockwise 300°.

4.1.2 The structural alignment of ABCA1 TMD1 and ABCG2 TMD

The best structural alignment of ABCA1 against ABCG2 was achieved at an RMSD 6.981Å across 215 pairs, as seen in figure 20. The structural alignment between ABCA1 and ABCG2 conveyed similar topological structure and folds for the TMDs and NBDs, however this alignment did not show the details of which transmembrane helices of ABCA1 align well with the transmembrane helices of ABCG2. Therefore, structural alignment between the transmembrane helices of ABCA1 TMD1 and ABCG2 TMD were made in order to look at the relationship between the transmembrane helices of each domain. The conserved residues mentioned in the alignment of ABCG2 TMD and ABCA1 TMD1 are referred to in appendix 2 and quantitative results of the specific alignments are represented table 7, which shows a quick and comparative overview of the fully and weak conserved residues of the alignments.

Figure 21 presents the best alignment of ABCA1 TMD1 with ABCG2 TMD at RMSD 4.504**Å** (across 190 C atom pairs). Through observation, there is significant alignment between IH1 and TM1a as seen in figure 21A. Using the Match and Align tool in chimera, IH1 and TM1a, represented between residue 407-430, had 3 fully conserved and 3 weakly conserved residues**.** In figure 21A, B and 22A there is significant alignment between TM1b and TM1, with only 1 fully conserved and 18 weakly conserved residues between residues 431-451.

In figure 21B, 21C and 22A, ABCG2 TM2 and ABCA1 TM2 are shown to align well as they take similar form and topological structure; the structural sequence alignment in figure 23 further confirms this as there are 7 fully conserved and 28 weakly conserved residues. As ABCA1 IH2 is very close to ABCA1 TM2, there is significant alignment of ABCA1 IH2 and segments of ABCG2 TM2.

ABCA1 TM3 and ABCG2 TM3 are shown to align significantly, as seen in figure 21a and 22a, the structural sequence alignment shows from residues 1096-1130 that there are 4 fully conserved and 28 weakly conserved residues. Similar to the alignment between TM3, ABCA1 TM4 and ABCG2 TM4 also shows significant alignment, as seen in figure 21A, 21C and 22A, there are 4 fully conserved and 29 weakly conserved residues between residues 1131-1162. Furthermore, in this structural alignment only at one point are there no conservation of the residues, indicating good structural similarity. So far, the results show conservation between ABCG2 TM1-4 and ABCA1 IH1 TM1-4 indicating an evolutionary relationship which may exist.

ABCG2 TM5a shows good alignment with ABCA1 TM5, as seen in figure 21a, 22c and 22d, which is further reflected in the structural sequence alignment which contains 4 fully conserved and 4 weak conserved residues between residues 1161-1181. ABCG2 TM5c aligns well with ABCA1 EH2, as seen in figure 21C, 21D and 22C, the associated structural sequence alignment shows 1 fully conserved and 16 weak conserved residues between residues 1221-1240. ABCG2 TM6b is shown to align with ABCA1 TM6, as seen in figures 21A, 21C, 22C and 22d; the similarity in the alignment is seen in the structural sequence alignment where between the residues 1291-1330 there is 1 fully conserved and 22 weak conserved residues.
Overall, in the structural alignment of ABCA1 TMD1 and ABCG2 TMD the Match and Align programme found overall RMSD was 2.201 across sequence length of 266248, the

SDM score was 44.827 and the Q score was 0.401 (appendix 1). The overall RMSD score of the structural alignment is within a few angstroms, showing the structural alignment to be significant. Furthermore, the SDM score is less than 50.00 which indicates more structural similarity than difference according to Johnson *et al* study. Finally, the Q score shows some significant similarity in structure to consider evolutionary links according to Krissinel *et al* study (Johnson *et al*., 1990; Krissinel *et al., 2004)*.

The results of the alignment between ABCA1 TMD1 and ABCG2 TMD, shows only segments of TM5 and TM6 align well, interestingly there was no structural alignment between ABCG2 TM5b and ABCA1 TM5, as seen figure 21C, 22C, 22D; through observation ABCG2 TM5b would have been expected to align with EH1 of ABCA1, however this was not the result. ABCG2 TM6a had no alignment with any segment of ABCA1 TM6. The results from the structural alignment of ABCA1 TMD1 and ABCG2 TMD suggests that ABCA1 and ABCG2 may have evolved from the same or similar ancestor especially as the percentage identity was 15.30%. Although in figure 22 the structural alignment demonstrates that regions ABCG2 TM5b and TM6a do not have any alignment to ABCA1, all the other helices do have significant alignment. The alignment of the NBD of ABCA1 and ABCG2 showed high structural similarity as expected (appendix 8,9,10,11,12).

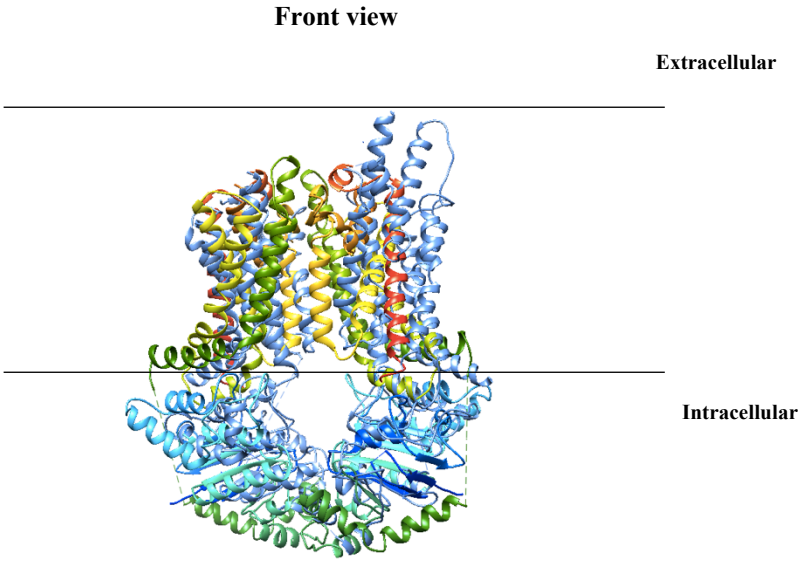


Figure 20: The alignment between ABCA1 and ABCG2

Front view: The structure of ABCA1 is represented in cornflour blue ribbons and ABCG2 transmembrane helices are represented as followed: TM1a – green, TM1b – green, TM2 – light green, TM3 – yellow, TM4 – yellow, TM5a – mustard, TM5b- mustard, TM5c – orange, TM6a- red and TM6b- Red. The initial alignment made between ABCG2 and ABCA1, the structural alignment in this figure shows how there is a similarity in size and shape, however with this structure it is hard to study the alignment of the transmembrane helices and nucleotide binding domain alignments.

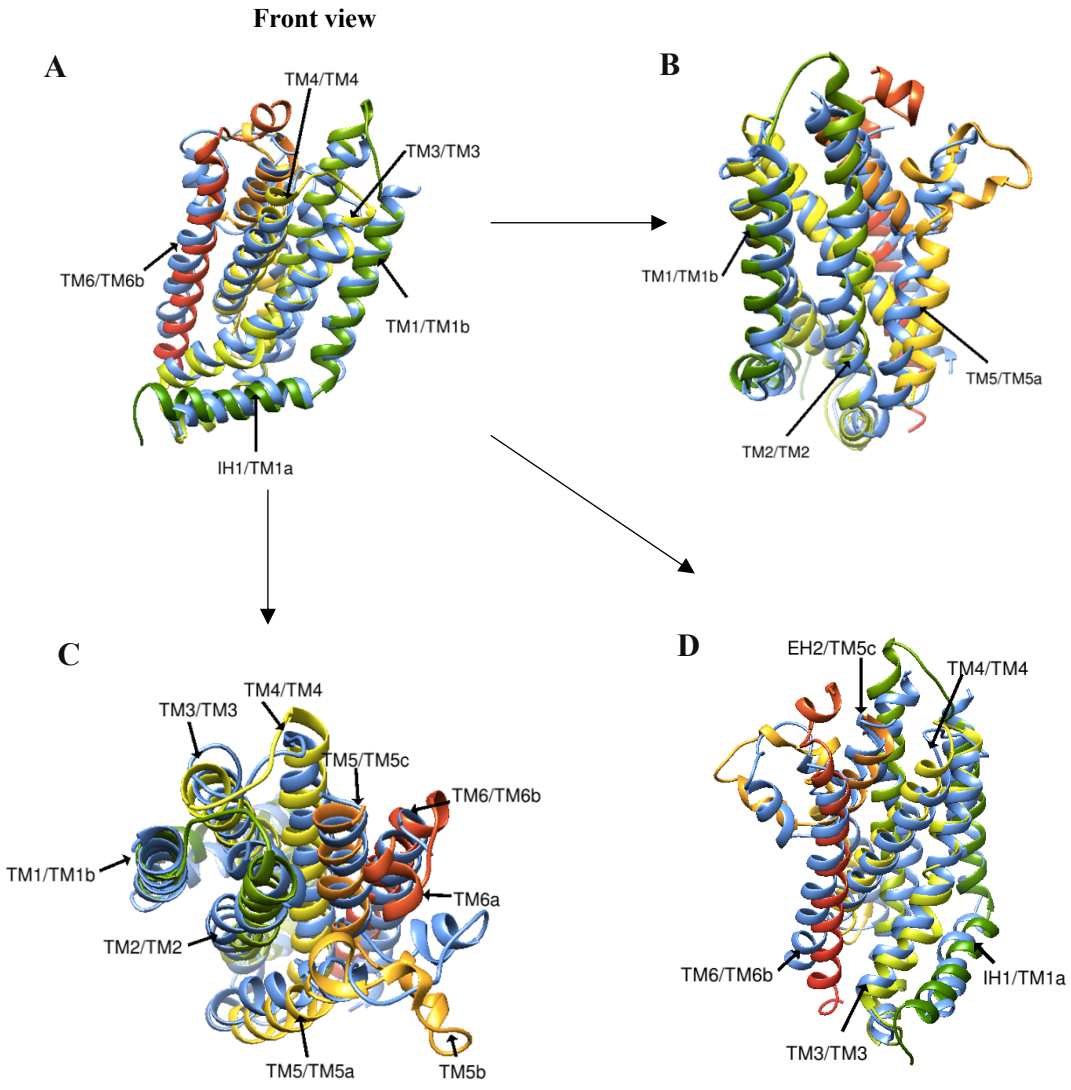


Figure 21:The alignment between ABCA1 TMD1 and ABCG2 TMD

A) *Front view.* ***B)*** *Rotation of front view by clockwise 90°.****C)*** *Overview.* ***D****)Rotation of front view by 60°. The structure of ABCA1 is represented in cornflour blue ribbons and ABCG2 transmembrane helices are represented as followed: TM1a – green, TM1b – green, TM2 – light green, TM3 – yellow, TM4 – yellow, TM5a – mustard, TM5b- mustard, TM5c – orange, TM6a- red and TM6b- red. This figure shows the detailed alignment of the specific transmembrane helices between ABCA1 TMD1 and ABCG2 TMD as followed: ABCA1 IH1/ ABCG2 TM1a, ABCA1 TM1/ ABCG2 TM1b, ABCA1 TM2/ ABCG2 TM2, ABCA1 TM3/ ABCG2 TM3, ABCA1 TM4/ ABCG2 TM4, ABCA1 TM5/ ABCG2 TM5a, ABCA1 EH2/ ABCG2 TM5c, ABCA1 TM6/ ABCG2 TM6b*

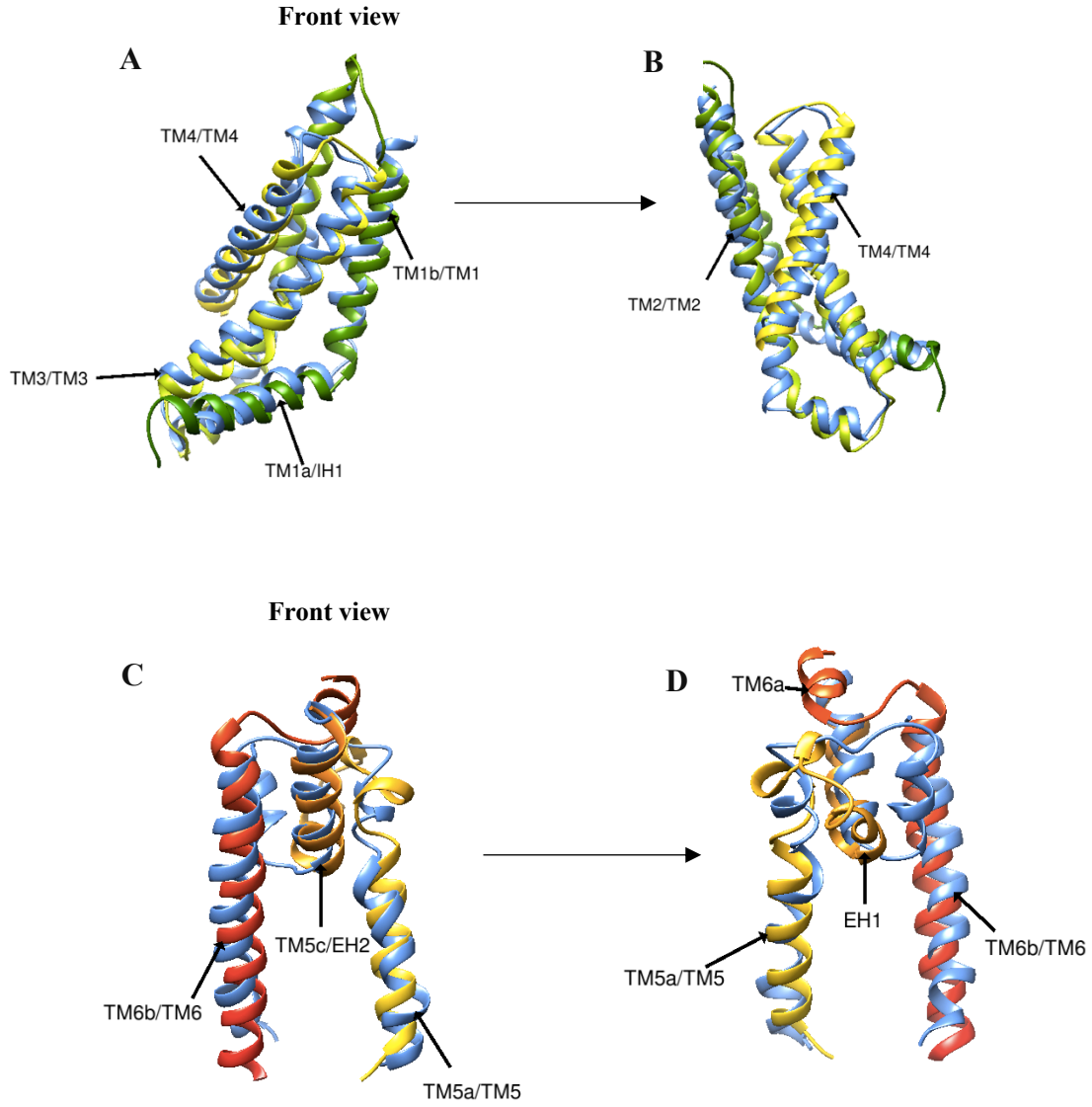


Figure 22: Alignment between ABCA1 IH1 TM1-4 and ABCG2 TM1-4; Alignment of ABCA1 TM5 and TM6 to ABCG2 TM5 and TM6

The structure of ABCA1 is represented in cornflour blue ribbons and ABCG2 transmembrane helices are represented as followed: TM1a – green, TM1b – green, TM2 – light green, TM3 – yellow, TM4 – yellow, TM5a – mustard, TM5b- mustard, TM5c – orange, TM6a- red and TM6b- red. ***A)*** *Front view: the alignment between ABCA1 IH1 TM1-4 and ABCG2 TM1-4 shows the specific transmembrane alignment as followed: ABCA1 IH1/ ABCG2 TM1a, ABCA1 TM1/ ABCG2 TM1b, ABCA1 TM2/ ABCG2 TM2, ABCA1 TM3/ ABCG2 TM3, ABCA1 TM4/ ABCG2 TM4.* ***B)*** *Alignment rotated by anticlockwise 90° front the front view (**A**).* ***C)****Front view: the alignment between ABCA1 TM5 and TM6 to ABCG2 TM5 and TM6 shows the specific transmembrane alignment as followed: ABCA1 TM5/ ABCG2 TM5a, ABCA1 EH2/ ABCG2 TM5c, ABCA1 TM6/ ABCG2 TM6b.****D)*** *Alignment rotated by anticlockwise 180° front the front view(**C**).*

Alignments and residue regions - ABCA1: ABCG2	Fully conserved residues	Weakly conserved residues
IH1:TM1a (407 -430)	3	3
TM1: TM1b (431-451)	1	18
TM2:TM2 (456-700)	7	28
IH2: TM2 (702-1076)	8	34
TM3-TM3 (1096-1130)	4	28
TM4-TM4 (1131-1162)	4	29
TM5: TM5 (1161-1181)	4	4
EH2:TM5c (1182-1290)	1	16
TM6:TM6b (1291-1330)	1	22

Table 7:Quantitative data from the alignment between ABCA1 TMD1 and ABCG2 TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TMD1 and ABCG2 TMD.

4.1.3 The structural alignment between ABCA1 and ABCG8

To evaluate the relationship and evolution of the ABCA and ABCG family, in addition to the structural and sequence alignment between ABCG2 and ABCA1, a structural and sequence alignment between ABCA1 and ABCG5/G8 may further validate the evolutionary relationship between the ABCA and ABCG family. The quantitative structural alignment data mentioned is referred to appendix 13, hence a summary of the fully and weakly conserved regions are referred to table 8.

In the initial structural alignment between ABCA1 and ABCG5/G8, ABCG5 aligned with ABCA1 TMD1 and NBD1, hence, ABCG8 TMD and NBD aligned with ABCA1 TMD2 and NBD2 as seen in figure 23. When using the chimera programme for structural alignment, many factors, such as loops and helices between the main transmembrane domain and nucleotide binding domain can affect the alignment. As a result, ABCG5 and ABCG8 were aligned with ABCA1 separately without the presence of loops and helices. ABCG8 aligned

well with ABCA1 TMD1 and NBD1, as seen in figure 24. Looking specifically at the transmembrane helices alignment between ABCG8 and ABCA1 TMD1, ABCA1 IH1 had significant alignment with ABCG8 TM1, as seen in figure 25A and 26A. Similar to the structural alignment between ABCG2 TM1 and ABCA1 IH1, within the ABCA1 IH1 and ABCG8 TM1 structural sequence alignment, there were 2 fully conserved residues and 13 weak conserved groups between residues 421-450.

Furthermore, similar to the alignment of ABCG2 TM1b and ABCA1 TM1, ABCG8 TM1 has significant alignment with ABCA1 TM1, as seen in figure 25A and 26A; with 2 fully conserved residues and 17 weak conserved groups between residues 451- 480. In comparison to ABCG2 TM1b and ABCA1 TM1 the conserved residues and weak conserved groups are better between ABCG8 TM1 and ABCA1 TM1.

TM2 from both ABCG8 and ABCA1 showed to take similar topology and fold similar to the alignment between TM2 in both ABCA1 and ABCG2. Figure 25B and 26B showed significant alignment between TM2 of both ABCG8 and ABCA1, in the structural sequence alignment, there were 4 fully conserved residues and 28 weak conserved groups between 1061-1101; in comparison to the alignment of both ABCG2 and ABCA1 TM2 there are less fully conserved residues present.

The alignment between TM3 of both ABCG8 and ABCA1 shows to take similar topology and fold as seen in figure 25C and 26A, thus in the sequence alignment, there were 2 fully conserved residues, and 25 weak conserved residues present between residue 1161-1160. TM4 of ABCA1 and ABCG8 is also seen to take similar topology as seen in figure 25C and 26A, the structural sequence alignment showed 4 fully conserved residues and 30 weak conserved group between residues 1161-1190.

The structural alignment of TM1-TM4 between ABCG8 and ABCA1 showed similar structural alignment when compared to TM1-4 between ABCG2 and ABCA1. However, similar to TM5 and TM6 alignment between ABCG2 and ABCA1, only segments of ABCG8 TM5 align with ABCA1 TM5. ABCG8 TM5 has significant alignment structurally with ABCA1 TM 5 between the residues 1191-1200 as seen in figure 25C, 26C and 26D. In the structural sequence alignment, there were 2 fully conserved residues alongside 11 weak conserved. ABCG8 TM5 structurally aligns well with ABCA1 EH4 between residues 1251-1264, with 1 fully conserved site and 11 weak conserved residues. TM6 from ABCA1 and

ABCG8 is also seen to have similar alignment as seen in figure 25D,26C and 26D, the sequence structural alignment shows 2 fully conserved residues and 23 weak conserved residues between 1311-1341.

Overall, in the structural alignment of ABCG8 TMD and ABCA1 TMD1 the Match and Align programme found, overall RMSD was 2.264 across sequence length of 266248, the SDM score as 46.253 and the Q score as 0.453. Similar to the alignment between ABCG2 and ABCA1, there is significant alignment between TM1-TM4 between ABCG8 and ABCA1 compared to the alignments of TM5 and TM6.

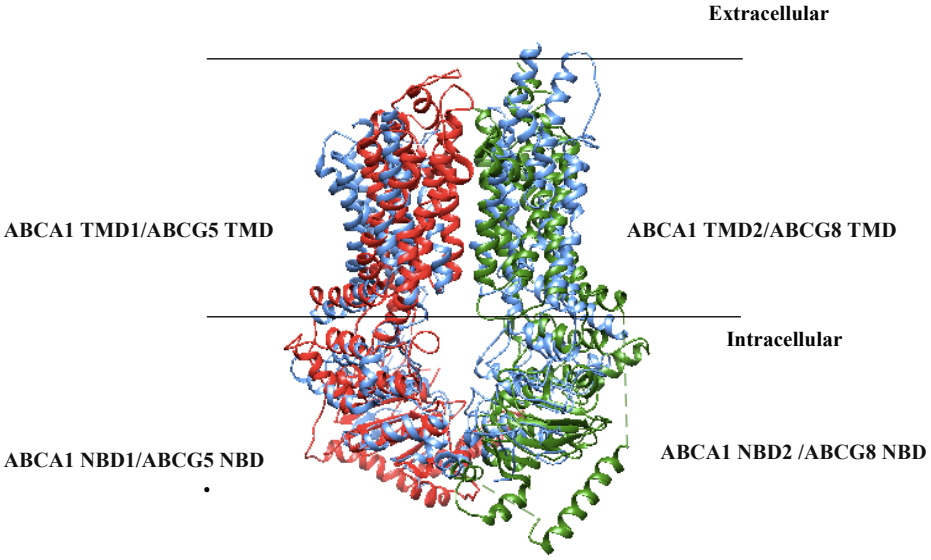


Figure 23: The structural alignment between ABCG5/G8 and ABCA1

The structure of ABCA1 is represented in cornflour blue. The structure of ABCG5 is represent in red ribbons and ABCG8 represented in forest green ribbons. This figure shows ABCG5 to align with ABCA1 TMD1 NBD1 and ABCG8 shows to align with ABCA1 TMD2 NBD2.

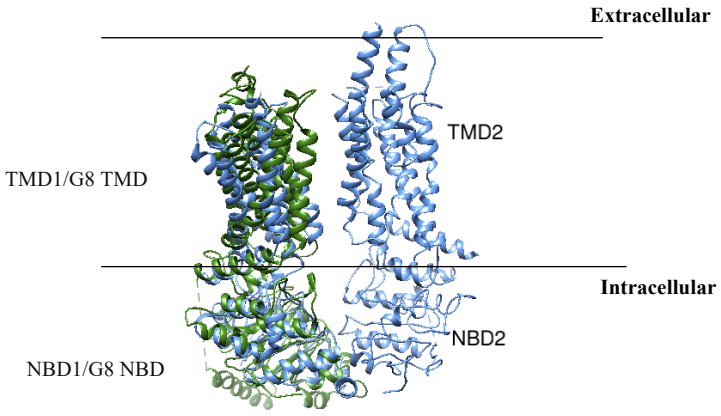


Figure 24:Alignment of ABCG8 andABCA1

ABCG8 represented as forest green and ABCA1 represented as cornflower blue. The alignment shows that ABCG8 TMD and NBD aligns well with ABCA1 TMD1 and NBD1 rather than ABCA1 TMD2 and NBD

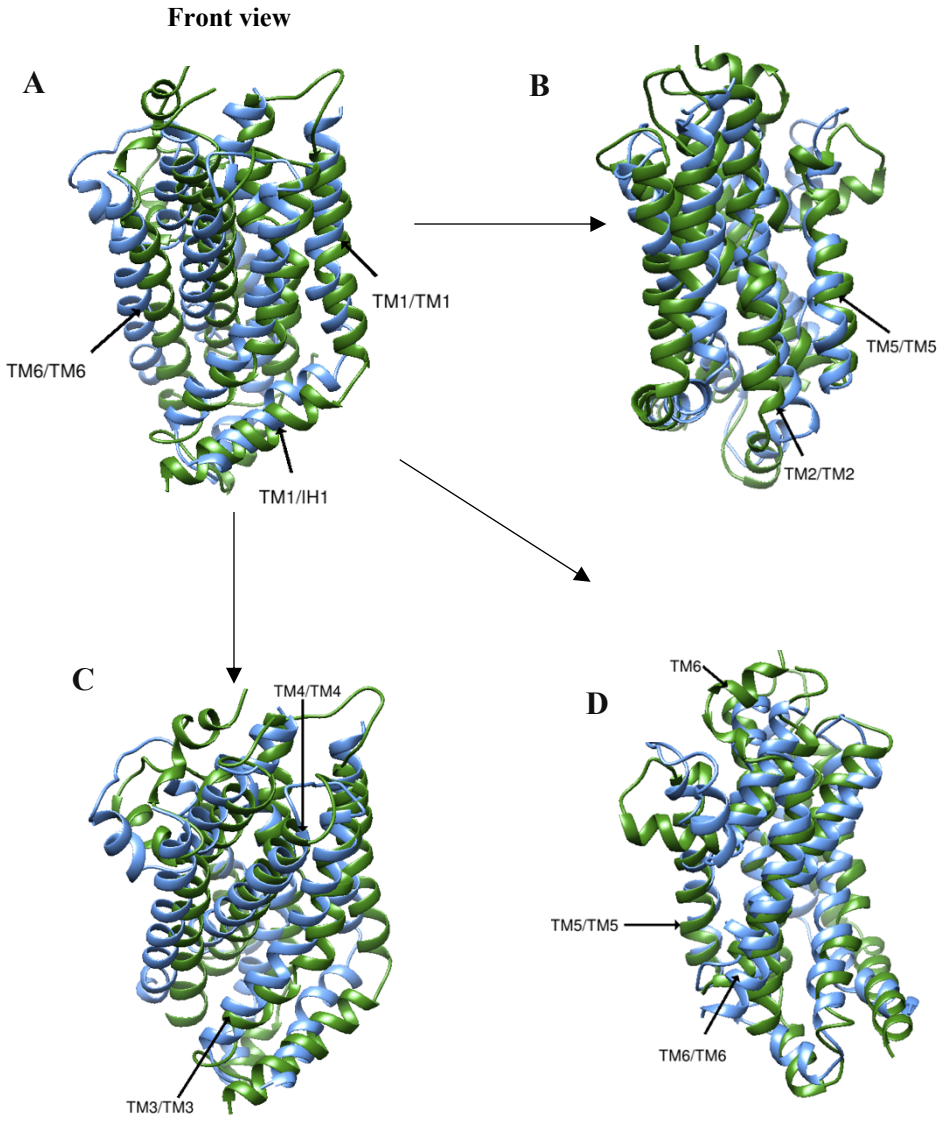


Figure 25: The alignment between ABCA1 TMD1 and ABCG8 TMD

A*) Front view.* ***B****) Rotation by clockwise 90° from front view.* ***C****) Rotation by clockwise 100° from front view* ***D****) Rotation by clockwise 200° from front view. The structure of ABCG8 TMD is represented in forest green and ABCA1 TMD1 is represented in cornflour blue This figure shows the detailed alignment of the specific transmembrane helices between ABCA1 TMD1 and ABCG2 TMD as followed: ABCA1 IH1/ ABCG8 TM1, ABCA1 TM1/ ABCG8 TM1, ABCA1 TM2/ ABCG8 TM2, ABCA1 TM3/ ABCG8 TM3, ABCA1 TM4/ ABCG8 TM4, ABCA1 TM5/ ABCG8 TM5a, ABCA1 EH2/ ABCG2 TM5, ABCA1 TM6/ ABCG8 TM6*

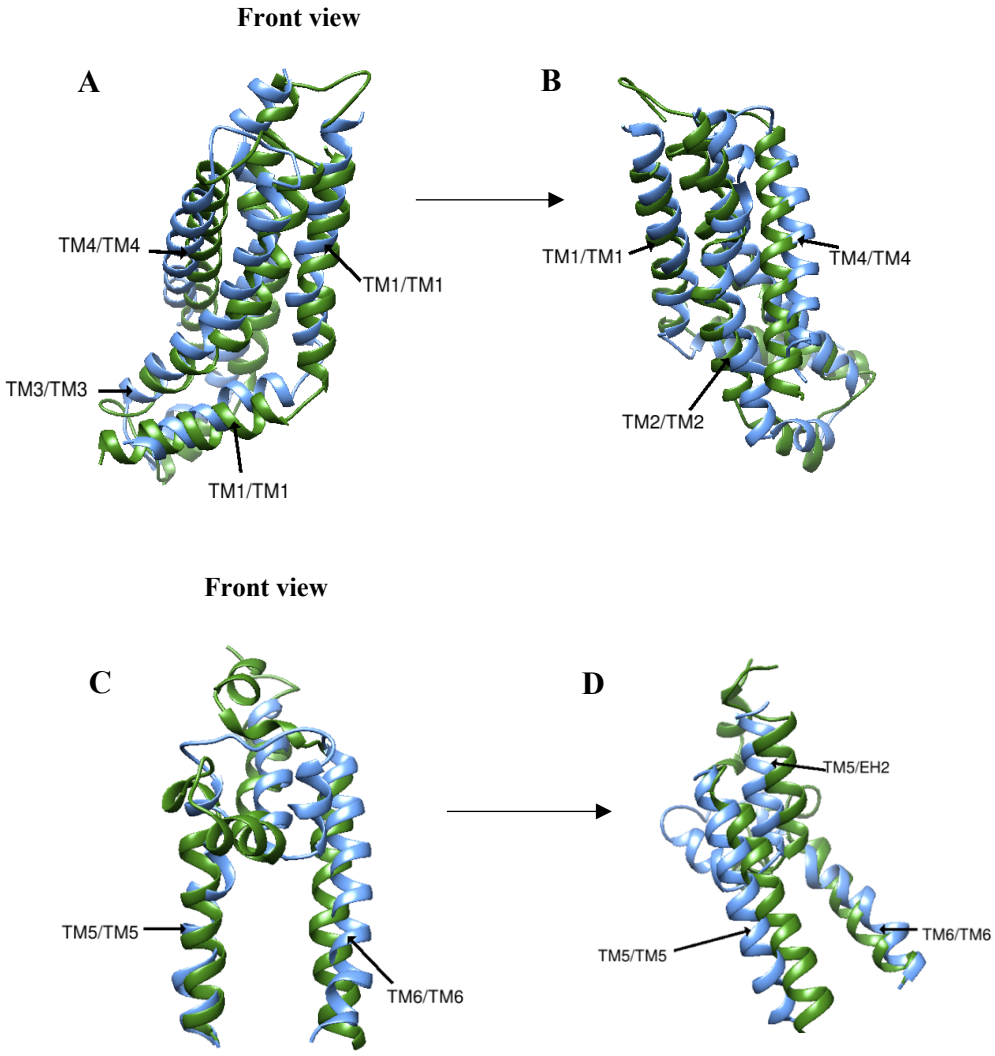


Figure 26:Alignment between ABCA1 IH1 TM1-4 and ABCG5 TM1-4; Alignment of ABCA1 TM5 and TM6 to ABCG8 TM5 and TM6

A) *Front view.* ***B)*** *Rotation by anticlockwise 90° from front view (**A**).* ***C)*** *Front view.* ***D)*** *Alignment Tilted upwards 45° front view (**C**). The structure of ABCG8 TMD is represented in forest green and ABCA1 TMD1 is represented in cornflour blue.* ***A, B)*** *The alignment between ABCA1 IH1 TM1-4 and ABCG8 TM1-4 shows the specific transmembrane alignment as followed: ABCA1 IH1/ ABCG8 TM1, ABCA1 TM1/ ABCG8 TM1, ABCA1 TM2/ ABCG8 TM2, ABCA1 TM3/ ABCG8 TM3, ABCA1 TM4/ ABCG8 TM4.* ***C, D)*** *The alignment between ABCA1 TM5 and TM6 to ABCG8 TM5 and TM6 shows the specific transmembrane alignment as: ABCA1 TM5/ ABCG8 TM5a, ABCA1 EH2/ ABCG2 TM5, ABCA1 TM6/ ABCG8 TM6*

Alignments and residue regions – ABCA1:ABCG8	Fully conserved residues	Weakly conserved residues
IH1:TM1 (421-450)	2	13
TM1:TM1 (451-480)	2	17
TM2:TM2 (1060-1101)	4	28
TM3:TM3 (1102-1160)	2	25
TM4:TM4 (1161-1190)	4	30
TM5:TM5 (1191-1200)	2	11
EH4:TM5 (1251-1264)	1	11
TM6:TM6 (1311-1341)	2	23

Table 8: Quantitative data from the alignment between ABCA1 TMD1 and ABCG8 TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TMD1 and ABCG8 TMD.

4.1.4 The structural alignment between ABCG5 TMD and ABCA1 TMD1

As previously mentioned, the initial alignment between ABCA1 and ABCG5/G8 showed that ABCG5 aligned with ABCA1 TMD1. Without the presence of ABCG8, ABCG5 continued to align with ABCA1 TMD1 as seen in figure 27. However, the alignment of the TM helices was better in the alignment between ABCG8 TMD and ABCA1 TMD1 than with ABCG5. As for the alignment of the TM helices of ABCG5 TMD and ABCA1 TMD1 the overall RMSD was 2.428 Å; SDM value was 41.883 and Q score being 0.299. The structural alignment data is referred from appendix 14, hence table 9 shows a summarised view of the fully and weakly conserved residues in the superimposed structural alignment.

ABCA1 IH1 structurally aligned significantly to ABCG5 TM1 as seen in figure 28A, which is similar to the alignment between ABCG8 TM1 and ABCA1 IH1 as expected. Furthermore, the sequence alignment of ABCG5 TM1 and ABCA1 IH1 had 3 fully conserved sites and 9 weak conserved residues between residues 360 and 404. Interestingly, ABCG5 TM1 and ABCA1 TM1 shown in figure 28A has some alignment but not as well as in the alignment between ABCG8 TM1and ABCA1 TM1; the structural sequence alignment shows between residues 405 and 442 , there were 2 fully conserved regions and 15 weak conserved residues.

The alignment of TM2 of ABCG5 and ABCA1 shows to have some structural alignment, however the helices are shown to not align as well as seen in figure 28A, this is reflected in the structural sequence alignment as there were no fully conserved residues and only 24 weak conserved residues between 1031-1079.

TM3 from both ABCA1 and ABCG5 is seen to align well as seen in figure 28C, where the helices overlap and convey similar structure; the sequence alignment shows there to be 4 fully conserved residues alongside 35 weak conserved residues 1080-1150. Surprisingly, TM4 from both ABCA1 and ABCG8 is seen to structurally align well as the helices overlap and convey similar structure as seen in figure 28C, the structural sequence alignment only shows 1 fully conserved residue and 24 weak conserved group between 1150-1200. As the alignments between ABCG5 TMD and ABCA1TMD1 show high similarity this supports the theory that ABCA and ABCG families may have a common ancestor.

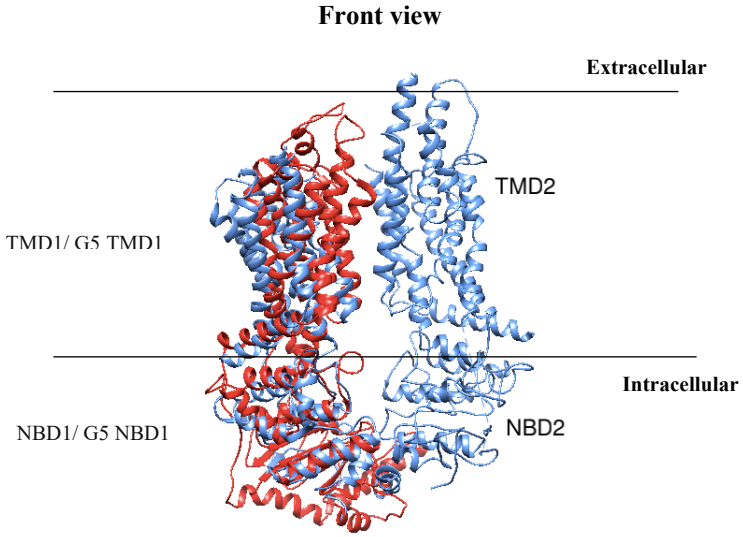


Figure 27: The structural alignment between ABCA1 and ABCG5

ABCG5 represented as red and ABCA1 represented as cornflower blue. The alignment shows that ABCG5 TMD and NBD aligns well with ABCA1 TMD1 and NBD1 rather than ABCA1 TMD2 and NBD2.

Front view

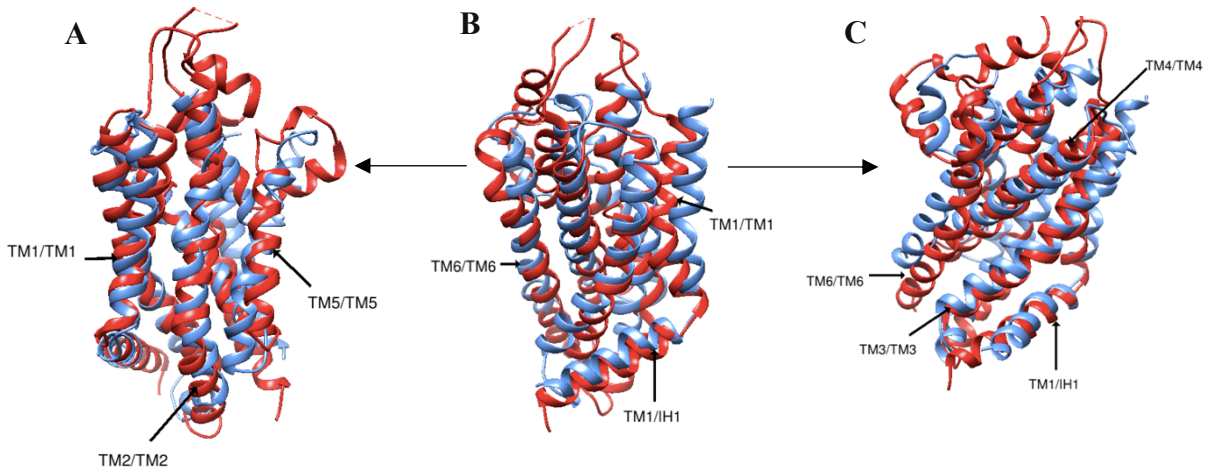


Figure 28:The structural alignment between ABCG5 TMD and ABCA1 TMD1

A) *Alignment rotated by clockwise 90° from front view.* ***B)*** *Front view.* ***C)*** *Alignment rotated by anticlockwise 60° from front view. .The structure of ABCG5 TMD is represented in red and ABCA1 TMD1 is represented in cornflour blue. This figure shows the detailed alignment of the specific transmembrane helices between ABCA1 TMD1 and ABCG5 TMD as followed: ABCA1 IH1/ ABCG5 TM1, ABCA1 TM1/ ABCG5 TM1, ABCA1 TM2/ ABCG5 TM2, ABCA1 TM3/ ABCG5 TM3, ABCA1 TM4/ ABCG5 TM4, ABCA1 TM5/ ABCG5 TM5, ABCA1 EH2/ ABCG5 TM5, ABCA1 TM6/ ABCG5 TM6*

Alignments and residue regions – ABCA1:ABCG5	**Fully conserved residues**	**Weakly conserved residues**
IH1:TM1 (360-404)	3	9
TM1:TM1 (405-442)	2	15
TM2:TM2 (1031-1079)	0	24
TM3:TM3 (1080-1150)	4	35
TM4:TM4 (1151-1200)	1	24
TM5-TM5 (1201-1260)	4	32
TM6-TM6 (1261-1300)	5	20

Table 9: Quantitative data from the alignment between ABCA1 TMD1 and ABCG5 TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TMD1 and ABCG5 TMD.

4.1.5 The structural alignment between ABCA1 NBD1 and ABCG5/G8 NBD

The alignment between ABCA1 NBD1 and ABCG5G8 NBD was carried out rather than with NBD2 as both ABCG5 and ABCG8 aligned significantly with NBD1. The alignment between ABCG5 NBD and ABCA1 NBD1, most of the $\alpha\ helices$ and β pleated sheets are seen to align (appendix 15), this reflected by the overall RMSD of 2.208, SDM as 45.973, Q score as 0.403. These results show high structural similarity, hence supporting the structural alignment. Similarly, the alignment between ABCG8 NBD and ABCA1 NBD1, most of the $\alpha\ helices$ and β pleated sheets (appendix 15), this reflected by the overall RMSD overall RMSD of 2.263, SDM as 41.204, Q score as 0.427. Similar to the alignment of ABCG5 NBD alignment to ABCA1 NBD, the results of the alignment between ABCG8 NBD to ABCA1 NBD shows high structural similarity, hence supporting the structural alignment.

Overall, the results showed a defined structural relationship between ABCA1 to ABCG2 and ABCG5G8, this establishes a clear indication that ABCA and ABCG families possibly originated from the same ancestor. Furthermore, as TMD are highly variable and there were significant structural alignment found between ABCA and ABCG families strongly supports this postulation.

4.2 The relationship between bacterial ABC transporter MacB to ABCA and ABCG families

4.2.1 The structural alignment between ABCG2 and 5WS4

The first structure used in the analysis of the potential relationship between bacterial mechanotransducers MacB and ABCG2 was the crystal structure of tripartite- type ABC transporter MacB from *Acinetobacter baumanni* represented as 5WS4 from RCSB protein data bank. 5WS4 is homodimer with the presence of a periplasmic domain, 4 transmembrane domains on one protomer and a nucleotide binding domain, for the structural alignment of 5WS4 with ABCG2, ABCA1 and ABCG5/G8 the periplasmic domain was removed, hence, the structure of 5WS4 without the periplasmic domain is seen in figure 29. The conserved and weak residues mentioned in the alignment of ABCG2 TMD and 5WS4 TMD is referred from appendix 16, thus the quantitative data for the conserved and weak residues mentioned in the alignments is represented in table 10.

In the structural alignment between TM helices of 5WS4 and ABCG2 there were very little alignment of the TM helices as seen in figure 30. The alignment in figure 30 showed that 5WS4 was angled towards TM1-4 of ABCG2 yet no alignment towards ABCG2 TM5 or TM6. Considering that in Ford *et al* review paper it suggested there were similar structure of ABCG2 TM1-4 and potentially bacterial mechanotransducers MacB TM1-4, TM5 and TM6 from ABCG2 was removed and the rest of the TM helices were aligned to 5WS4 TMD. In the structural alignment between 5WS4 TM helices and ABCG2 TM1-4 there was an overall RMSD of 2.316, SDM of 51.918 and Q score of 0.248, using Clustal omega alignment the percent identity worked out to be 19.28%

In figure 31A, the 5WS4 elbow is seen to take similar topological structure as ABCG2 TM1a however no overlap of the helices was present. In the structural sequence alignment between 651-680 there were 10 weak conserved, although no fully conserved residues, the weak conserved residues indicate some structural similarity. The alignment ABCG2 TM2 and 5WS4 TM2 showed to have some overlap of the helices, and overall seen to convey similar structure as seen in figure 31B. The structural sequence alignment showed 3 fully conserved residues and 31 weak conserved residues. The alignment of ABCG2 TM3 and 5WS4 TM3 showed an overlap of the helices which conveys to take similar structure as seen in figure 31C and 31D, the structural sequence alignment shows the presence of one fully conserved residue and 26 weak conserved residues. ABCG2 TM4 and 5WS4 TM4 showed to align well as seen in figure 31C and 31D, where there were similarity in the overlap of the helices. The structural sequence alignment for ABCG2 TM4 and 5WS4 TM4 presented 27 weak conserved residues between 1001-1090.

In the alignments between 5WS4 TM1-4 and ABCG2 TM1-4, the best structural alignment was seen with TM3 and TM4 of 5WS4 and ABCG2 even though the structural sequence alignment shows alignment of TM2 has higher fully conserved residues. Looking at the overall results so far, the alignment between ABCG2 TM1-4 and 5WS4 TM1-4 have good structural alignment considering that structural analysis of ABCG2 and 5WS4 has not been looked at in detail yet, this indicates that ABCG families may have MacB as the original ancestor where it evolved from.

The structural alignment of the 5WS4 NBD and ABCG2 NBD is shown to be highly similar especially as the 8 $\alpha - helices$ overlap, the $\beta -$ pleated sheets are also seen to take similar

structure and conformation (appendix17); the overall RMSD as 1.930, SDM as 40.556, Q score as 0.394 and the percentage identity worked out to be 20.01%. The structural sequence alignment showed 49 fully conserved residues and 139 weak conserved residues between residue 651-1100, indicating high structural similarity (appendix 18). The structural alignment of the NBD of 5WS4 and ABCG2 sheds light in supporting the theory suggested by Ford *et al* that ABCG family may have evolved from bacterial mechanotransducers MacB.

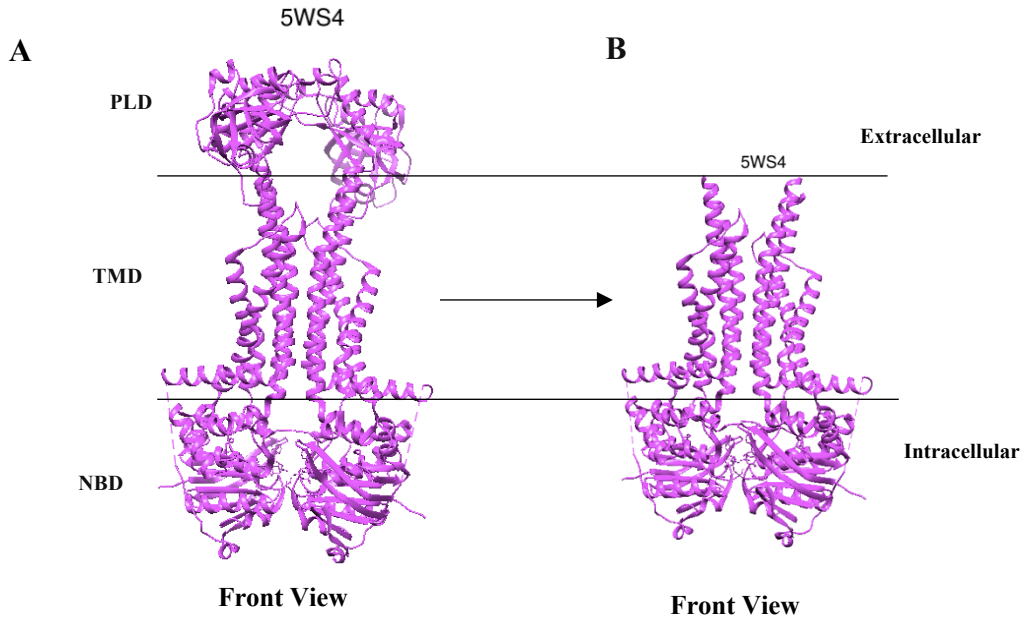


Figure 29:Structure of 5WS4 present in chimera VS the structure used in the alignments to ABCA1, ABCG2 and ABCG5/G8.

The structure of 5WS4 in this figure is represented in magenta. ***A)*** *The structure of 5WS4 with the presence of PLD, MD and NBD.* ***B)*** *The structure of 5WS4 used in the alignments of ABCA1, ABCG2 and ABCG5/G8.*

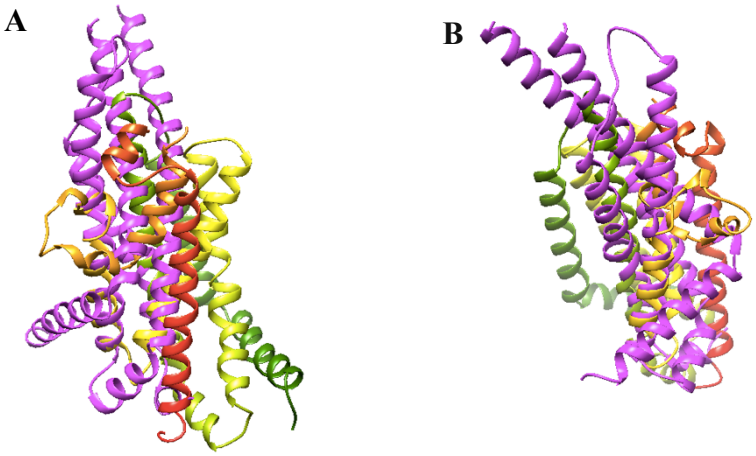


Figure 30:The initial alignment of ABCG2 TMD and 5WS4 TMD

In this figure 5WS4 is represented in magenta and the ABCG2 transmembrane helices are represented in various colours of the rainbow. The initial alignment between the transmembrane domain of ABCG2 and 5WS4 showed the alignment to be slightly distorted, however 5WS4 tended towards ABCG2 TM1-4.

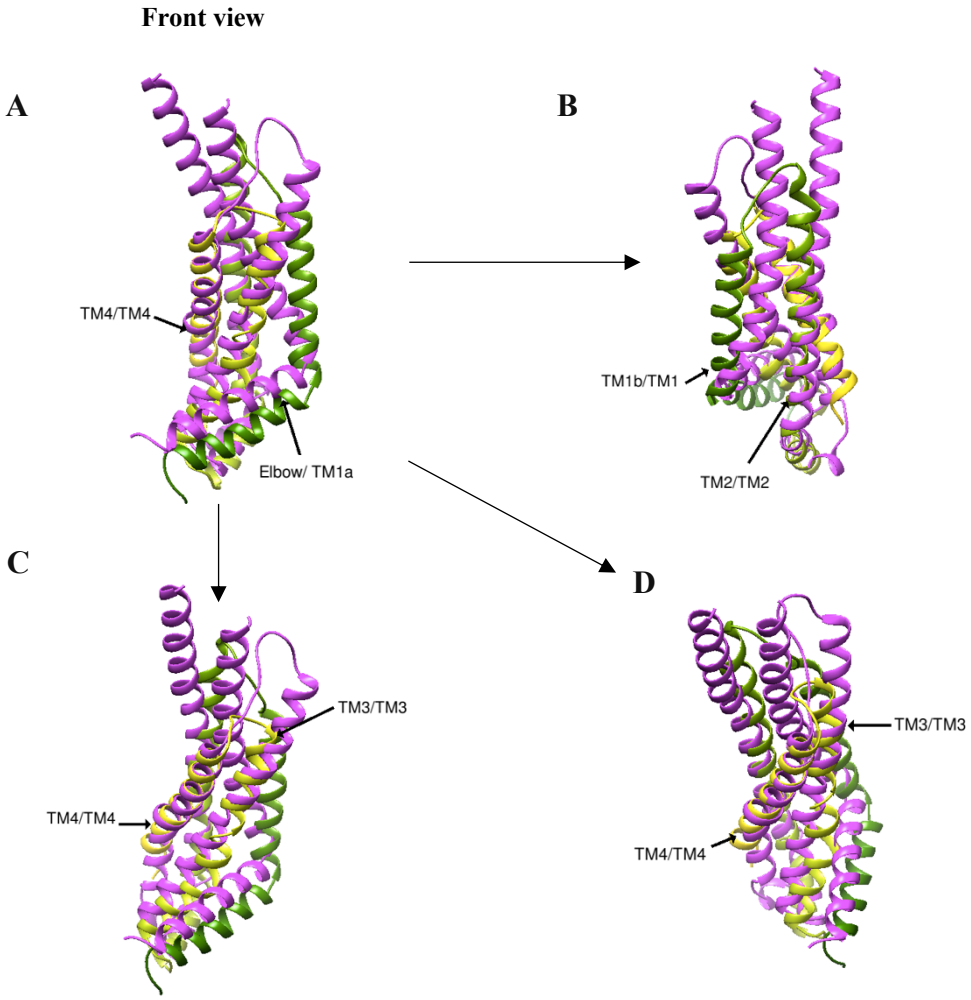


Figure 31:Alignments between ABCG2 TM1-4 and 5WS4 TM1-4

The structure of 5WS4 is represented in magenta, structure of ABCG2 is represented in yellow and green. ***A)*** *Front view: the structural alignment between; 5WS4 elbow and ABCG2 TM1a, 5WS4 TM4 and ABCG2 TM4.****B)*** *Rotation by clockwise 90° from front view: the structural alignment between; ABCG2 TM2 and 5WS4 TM2, ABCG2 TM1b and 5WS4 TM1. C) Rotation by clockwise 60° from front view.* ***D****) Rotation by anticlockwise 90° from front view.* ***C, D)*** *The structural alignment between; ABCG2 TM3 and 5WS4 TM3, ABCG2 TM4 and 5WS4 TM4.*

Alignments and residue regions – 5WS4 TMD: ABCG2 TM1-4	**Fully conserved residues**	**Weakly conserved residues**
Elbow:TM1a (651-680)	0	10
TM2:TM2 (681-898)	3	31
TM3:TM3 (900-1001)	1	26
TM4:TM4 (1001-1090	0	27

Table 10: Quantitative data for the alignment between ABCG2 TM1-4 and 5WS4 TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG2 TM1-4 and 5WS4 TMD

4.2.2 The structural alignment between ABCG2 and 5LIL

The second structure used in the analysis of the potential relationship between bacterial mechanotransducers MacB and ABCG2, was the structure of *Aggregatibacter actinomycetemcomitans* MacB bound to ATPyS (5LIL). Similar to 5WS4, 5LIL is a homodimer with the presences of a periplasmic domain, 4 transmembrane and a nucleotide binding domain on one protomer. There are 2 protomers which make up 5LIL, in the structural alignment made between 5LIL and ABCG2, ABCG5/G8 and ABCA1 the periplasmic domain was removed as seen in figure 32. Furthermore, only one protomer of 5LIL was used in the alignment; the conserved residues mentioned in the alignment of ABCG2 TM1-4 and 5LIL TMD is referred from appendix 19; table 11 represents the summarised conserved regions in the alignment between ABCG2 TM1-4 and 5LIL TMD.

When 5LIL was first aligned with ABCG2, there were no significant alignment of the TMD, however there were alignment of the NBD. When there was just alignment of the TMD of ABCG2 and 5LIL, there were some alignments; noticing that structurally 5LIL has 4 transmembrane which aligns with ABCG2 TM1-4, TM5 and TM6 were removed from ABCG2 and the alignments of ABCG2 TM1-4 and 5LIL TMD were made. In the structural alignment between ABCG2 TM1-4 and 5LIL TMD there was significant alignment between ABCG2 TM1a and the amphipathic helix of 5LIL as seen in figure 33A. Through analysis within that structurally superimposed region there were 2 fully conserved residues, and 20 weak conserved groups present between residues 646-670. Compared to the previous alignment between ABCG2 TM1b to 5WS4, it was expected that ABCG2 TM1b aligns with 5LIL TM1; interestingly in the alignment 5LIL TM1 and ABCG2 do not align as well and diverge outwards away from each as seen figure 33A.

There is partial alignment of ABCG2 TM2 and 5LIL TM2 as seen in figure 33B, where both the helices are seen to converge towards the end of the alignment. The structural sequence alignment showed 2 fully conserved residues and 23 weak conserved groups between residues 931-980. 5LIL TM3 and ABCG2 TM3 is seen to align the best as they both convey similar topology and fold as seen in figure 33C, the structural sequence analysis showed 3 fully conserved regions and 29 weak conserved groups between residues 981-1030. 5LIL TM4 and ABCG2 TM4 aligned well in the beginning however both the transmembrane diverge out towards the end as seen in figure 33C, the structural sequence alignment showed

only one fully conserved residue present alongside 21 other weak conserved groups between residues 1031-1070.

The results of the alignment between 5LIL TMD and ABCG2 TM1-4 take similar structure and form as reflected by the conserved and weak residues of each helices alignment. Overall TM3 and TM1a/amphipathic helix are seen to be the best aligned between 5LIL and ABCG2 as structurally observed both the helix overlap and align better and have higher fully conserved residues in the structural sequence alignment compared to in the alignment of TM1, TM2 and TM4. There was significant structural alignment of the NBD between ABCG2 and 5LIL (appendix 20), most of the α helices and β pleated sheets are aligned well. This is supported by the structural sequence alignment as there were 46 fully conserved residues between residues 51-350; the overall RMSD as 2.413, SDM as 49.304, Q score as 0.329 (appendix 21). Although small divergences were noticed between 5LIL TM1 and ABCG2 TM1b it can be concluded there to be high structural similarity especially as the showed overall RMSD as 2.309, SDM as 50.907, Q score as 0.364 and the percentage identity worked out to be 15.36%. Henceforth, these results indicate high structural similarity therefore indicating a high possibility that ABCG family originates from MacB transporters.

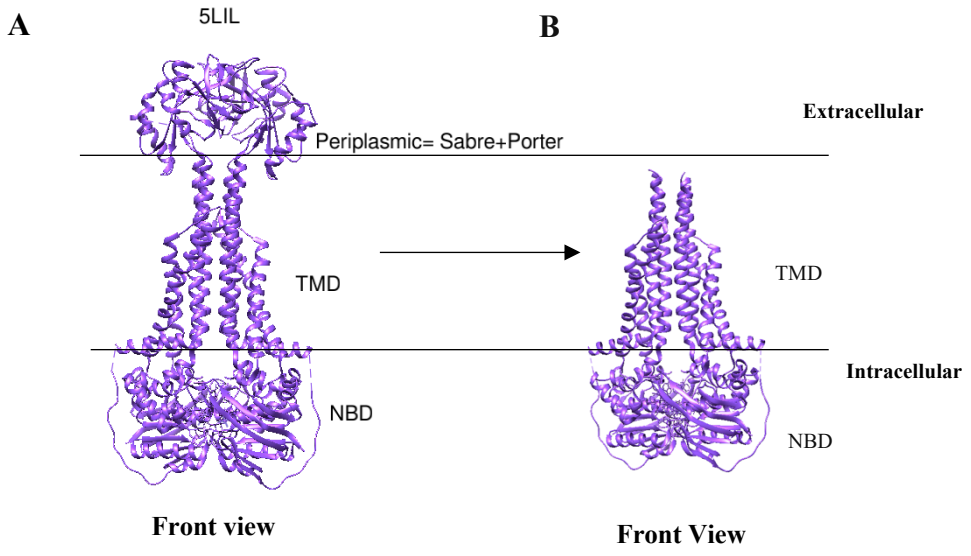


Figure 32:The structure of 5LIL in chimera VS the structure of 5LIL used in the structural alignment to ABCA1, ABCG2, ABCG5/G8.

The structure of 5LIL in this figure is represented in purple. ***A)*** *The structure of 5LIL with the presence of periplasmic domain, TMD and NBD.* ***B)*** *The structure of 5LIL used in the alignments of ABCA1, ABCG2 and ABCG5/G8.*

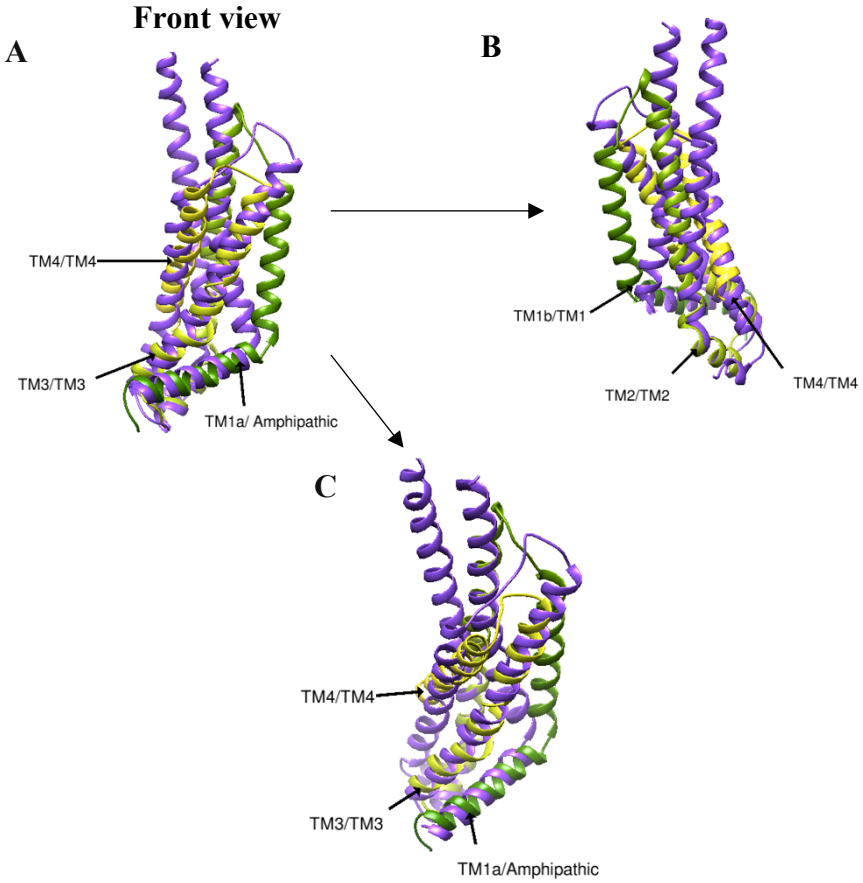


Figure 33:The structural alignment of 5LIL TMD and ABCG2 TMD

The structure of 5LIL is represented in purple, structure of ABCG2 is represented in yellow and green. ***A)*** *Front view: the structural alignment between; 5LIL amphipathic helix and ABCG2 TM1a, 5LIL TM3 and ABCG2 TM3,5LIL TM4 and ABCG2 TM4.* ***B)*** *Rotation by anticlockwise 90° from front view: the structural alignment between; ABCG2 TM2 and 5LIL TM2, ABCG2 TM1b and 5LIL TM1.****C)*** *Rotation by clockwise 90° and tilt 25° from front view: the structural alignment between; ABCG2 TM3 and 5LIL TM3, ABCG2 TM4 and 5LIL TM4, 5LIL amphipathic helix and ABCG2 TM1.*

Alignments and residue regions – 5LIL TMD: ABCG2 TM1-4	**Fully conserved residues**	**Weakly conserved residues**
Amphipathic helix:TM1a (646:670)	2	20
TM2:TM2 (931:980)	2	23
TM3:TM3 (981-1030)	3	29
TM4:TM4 (1031-1070)	1	21

Table 11: Quantitative data for the alignment between ABCG2 TM1-4 and 5LIL TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG2 TM1-4 and 5LIL TMD

4.2.3 The structural alignment between ABCG2 and 5NIK

The third structure used in the analysis of the potential relationship between mechanotransducers MacB and ABCG2, is the structure of MacAB-TolC ABC type tripartite multidrug efflux pump from *E.coli* -K12. In the structural alignment only the MacB component from the MacA-MacB-TolC assembly was used, the periplasmic domain present as part of MacB was removed as seen in figure 34. As MacB is a homodimer, one protomer was aligned with one protomer of ABCG2. The alignment of the TMD between 5NIK and ABCG2, the conserved residues mentioned are referred from appendix 22 and table 12 represents the summary for the conserved residues in the alignment.

Similar to the alignment of 5LIL to ABCG2, there were no defined TM helices alignment of 5NIK TMD and ABCG2 TMD as seen in figure 35, however there was alignment of the NBD (appendix 23). When there was just the alignment ABCG2 TMD and 5NIK TMD there were some alignments of the TM helices; noticing that structurally 5NIK has 4 transmembrane which aligns with ABCG2 TM1-4, TM5 and TM6 were removed from ABCG2 and the alignment between ABCG2 TM1-4 and 5NIK TM were made. Figure 36 showed significant alignment between ABCG2 TM1a and the connective helix of 5LIL. Through analysis the structural sequence alignment showed the 2 fully conserved residues and 16 weak conserved groups between residues 630-660. The alignment of ABCG2 TM1b and 5NIK TM1 is similar to the alignment of 5WS4 TM1 and ABCG2 TM1b, where 5NIK TM1 and ABCG2 TM1b is seen to not align and diverge away from each other as seen in figure 36B; the structural sequence alignment showed only 9 weak conserved residues.

There is partial alignment of ABCG2 TM2 and 5NIK TM2 as seen in figure 36B, where both the helices are seen to converge towards the end of the alignment, the structurally sequence alignment showed 1 fully conserved region and 24 weak conserved residues between residues 921-970.

5NIK TM3 and ABCG2 TM3 is seen to align the best, similar to the alignment of ABCG2 TM3 and 5WS4 TM3. The alignment has significant similarity as there is overlap of the helices as seen in figure 36C. The structural sequence alignment showed the highest fully conserved residues compared to TM1,2 and 4; there were 4 fully conserved resides and 30 weak conserved groups between residues 971-1020. Although visually there was

considerable alignment of ABCG2 TM4 and 5NIK TM4 as seen from figure 36C, interestingly structural sequence alignment only showed 22 weak conserved groups between residue 1031-1061. There was significant alignment of the NBD similar to the other alignments made by ABCG2 and 5WS4 and 5LIL; the alignment of 5NIK NBD and ABCG2 NBD showed most of the α helices and β pleated sheets to be aligned well (appendix 23). This is supported by the structural sequence alignment as there were 47 fully conserved residues between residues 1 - 400; the overall RMSD was 1.921, SDM as 38.435, Q score as 0.438 and percentage identity of 16.82% (appendix 24).

Although small divergences were noticed between 5NIK TM1 and ABCG2 TM1b it can be concluded there to be high structural similarity especially as the showed overall RMSD as 2.476, SDM as 56.229, Q score as 0.227 and the percentage identity as 16.82%. Henceforth, these results indicate high structural similarity therefore indicating once again high possibility that ABCG family's originates from MacB transporters.

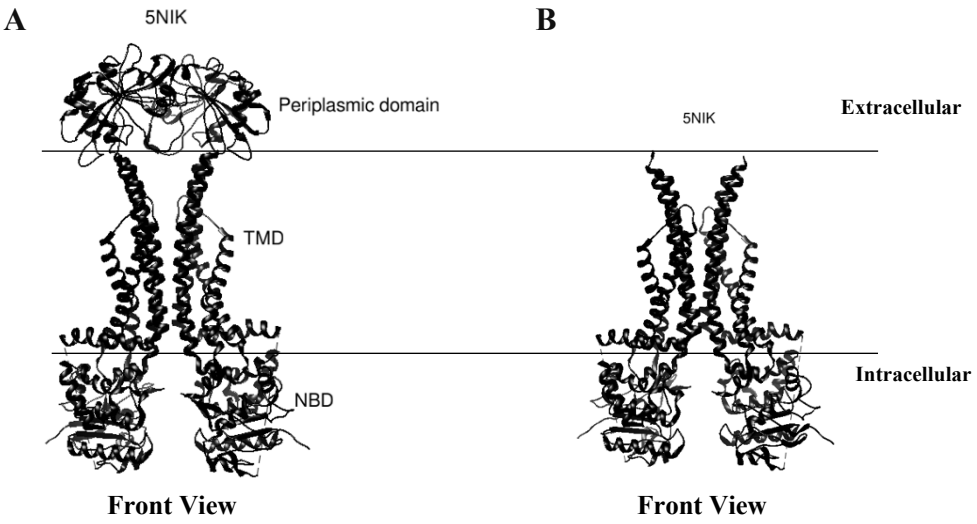


Figure 34:The structure of 5NIK in Chimera VS the structure used in the alignment to ABCA1, ABCG2 and ABCG5/G8

The structure of 5NIK in this figure is represented in black. **A)** *The structure of 5NIK with the presence of periplasmic domain, TMD and NBD.* ***B)*** *The structure of 5NIK used in the alignments of ABCA1, ABCG2 and ABCG5G8.*

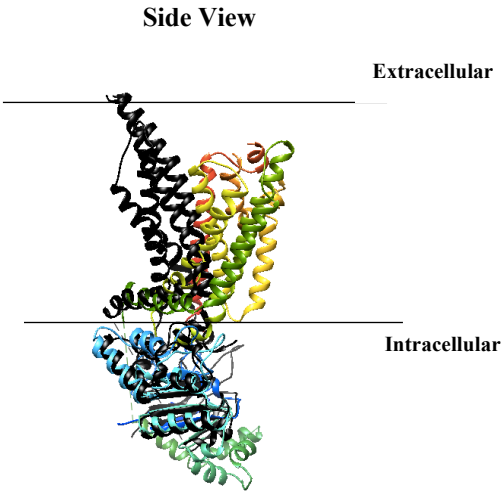


Figure 35: The initial alignment of 5NIK and ABCG2

In this figure, 5NIK is represented in magenta and the ABCG2 transmembrane helices are represented in various colours of the rainbow. The initial alignment between the transmembrane domain of ABCG2 and 5NIK showed the NBD to be aligned well, however the TMD to be distorted in the alignment.

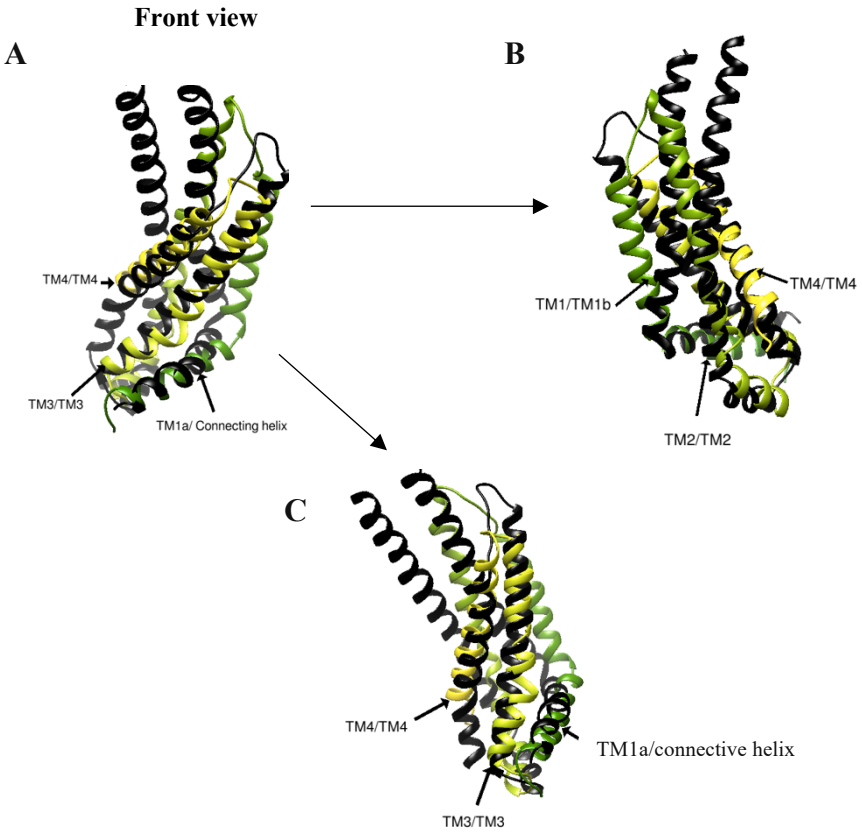


Figure 36:The alignment of 5NIK TMD and ABCG2 TM1-4

The structure of 5NIK is represented in black, structure of ABCG2 is represented in yellow and green. ***A)*** *Front view: the structural alignment between; 5NIK connective helix and ABCG2 TM1a, 5NIK TM3 and ABCG2 TM3,5NIK TM4 and ABCG2 TM4.* ***B)*** *Rotated anticlockwise 90° from front view: the structural alignment between; ABCG2 TM2 and 5NIK TM2, ABCG2 TM1b and 5LIL TM1, ABCG2 TM4 and 5NIK TM4.****C)*** *Rotated clockwise 90° from front view****:*** *the structural alignment between; ABCG2 TM3, 5NIK TM3, ABCG2 TM4 and 5NIK TM4 and 5NIK connective helix and ABCG2 TM1a*

Alignments and residue regions – 5NIK TMD: ABCG2 TM1-4	Fully conserved residues	Weakly conserved residues
Connective Helix:TM1a (630-660)	2	16
TM1b:TM1 (661-920)	0	9
TM2:TM2 (921-970)	1	24
TM3:TM3 (971-1020)	4	30
TM4:TM4 (1031-1061)	0	22

Table 12: Quantitative data for the alignment between ABCG2 TM1-4 and 5NIK TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG2 TM1-4 and 5NIK TMD

4.2.4 The structural alignment between ABCG5/G8 and 5WS4

5WS4 was aligned with ABCG5/G8, 5WS4 aligned better with ABCG8 rather than ABCG5 as seen in figure 37. Therefore, the alignment between ABCG8 and 5WS4 was used to evaluate the potential structural relationship between ABCG8 and mechanotransducer MacB. The following conserved and weak residues mentioned with the alignments between ABCG8 TM1-4 and 5WS4 TMD from appendix 25; the conserved residues mentioned in the alignment of ABCG8 TMD and 5WS4 TMD is summarised in table 13.

The elbow of 5WS4 aligned close to ABCG8 TM1 as seen in figure 38A, thus the corresponding structural sequence alignment showed only 4 weak conserved residues, this showed that the structural similarity is not as high as expected however that may be as a result from both helices not overlapping. The alignment of TM2 from 5WS4 and ABCG8 showed to convey similar structure and direction as seen in figure 38B, furthermore the sequence alignment shows there are 2 full conserved residues and 26 weak conserved residues within the aligned regions.TM3 from both ABCG8 and 5WS4 shows significant alignment as seen in figure 38A; the structural sequence alignment from 1001-1060 showed 3 fully conserved residue and 24 weak conserved residues. From figure 38A, the alignment of TM4 from 5WS4 and ABCG8 showed similarity in the alignment of ABCG2 TM4 and 5WS4 TM4. The structural sequence alignment showed between 1075-1110 there are 2 fully conserved residues and 23 weak conserved groups.

The alignments of 5WS4 TM1-4 and ABCG8 TM1-4 were seen to be better when TM5 and TM6 were not present, which is similar to the previous alignments made between ABCG2 and 5WS4. Although the alignment between ABCG8 TM1 and 5WS4 Elbow was not as significant compared to other TM helices alignment between the two transporters, overall, the results showed the ABCG8 TM1-4 and 5WS4 TMD do have reasonable structural alignment. This is supported by the overall RMSD being 2.408, SDM as 47.909, Q score as 0.388 and the percentage identity being 15.95%.

5WS4 NBD was aligned to ABCG8 NBD and the results showed the 8 α helices and the β pleated sheets to overlap and align well (appendix 26). The structural sequence alignment showed 25 fully conserved residues and 72 weak conserved residues between residues 51-450, reflecting high structural similarity as expected (appendix 27). Furthermore, this was supported by the structural alignment having an over RMSD of 2.268, SDM as 48.933, Q score as 0.381 and a percentage identity of 16.82%. Normally in literature the NBD sequence alignment are suggested 20% or more, however when using chimera some loops external to the area of focus are truncated, as a result the percentage identity is slightly lower. Overall considerable structural alignment between 5WS4 and ABCG8.

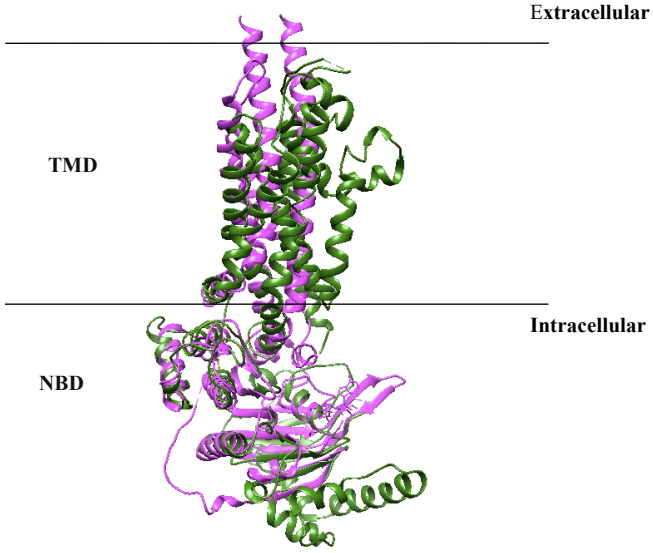


Figure 37:The structural alignment ABCG8 and 5WS4

The structure of 5WS4 is represented in magenta and ABCG8 is represented in forest green. The figure represents the alignment between ABCG8 and 5WS4, where the NBD are seen to align well.

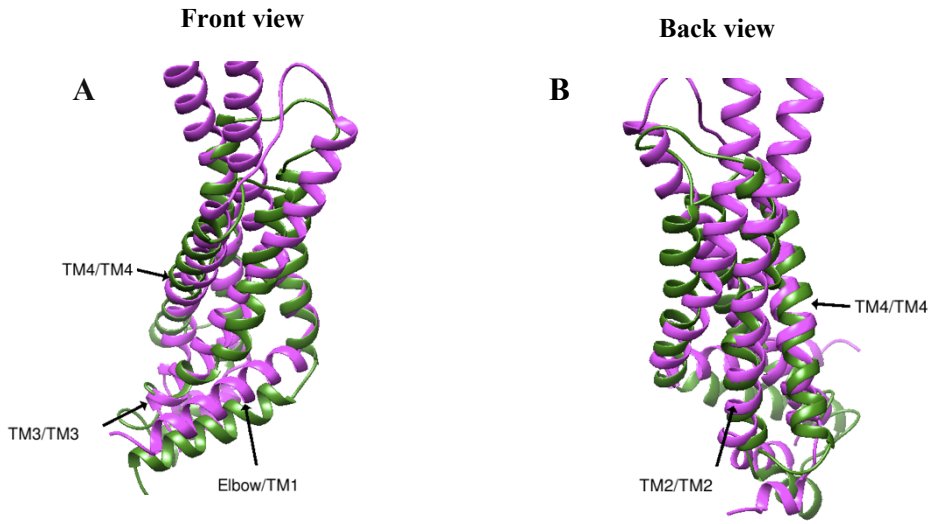


Figure 38: The structural alignment between ABCG8 TM1-4 and 5WS4 TMD

A) *The visible structural alignment between 5WS4 elbow and ABCG8 TM1, ABCG8 TM3 and 5WS4 TM3, 5WS4 TM4 and ABCG8 TM4.* ***B)*** *The structural alignment between 5WS4 TM2 and ABCG8 TM2, 5WS4 TM4 and ABCG8 TM4*

Alignments and residue regions – 5WS4 TMD: ABCG8 TM1-4	**Fully conserved residues**	**Weakly conserved residues**
Elbow:TM1 (660-721)	0	4
TM2:TM2 (733-923)	2	26
TM3:TM3 (1001-1060)	3	24
TM4:TM4 (1075-110)	2	23

Table 13: Quantitative data for the alignment between ABCG8 TM1-4 and 5WS4

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TM1-4 and 5WS4 TMD.

4.2.5 The structural alignment between ABCG5/G8 and 5LIL

5LIL was used as the second MacB structure to look at the structural relationship between ABCG5G8. The alignment was made between 5LIL with no periplasmic domain; the conserved residues mentioned between ABCG2 TMD and 5LIL TMD is referred from appendix 28. Table 14 represents the summary of the conserved residues in the alignment between ABCG2 TMD and 5LIL TMD.

5LIL aligned with ABCG8 rather than ABCG5 similar to how 5WS4 aligned to ABCG8. Figure 39 showed ABCG8 NBD and 5LIL aligned well however the TMD alignment was not as significantly aligned. When the TMD of ABCG8 and 5LIL were aligned, there was alignment of ABCG8 TM1-4 and 5LIL TMD, but there was no alignment of TM5 or TM6. To look closely at the alignments of ABCG8 TM1-4 and 5LIL TMD, TM5 and TM6 were removed.

The results showed some alignment of amphipathic helix and TM1 however it was not significant as seen in figure 40A, this was reflected by the structural sequence alignment with only 10 weak conserved residues between 651-690. In figure 40B, 5LIL TM1 is seen to diverge outwards and away from ABCG8 TM1, hence very little structural alignment is seen, however the structural sequence alignment shows 16 weak conserved residues between 690-741. The alignment between TM2 showed both helices started to converge and align well as seen in figure 40B, this was reflected by the 3 fully conserved residues and 26 weak conserved residues present in the structural sequence alignment between 940-990. ABCG8 TM3 and 5LIL TM3 was seen to align well as seen in figure 40A and 40C, the structural sequence alignment showed 3 fully conserved residues and 22 weak conserved resides between 1001-1040. The best alignment was seen with TM4 structurally as there was a defined similarity of both structures aligning as seen in figure 40C, the structural sequence alignment showed were 4 fully conserved residues and 20 weak conserved residues between 1061-1200. Overall, the size and shaped of ABCG8 TM1-4 and 5LIL TMD is shown to be highly similar as reflected but the overall RMSD being 2.445, SDM was 56.907, Q score was 0.289 and percentage identity of 16.72%. However, the alignment was better alignment between ABCG2 and 5LIL than between ABCG8 and 5LIL. Nevertheless, the overall structural alignment is supportive towards the theory of ABCG families originating from MacB transporter, thus may adopt a more classical transmembrane pumping function rather than the alternative access mechanism.

As expected from previous results, the alignment of ABCG8 NBD and 5LIL NBD showed high similarity especially as there are significance overlap of the α helices and β pleated sheets (appendix 29). This was reflected by the 37 fully conserved residues between 651-1150, an overall RMSD of 2.167, SDM was 47.210, Q score of 0.289 and a percentage identity of 18.72% (appendix 30).

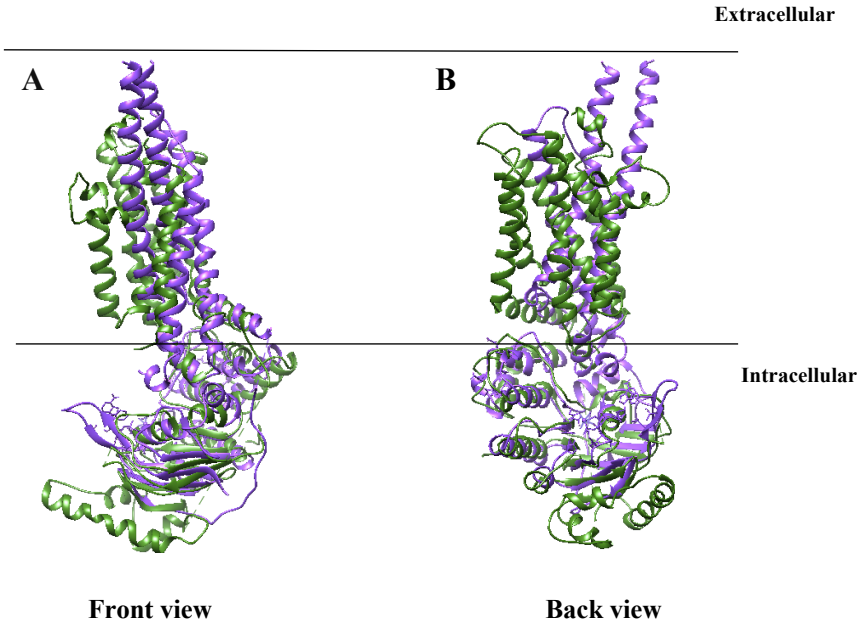


Figure 39: The structural alignment between ABCG8 and 5WS4

The structure of 5LIL is represented in purple and ABCG8 is represented in forest green. The figure represents the alignment between ABCG8 and 5LIL, where the NBD are seen to align well. ***A)*** *The alignment shows the front view, whilst* ***B)*** *represents the side view once the structure is rotated 90°.*

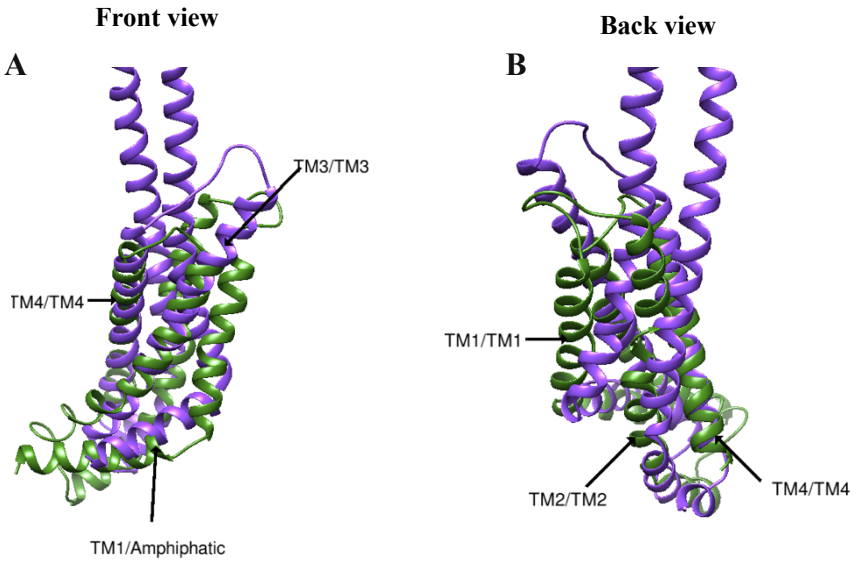


Figure 40: The structural alignment between ABCG8 TM1-4 and 5LIL

A) *The visible structural alignment between: 5LIL amphipathic helix and ABCG8 TM1, ABCG8 TM3 and 5LIL TM3, 5LIL TM4 and ABCG8 TM4.* ***B)*** *The structural alignment between: 5LIL TM1 and ABCG8 TM1, 5LIL TM2 and ABCG8 TM2, 5LIL TM4 and ABCG8 TM4*

Alignments and residue regions – 5LIL TMD: ABCG8 TM1-4	Fully conserved residues	Weakly conserved residues
Amphipathic helix: TM1 (651-690)	0	10
TM1:TM1 (690-741)	0	16
TM2:TM2 (940-990)	3	26
TM3:TM3 (1001-1040)	3	22
TM4:TM4 (1061-1200)	4	20

Table 14: Quantitative data for the alignment between ABCG8 TM1-4 and 5LIL TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TM1-4 and 5LIL TMD.

4.2.6 The structural alignment between ABCG8 and 5NIK

The third structural alignment to look at the relationship between ABCG5/G8 and MacB, ABCG5/G8 was aligned to 5NIK. Similar to the alignment made between ABCG2 and 5NIK, only the MacB component from the MacA-MacB-ToLC assembly, thus the periplasmic domain which is present as part of MacB was removed. The fully conserved and weakly conserved residues mentioned in the alignment of ABCG8 TMD and 5NIK TMD are referred from appendix 31; table 15 represents the summarised data for the conserved residues in the alignment.

Similar to 5WS4 and 5LIL, 5NIK aligned to ABCG8 rather than ABCG5 as seen in figure 41. In the alignment, 5NIK connective helix was seen to fold in the similar direction as ABCG8 TM1 as seen in figure 42A, however there were no overlap of the helices and this was reflected in the structural sequence alignment as there is only 16 weak conserved residues present between residues 650-671. There was significant alignment of 5NIK TM2 and 5NIK TM2 as seen in figure 42B, hence, the structural sequence alignment showed 2 fully conserved residues and 13 weak conserved residues present between residues 950-990. Initially at the start of both TM3 there was some alignment, however there was some divergence towards the end of the alignment as seen in figure 42B, the structural sequence alignment showed there to be 2 fully conserved residues and 16 weak conserved groups between residues 1001-1040. The alignment of TM4 was seen to be the best compared to TM1,TM2 and TM3 as seen in figure 42A, this was reflected in the structural sequence

alignment where there was 4 fully conserved residues and 22 weak conserved residues between 1061-1093. Overall, there was high structural similarity between ABCG8 and 5NIK, all the transmembrane helices were seen to align well, this was reflected as the overall RMSD was 2.452, SDM was 58.425, Q score was 0.286 and percentage identity was 16.82%.

There was significant alignment of the 5NIK NBD similar to the other alignments made by ABCG8 to 5WS4 and 5LIL. The alignment of 5NIK NBD and ABCG5 NBD showed most of the α helices and β pleated sheets to be aligned well (appendix 32), the structural sequence alignment shows there to be 55 fully conserved residues between residues 1- 450 with an overall RMSD of 2.196, SDM was 42.937, Q score of 0.398 and percentage identity of 17.13% (appendix 33). As there is significant alignment of 5NIK and ABCG8, this further supports that ABCG families may have originated from MacB transporters.

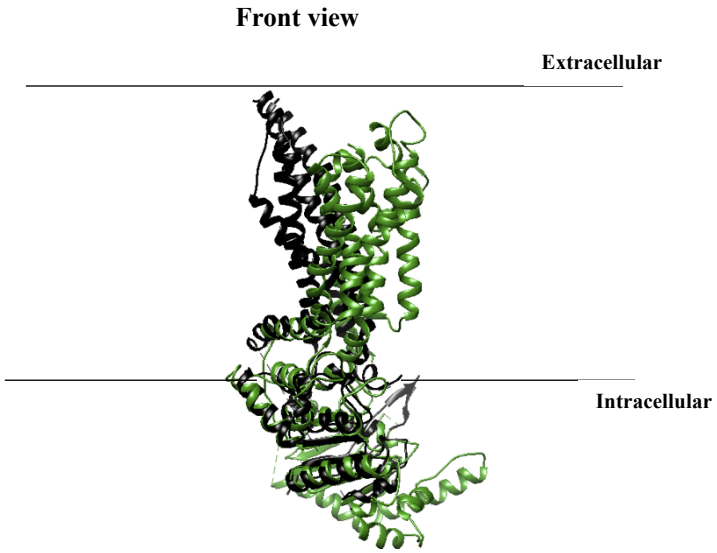


Figure 41: The structural alignment between ABCG8 and 5NIK

The structure of 5NIK is represented in black and ABCG8 is represented in forest green. The figure represents the alignment between ABCG8 and 5NIK, where the NBD are seen to align well.

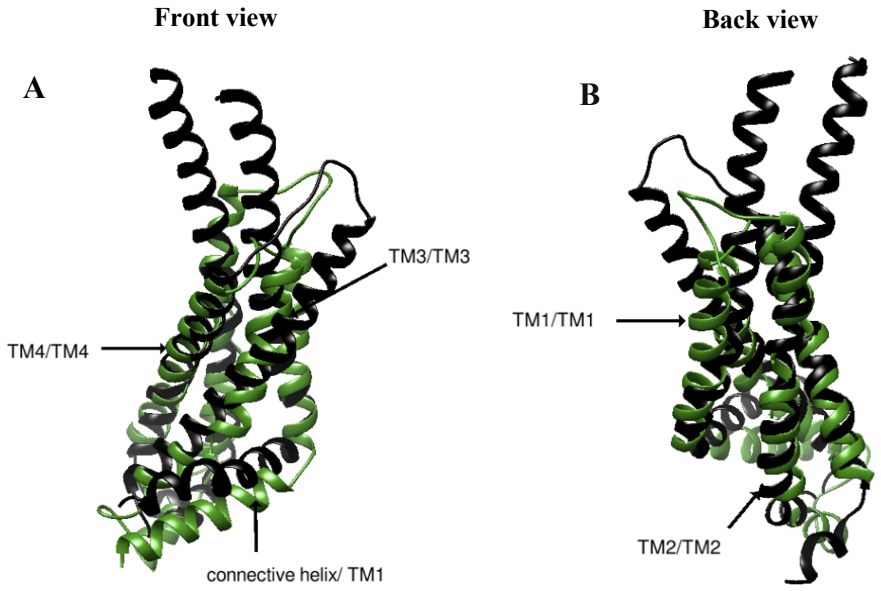


Figure 42:The structural alignment between ABCG8 TM1-4 and 5NIK TMD

A) *The visible structural alignment between: 5NIK connective helix and ABCG8 TM1, ABCG8 TM3 and 5NIK TM3, 5NIK TM4 and ABCG8 TM4.* ***B)*** *The structural alignment between: 5NIK TM1 and ABCG8 TM1, 5NIK TM2 and ABCG8 TM2*

Alignments and residue regions – 5NIK TMD: ABCG8 TM1-4	**Fully conserved residues**	**Weakly conserved residues**
Connective helix: TM1 (650-671)	0	16
TM2:TM2 (950-990)	2	13
TM3:TM3 (1001-1040)	2	16
TM4-TM4 (1061-1093)	4	22

Table 15: Quantitative data for the alignment between ABCG8 TM1-4 and 5NIK TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TM1-4 and 5NIK TMD

4.2.7 The structural alignment between ABCA1 and 5WS4

As mentioned previously, Ford *et al* suggested there is a possibility that both ABCA and ABCG family both might have evolved from the same ancestor and bacterial mechanotransducers MacB may have been this ancestor as. Therefore, to find whether ABCA1 relationship with bacterial mechanotransducers MacB the alignment between 5WS4 and ABCA1 was determined. The fully conserved and weakly conserved residues mentioned in the alignment of ABCA1 TMD1 and 5WS4 TMD is referred from appendix 34; table 16 represents a summary of the quantitative data of this alignment.

The results showed that one protomer of 5WS4 TMD NBD aligned with ABCA1 TMD1 NBD rather than ABCA1 TMD2 and NBD2 as seen in figure 43. The alignment of 5WS4 TM and ABA1 TM showed that ABCA1 IH1 aligned with 5WS4 TM 1 as seen in figure 44A, the sequence alignment showed that within that aligned region there were 2 fully conserved residues and 16 weak conserved residues between 291-331.

TM2 from both ABCA1 and 5WS4 showed to align well as it conveyed similar structure as seen in figure 44B. The structural sequence alignment showed within the aligned region (from 1171-1211 residue) there were 2 fully conserved regions and 30 weak conserved residues. The alignment of TM3 for both ABCA1 and 5WS4 showed to have similar structure as seen in figure 44A, thus, the structural sequence alignment showed to have 8 fully conserved and 30 weak conserved residues between 1221-1260.

Similar to the alignment of TM3, TM4 from both ABCA1 and 5WS4 showed significant alignment as seen in figure 44A. The structural sequence alignment showed between residues 1271-1321 there were 23 weak conserved residues. TM5 and TM6 were shown to have no alignment with any of the transmembrane helices of 5WS4, which is similar to the alignment made between ABCG2 and 5WS4; the attempt of removing TM5 and TM6 of ABCA1 were made in order to get the alignment of just ABCA1 TM1-4 and 5WS4 TM1-4, however chimera concluded that less than 3 residues present to match. Although chimera stated there were 3 resides to match it was evident that 5WS4 TM1-4 and ABCA1 TM1-4 aligned significantly as the overall RMSD was 2.330, SDM was 58.922, Q score of 0.156; all of which indicating high structural. Furthermore, this suggests a structural evolutionary relationship between ABCA1 and MacB, hence, further implicating both ABCA and ABCG families both originate from a MacB transporter.

ABCA1 NBD and 5WS4 NBD was shown to align well (appendix 35), where 8 of the α helices aligned well between 5WS4 and ABCA1, the structural sequence alignment showed between residues 951-1200 there were 48 fully conserved residues alongside 155 weak conserved residues (appendix 36); high structural similarity was implicated as the overall RMSD was 2.186, SDM was 47.986, Q score was 0.390. Clustal Omega was used to display the percentage identity, as on chimera ABCA1 is a single chain the percentage identity of the TMD and NBD was not worked out separately in the alignment with 5WS4; the percentage identity of entire ABCA1 and 5WS4 structure was 21.20%.

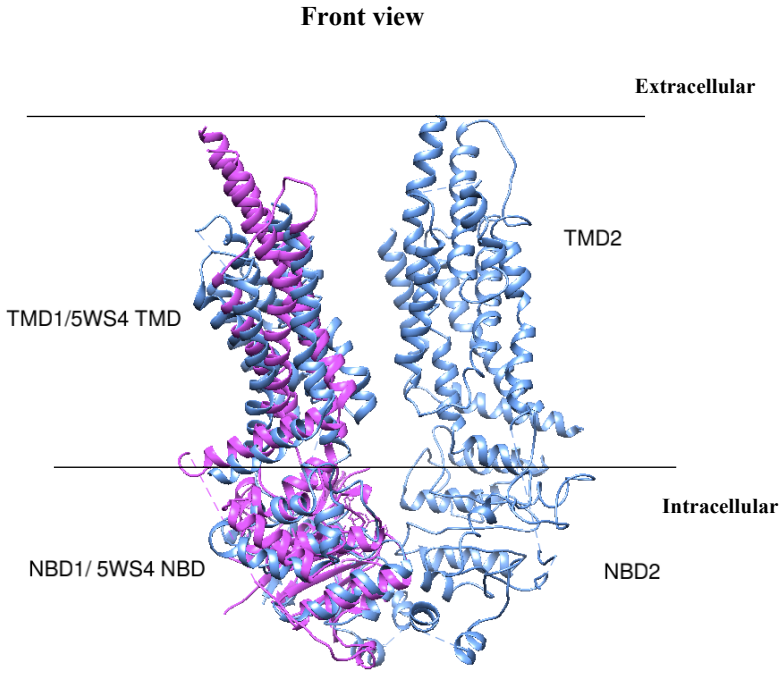


Figure 43:The structural alignment between 5WS4 and ABCA1

The structure of 5WS4 is represented in magenta and ABCA1 is represented in cornflour blue. The figure shows 5WS4 aligning with ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 NBD2.

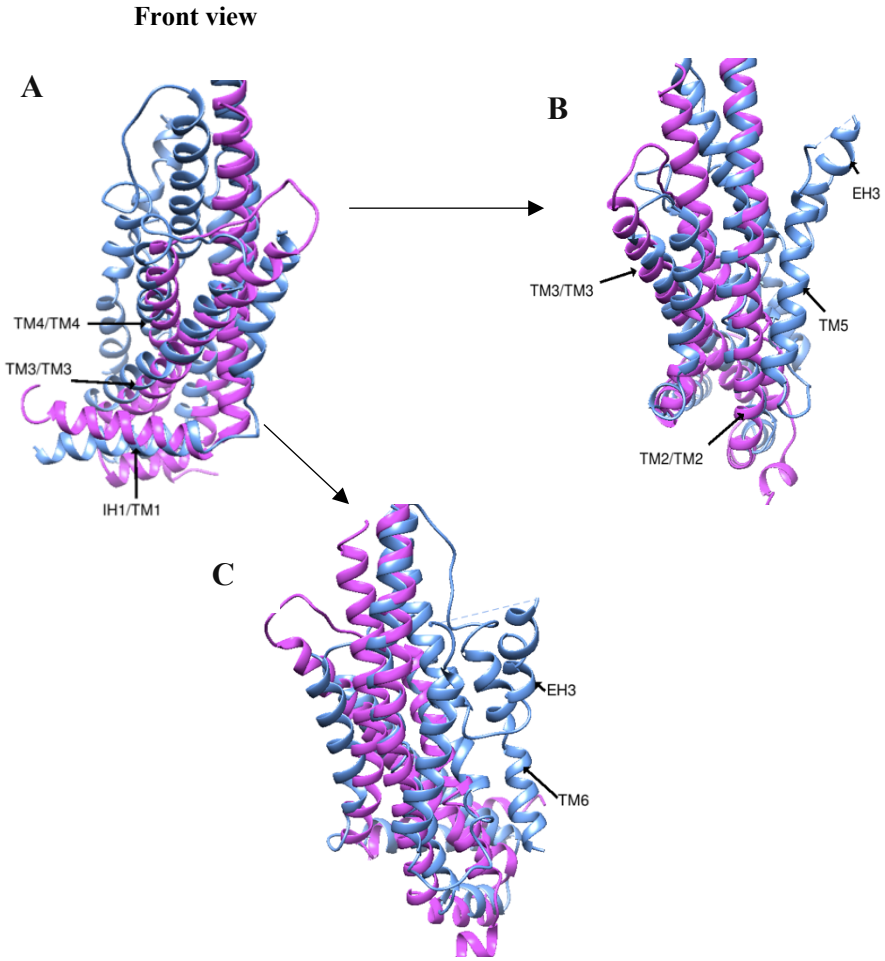


Figure 44:The structural alignment between ABCA1 TMD1 and 5WS4 TMD

The structure of ABCA1 NBD is represented in cornflour blue and the structure of 5WS4 NBD is represented in magenta ***A)*** *Front view: The visible structural alignment between 5WS4 elbow and ABCA1 IH1, ABCA1 TM3 and 5WS4 TM3, 5WS4 TM4 and ABCG8 TM4.* ***B)*** *Rotated clockwise 90° from front view: the structural alignment between 5WS4 TM2 and ABCA1 TM2, 5WS4 TM3 and ABCA1 TM3.* ***C)****Rotated anticlockwise 90°from front view: no alignment of TM5 and TM6.*

Alignments and residue regions – 5WS4 TMD: ABCA1 TM1-4	**Fully conserved residues**	**Weakly conserved residues**
IH1:TM1 (291-331)	2	16
TM2:TM2 (1171-1211)	2	30
TM3:TM3 (1221-1260)	8	30
TM4:TM4 (1271-1321)	0	23

Table 16: Quantitative data for the alignment between ABCA1 TM1-4 and 5WS4 TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TM1-4 and 5WS4 TMD

4.2.8 The structural alignment between ABCA1 and 5LIL

5LIL was the second structure used to look at the structural evolutionary relationship between ABCA1 and MacB. Similar to the alignment made between ABCG2 and 5LIL, the periplasmic domain of 5LIL was removed and one protomer of it was aligned. The results showed that 5LIL aligned with ABCA1 TMD1 NBD1 rather than ABCG2 TMD2 NBD2 as seen in figure 45. The NBD was seen to align better than the TMD in the structural alignment between ABCG2 and 5LIL. In the attempt to look closely at the detailed alignment between ABCG2 TMD and 5LIL TMD, the alignment between just the TMD were made. Surprisingly, according to chimera there were less than 3 residues able to align. Therefore, looking at the comparison of the TM helices was not made. However, ABCA1 NBD1 and 5LIL NBD showed significant alignment (appendix 37), where the α helices and β pleated sheets aligned well. This was reflected by the 40 fully conserved residues between residues 901- 1250 (appendix 38). To support the high structural similarity the overall RMSD 2.004, SDM was 41.460, Q score was 0.435 and the percentage identity was 20.33%. Although there were difficulties obtaining the detailed structural alignment of ABCA1 TMD and 5LIL TMD, the structural alignment seen in figure 45 shows overall shape and size to be significant. As there was considerable alignment between ABCA1 NBD1 and 5LIL NBD, therefore this suggest ABCA1 originated from MacB transporters.

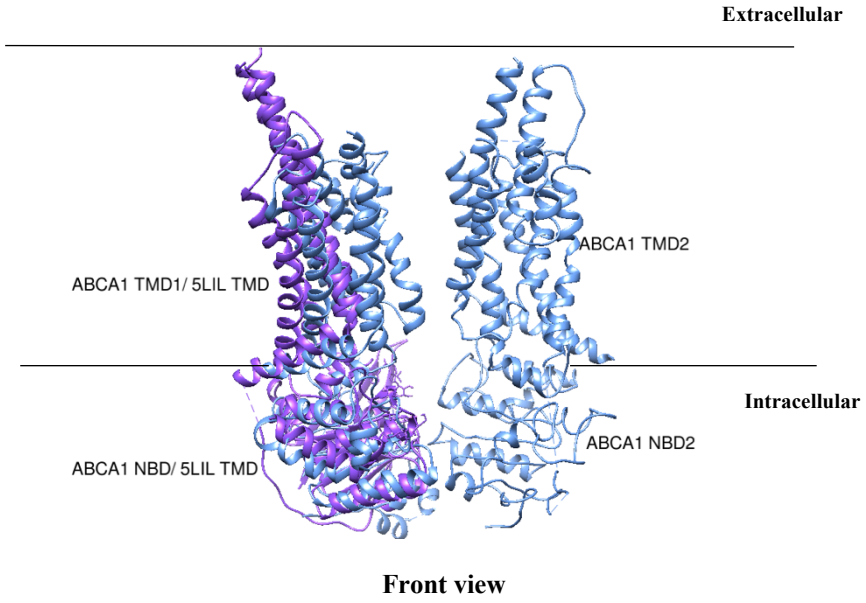


Figure 45: The structural alignment between ABCA1 and 5LIL

The structure of ABCA1 is represented in cornflour blue and 5LIL is represented in purple. This figure illustrates 5LIL aligned to ABCA1 TMD1 NBD1 rather than TMD2 and NBD2.

4.2.7 The structural alignment between ABCA1 and 5NIK

The third structure that ABCA1 was aligned with to evaluate the potential evolutionary relationship between ABCA1 and MacB was 5NIK. Similar to the alignment made with ABCG2, only the MacB component from the MacA-MacB-TolC assembly was used, hence the periplasmic domain present as part of MacB was removed. Furthermore, the alignment of the TMD between 5NIK and ABCA1, the conserved residues are referred from appendix 39, hence table 17 represents the summarised quantitative data for this alignment.

The results showed that one protomer of 5NIK TMD NBD aligned with ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 NBD2 as seen in figure 46. The alignment of 5NIK TMD and ABCA1 TMD1 showed that ABCA1 IH1 aligned with 5NIK connective helix as seen in figure 47A, the structural sequence alignment showed that within the alignment region (residues 271-300) there were only 18 weak conserved residues.

The alignment of 5NIK TM2 and ABCA1 TM2 shown in figure 47C conveys similar structure, the structural sequence alignment showed there to be 1 fully conserved residue and 30 weak conserved residues between residues 1151-1190. ABCA1 TM3 and 5NIK TM3 showed to have the best alignment compared to the alignment of TM1, TM2 and TM4. The alignment showed the helix conveying similar topology, this is reflected in the structural sequence alignment where there were 3 fully conserved residues and 20 weak conserved residues between residues 1195-1220. Visually the alignment of 5NIK TM4 and ABCA1 TM4 was seen to have significant alignment as seen in figure 47B, however the structural sequence alignment showed the presence of 2 fully conserved and 18 weak conserved residues between 1262-1297. There was significant alignment of the NBD, similar to the other alignments made by ABCA1 NBD to 5LIL, 5WS4; in the alignment of 5NIK NBD and ABCG2 NBD most of the α helices and β pleated sheets are aligned well (appendix 40), the structural sequence alignment showed 42 fully conserved residues between 901-1200 (appendix 41). Overall, the alignment between ABCA1 and 5NIK to convey very similar structures which further indicates ABCA families to have originated from MacB transporters.

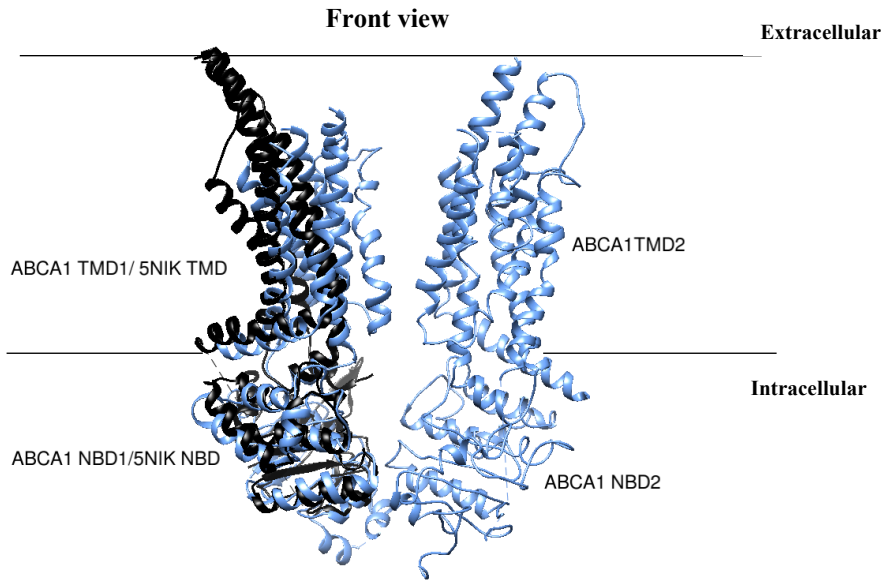


Figure 46: The alignment between ABCA1 and 5NIK

The structure of ABCA1 is represented in cornflour blue and 5NIK is represented in black. This figure illustrates, 5NIK aligned to ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 and NBD2.

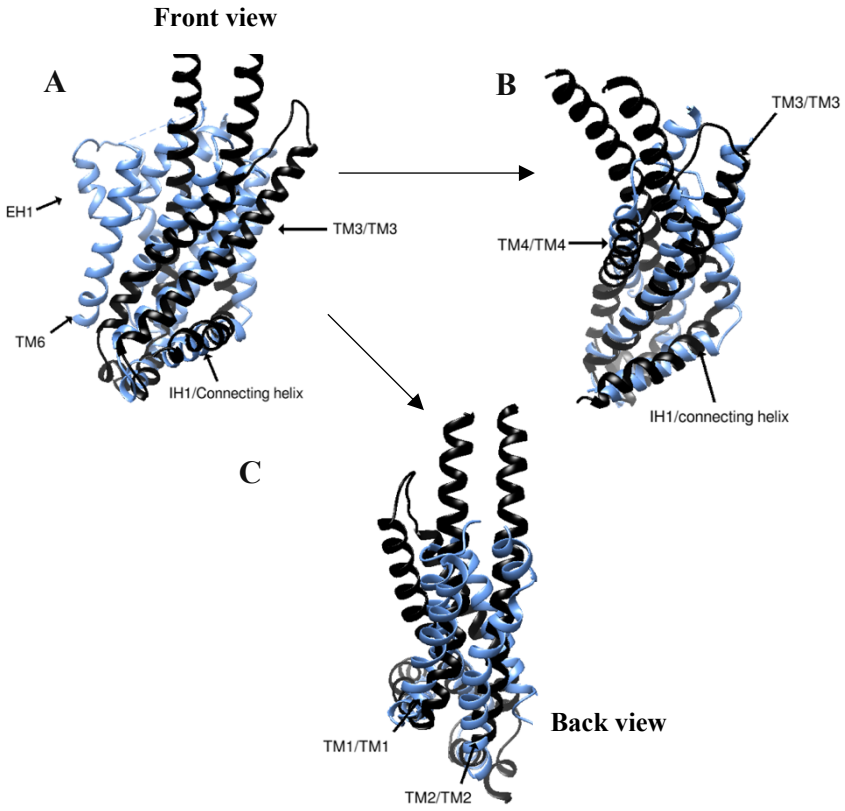


Figure 47: The structural alignment between ABCA1 TM1-4 and 5NIK TMD

The structure of ABCA1 TMD1 is represented in cornflour blue and the structure of 5NIK TMD is represented in black ***A)*** *Front view: the visible structural alignment between 5NIK connective helix and ABCA1 IH1, ABCA1 TM3 and 5NIK TM3, 5NIK TM4 and ABCA1 TM4.* ***B)*** *Alignment rotated rotation by clockwise 90° from the front view the structural alignment between: 5NIK TM4 and ABCA1 TM4, 5NIK connective helix and ABCA1 IH1.* ***C)****Back view****:*** *the structural alignment between: 5NIK TM1 and ABCA1 TM1, 5NIK TM2 and ABCA1 TM2.*

Alignments and residue regions – 5NIK TMD: ABCA1 TM1-4	Fully conserved residues	Weakly conserved residues
IH1:TM1 (271-300)	0	18
TM2:TM2 (1151-1190)	1	30
TM3:TM3 (1195-1220)	3	20
TM4:TM4 (1262-1297)	2	18

Table 17: Quantitative data for the alignment between ABCA1 TM1-4 and 5NIK TMD

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TM1-4 and 5NIK TMD

4.3 The structural relationship between MlaE and MlaF to ABCA and ABCG families

4.3.1 The structural alignment between ABCG2 with MlaE and MlaF

Recently in 2020 there has been a high-resolution structure of MlaFEDB determined by Chi *et al*, where MlaE represents the TMD and MlaF represents the NBD. The structural alignment of the TMD and NBD would shed some light into any evolutionary relationship that may exist between MlaE and MlaF to ABCA and ABCG families.

The last 4 alignments with bacterial ABC transporters made with ABCG2 have had 4 transmembrane helices as part of their TMD, however, MlaE has 5 transmembrane, one periplasmic helix (PH), one coupling helix and one EH which is attached to MlaE TM1 running parallel to the inner membrane plane as seen in figure 48. The EH is very similar to the connecting helix of human ABC transporter (Chi *et al*., 2020). Although EH may seem to have structural similarities to connecting helix, the alignment of MlaE TM1-5 and ABCG2 TM1-6 showed the presence of the EH helix completely distorted the alignment of the TM helices. Once the removal of the EH helix, the alignment of ABCG2 TM1-6 and MlaE TM1-5 was better however not as prominent. There was no alignment of the ABCG2 TM6 with any MlaE transmembrane helices. Furthermore, in the following alignments the conserved residues mentioned in the alignment between ABCG2 TM1-5 and MlaE TM1-5 is referred from appendix 42; hence table 18 represents the quantitative summarised data for this alignment.

Removing ABCG2 TM6 and the EH helix from MlaE showed significant alignment between MlaE TM1-5 and ABCG2 TM1-5 as seen in figure 49A and 49B. The first alignment between MlaE TM 1-5 and ABCG2 TM1-5 showed ABCG2 TM1a and MlaE TM1 were well aligned as seen in figure 49A, the structural sequence alignment showed 1 fully conserved residue and 12 weak conserved residues present between residue 421-440. ABCG2 TM1b and MlaE TM1 was seen to align in the beginning but the helices started to diverge towards the end, within the structural sequence alignment there were only fully conserved residue and 16 weak conserved residues. In figure 49B, the beginning of the alignment of ABCG2 TM2 and MlaE TM2, initially both helices were seen to converge from middle to the end of the alignment. The structural sequence alignment reflects there to be higher sequence alignment from middle of the alignment between residues 480-530 hence within the region there were 3 fully conserved region and 14 weak conserved residues.

In the alignment between ABCG2 TM3 and MlaE TM3 it was seen that both helices converge towards each other in the alignment, both the helices did not overlap significantly compared to the other TM3 alignment of the bacterial ABC transporters however there are 4 fully conserved and 24 weak conserved residues between 531-585. The alignment of ABCG2 TM4 and MlaE TM4 was by far the best aligned transmembrane within the alignment of ABCG2 TM1-5 and MlaE TM1-5 as seen in figure 49A; the structural sequence alignment showed there to be 2 fully conserved residues and 31 weak conserved residues.

In figure 49B, there was some initial alignment between ABCG2 TM5a and MlaE TM5, however both helices were seen to diverge away from each other, this was supported by the structural sequence alignment where only having 12 weak conserved residues between residues 661-681. Overall, there was considerable alignment between the transmembrane helices of ABCG2 TM1-5 and MlaE, especially as the percentage identity was 16.92%. The structural similarity between ABCG2 TM1-5 and MlaE suggest that convergent evolution may be taking place.

The alignment of the ABCG2 NBD and MlaF showed the β- pleated sheets and α helcies to be aligned well as expected (appendix 43). High structural similarity was seen with this alignment as Clustal omega sequence alignment showed 19.70 % percentage identity (appendix 44).

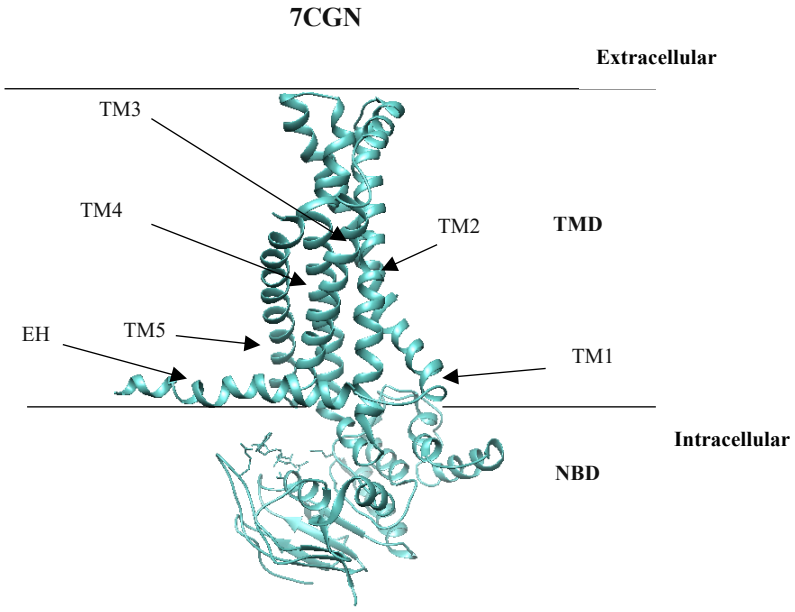


Figure 48: The structure of MlaE and MlaF

The structure of MlaE is represented in light sea green. This figure shows MlaE to have 5 transmembrane, one periplasmic helix (PH), one coupling helix and one EH which is attached to MlaE TM1 running parallel to the inner membrane plane

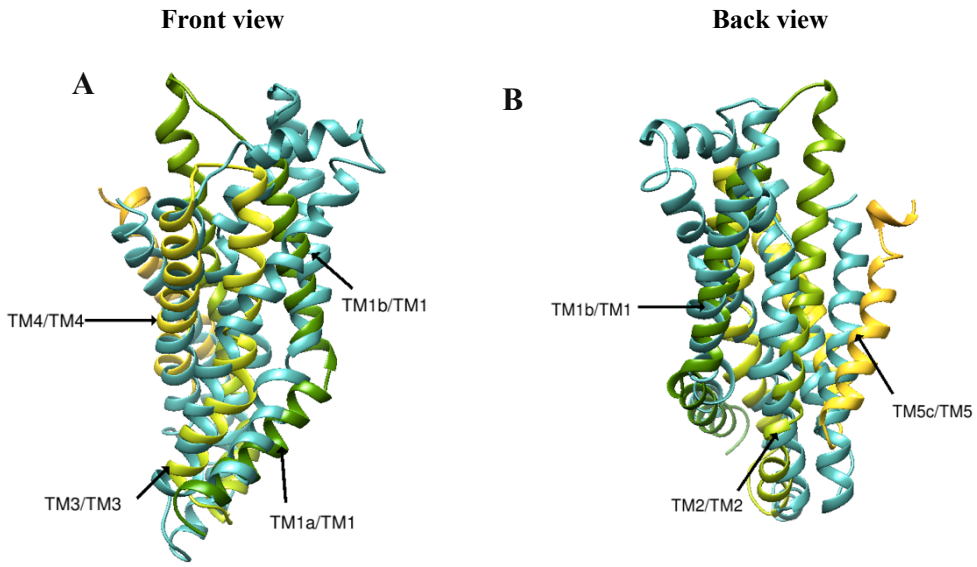


Figure 49:The structural alignment between ABCG2 TM1-5 and MlaE

The structure of MlaE is represented in light sea green. ABCG2 transmembrane helices represent in green, and yellow ribbons. ***A)*** *Front view: the structural alignment between: ABCG2 TM1a and MlaE TM1, ABCG2 TM1b and MlaE TM1, ABCG2 TM3 and MlaE TM3, ABCG2 TM4 and MlaE TM4.* ***B)*** *Back view: the structural alignment between: ABCG2 TM1b and MlaE TM1, ABCG2 TM2 and MlaE TM2, ABCG2 TM5a and MlaE TM5.*

Alignments and residue regions – MlaE: ABCG2 TM1-5	Fully conserved residues	Weakly conserved residues
TM1a:TM1 (421-440)	1	12
TM1b:TM1 (441-479)	1	16
TM2:TM2 (480-530)	3	14
TM3:TM3 (531-585)	4	24
TM4:TM4 (630-660)	2	31
TM5a:TM5 (661-681)	0	12

Table 18: Quantitative data for the alignment between ABCG2 TM1-5 and MlaE

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG2 TM1-5 and MlaE.

4.3.2 The structural alignment between ABCG5/G8 with MlaE and MlaF

The MlaE used in the alignment is the same as the one used in the alignment with ABCG2, hence the MlaE EH helix was removed when making the alignment of ABCG8 TMD and MlaE. Furthermore, the conserved residues in the alignment of ABCG8 TMD and MlaE TMD is referred from appendix 45, thus table 19 represents the summarised quantitative data from this alignment.

MlaE showed to be aligned with ABCG8 rather than ABCG5. Once ABCG8 TMD and MlaE were aligned the structural alignment showed ABCG8 TM1 and MlaE TM1 to be aligned well as seen in figure 50A. The structural sequence alignment of ABCG8 TM1 and MlaE TM1 showed 3 fully conserved and 10 weak conserved residues. ABCG8 TM2 and MlaE TM2 were seen to converge towards each other and have significant structural alignment. The structural sequence alignment showed 3 fully conserved residues and 20 weak conserved residues between residues 460-483. ABCG8 TM3 and MlaE TM3 were also seen to align significantly well as seen in figure 50C, this was reflected in the structural sequence alignment where there are 2 fully conserved and 22 weak conserved residues between 500-620. Although ABCG5 TM4 and MlaE TM4 showed visually the best alignment, the structural sequence alignment showed 24 weak conserved residues between residues 631-670, which was lower than expected.

The structural alignment of ABCG8 TM5 and MlaE TM5 was structurally aligned better compared to the alignment between ABCG2 TM5 and MlaE TM5, however the ABCG8 TM5 and MlaE TM5 is seen to diverge towards the end. The structural sequence alignment showed 3 fully conserved and 8 weak conserved residues between residues 681- 710 for ABCG5 TM5 and MlaE TM5. The alignment of the ABCG8 NBD and MlaF showed the β-pleated sheets and α helcies to be aligned well as expected (appendix 46). High structural similarity was seen with this alignment as Clustal omega sequence alignment showed 23.05% percentage identity (appendix 47). Overall, the results showed great structural alignment between ABCG8 to MlaE and MlaF, indicating convergence structural evolution taking place within the ABCG family.

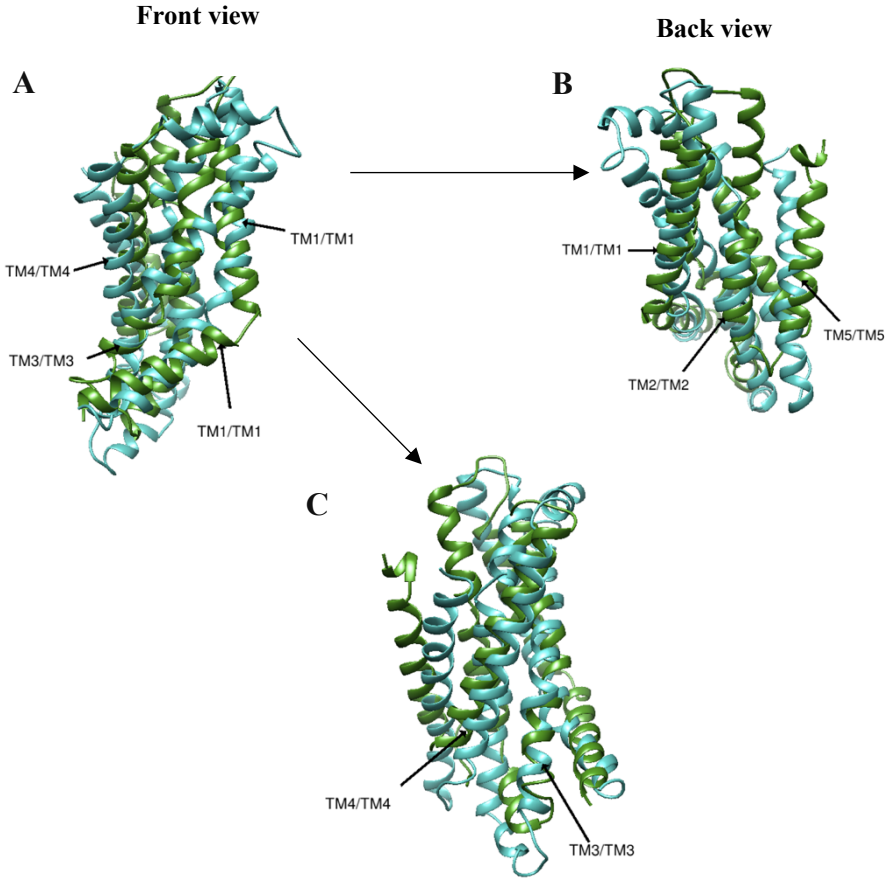


Figure 50:The structural alignment between ABCG5/G8 TM1-5 and MlaE

The structure of MlaE is represented in light sea green. ABCG8 transmembrane helices represent in forest ribbons. ***A)****Front view: The structural alignment between: ABCG8 TM1 and MlaE TM1, ABCG2 TM3 and MlaE TM3, ABCG2 TM4 and MlaE TM4* ***B)****Back view: the structural alignment between: ABCG8 TM1 and MlaE TM1, ABCG8 TM2 and MlaE TM2, ABCG8 TM5 and MlaE TM5.****C)*** Rotated clockwise by 300° from the front view *the structural alignment between: ABCG8 TM3 and MlaE TM3, ABCG2 TM4 and MlaE TM4.*

Alignments and residue regions – MlaE: ABCG8 TM1-5	Fully conserved residues	Weakly conserved residues
TM1:TM1 (400-459)	3	10
TM2:TM2 (460-483)	3	20
TM3:TM3 (490-500)	2	22
TM4:TM4 (631-670)	0	24
TM5a:TM5 (681-710)	3	8

Table 19: Quantitative data for the alignment between ABCG8 TM1-5 and MlaE

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TM1-5 and MlaE.

4.3.3 The structural alignment between ABCA1 with MlaE and MlaF

Similar to the other alignment made with MlaE, the EH helix was removed make better alignment with ABCA1 TM helices. The conserved residues mentioned in the alignment of ABCA1 TMD1 and MlaE is referred from appendix 48; hence table 20 represents the quantitative data for this alignment.

In the initial alignment made between ABCA1 and MlaE, MlaE aligned to ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 NBD2 as seen in figure 51. When the transmembrane domains were aligned, ABCA1 IH1 and MlaE TM1 were seen to align not as well as ABCA1 TM1 and MlaE TM1, between ABCA1 IH1 and MlaE TM1 there was some distance in the alignment as seen in figure 52A. The structural sequence alignment showed 1 fully conserved and 4 weak conserved residues between residues 21-90 for the alignment of ABCA1 IH1 and MlaE TM1. Furthermore, the alignment of ABCA1 TM1 and MlaE TM1 was better compared to the alignment of ABCA1 IH1 and MlaE TM1 as seen in figure 52B, hence was reflected in the structural sequence alignment where there were 2 fully conserved residues and 17 weak conserved residues between residues 91-115.

In the structural alignment of ABCA1 TM2 and MlaE TM2 there was a gap at the beginning, however both transmembrane helices were seen to converge and take similar form and structure as seen in figure 52B. The structural sequence alignment showed 2 full and 31 weak conserved residues between residues 721-706. Similar to ABCA1 TM2 and MlaE TM2 alignment, ABCA1 TM3 and MlaE TM3 were seen to align significantly well as seen in

figure 52C. The structural sequence alignment showed 4 fully conserved and 3 weak conserved residues between residues 741-789. The alignment of ABCA1 TM4 and MlaE TM4 was seen to align well as seen figure 52A. Furthermore, the structural sequence alignment shows there to be 3 fully conserved and 8 weak conserved residues between residues 780-900. The structural alignment of MlaE TM5 and ABCA1 TM5 showed to not align as well compared to the alignments of MlaE TM5 to ABCG2 TM5A and ABCG5 TM5. There was some overlap of the helices seen at the beginning of the alignment, but both helices were seen to diverge away from each other; the structural sequence alignment showed 1 fully conserved and 11 weak conserved residues between residues 950-1050. The alignment of the ABCA1 NBD and MlaF showed the β- pleated sheets and α helices to be aligned well as expected (appendix 49). High structural similarity was seen with this alignment as Clustal omega sequence alignment showed 21.29% percentage identity (appendix 50).

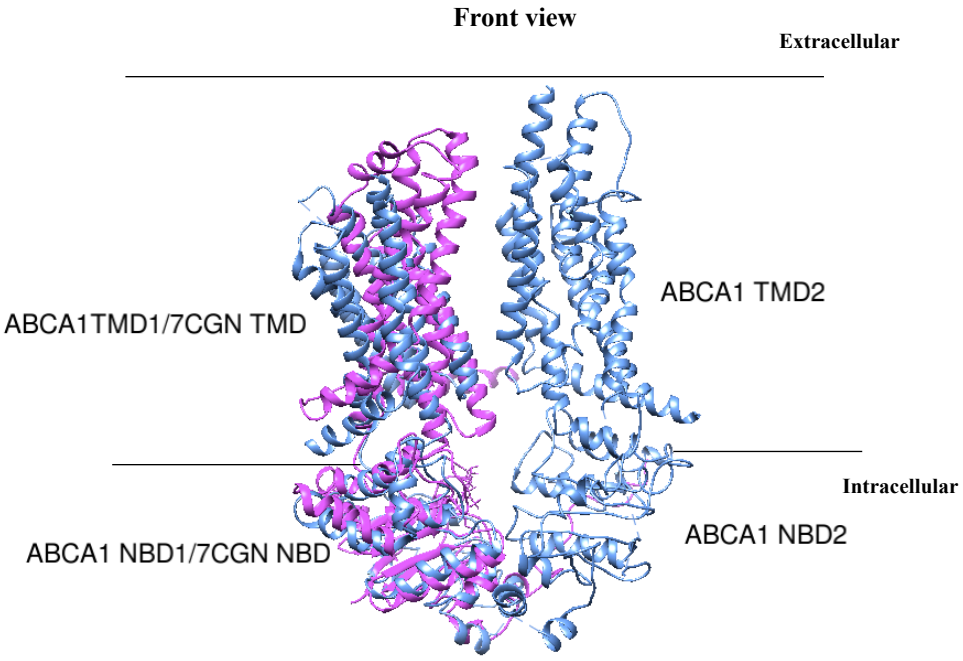


Figure 51:The structural alignment between ABCA1 and MlaE and MlaF

The structure represented in magenta is MlaE, structure ABCA1 is represented in cornflour blue. This figure illustrates MlaE and MlaF to have better alignment between ABCA1 TMD1 NBD1.

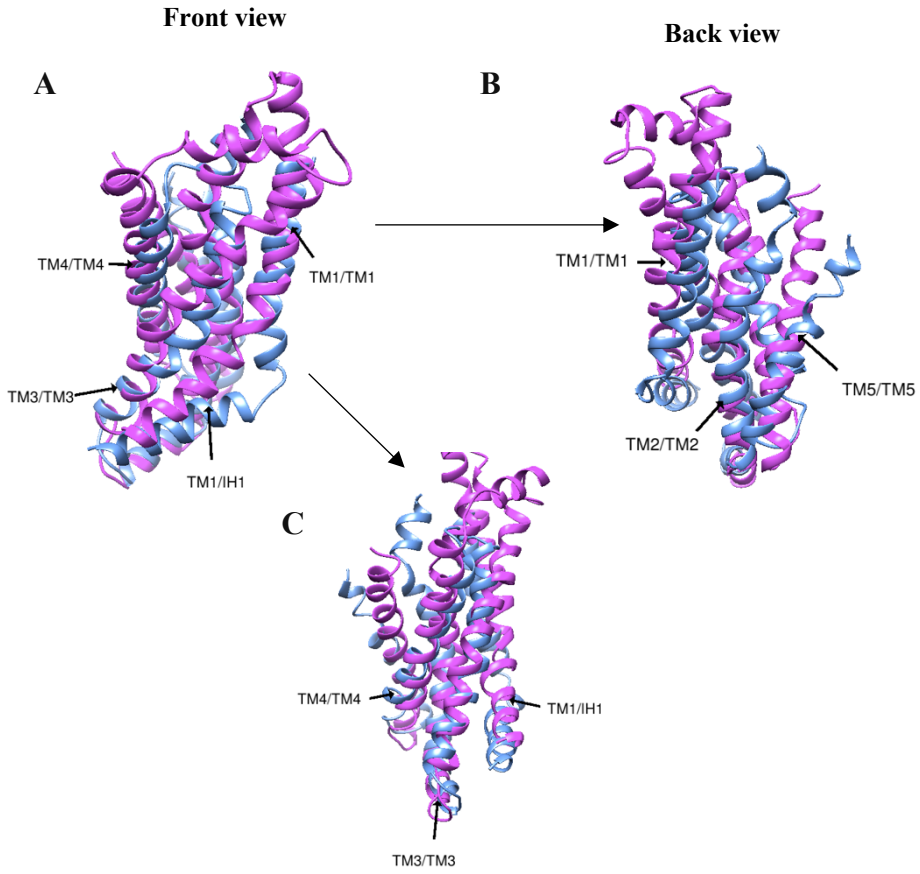


Figure 52: The structural alignment between ABCA1 TM1-5 and MlaE

A) *Front view.* ***B)*** *Back view.* ***C)*** *Rotation by 300 ° from the front view*
The structure of MlaE is represented in magenta and ABCA1 is represented in cornflour blue ribbons ***A,C)*** *The structural alignment between: ABCA1 IH1 and MlaE TM1, ABCA1 TM1 and MlaE TM1, ABCA1 TM3 and MlaE TM3 and ABCA1 TM4 and MlaE TM4.* ***B)*** *The structural alignment between: ABCA1 TM1 and MlaE TM1, ABCA1 TM2 and MlaE TM2 and ABCA1 TM5 and MlaE TM5.*

Alignments and residue regions – MlaE: ABCA1 TM1-5	**Fully conserved residues**	**Weakly conserved residues**
IH1:TM1 (21-90)	1	4
TM1-TM1 (91-115)	2	17
TM2:TM2 (721-736)	2	31
TM3:TM3 (741-789)	4	3
TM4:TM4 (790-900)	3	8
TM5-TM5 (950-1050)	1	11

Table 20:Quantitative data for the analysis of the alignment between ABCA1 TM1-5 and MlaE

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCA1 TM1-5 and MlaE.

4.4 The structural relationship between WzmWzt to ABCA and ABCG families

4.4.1 The structural alignment between ABCG2 and WzmWzt

The WzmWzt ABC transporter also known as the O antigen ABC transporter, unlike the transmembrane domain of MlaE where it contains 5 transmembrane helices, WzmWzt contains 6 transmembrane helices. Although previously WzmWzt was discussed in Ford *et al* paper about the mechanism of transport, the structural similarity and potential evolutionary relationship between WzmWzt to ABCA and ABCG families was not discussed. However, the detailed structural alignment between ABCG2 and WzmWzt may show some links.

The structure for WzmWzt present on chimera contained additional solvents and ligands, these extra solvents and ligands were removed as seen in figure 53. The conserved residues mentioned in the alignment between ABCG2 TMD is referred from appendix 51; table 21 represents the quantitative data for this alignment.

The structural alignment between ABCG2 and WzmWzt showed significant alignment of the TMD and NBD to some degree as seen in figure 54. The transmembrane helices were seen to take similar fold and topology; surprisingly the NBD was seen to not align as well in comparison as seen in figure 54. However, this might be as a result from the angle of the alignment made in chimera.

The alignment between ABCG2 TMD and Wzm gave a percentage identity of 20.31% which reflects a high structural and sequence similarity especially for the alignment of the transmembrane domain. ABCG2 TM1a and Wzm IF showed significant alignment as seen in figure 55a; the structural sequence alignment showed 2 fully conserved residues and 16 weak residues between residues 381-400. ABCG2 TM2 and Wzm TM2 was seen to convey similar structure as it showed good alignment as seen in figure 55B. The structural sequence alignment showed there to be only 23 weak conserved residues and 0 fully conserved residues between residues 491-520. Similar in the alignment of MlaE TM3 and ABCG2 TM3, Wzm TM3 and ABCG2 TM3 showed to converge in the similar direction and fold as seen in figure 103c, the structural sequence alignment indicated 2 fully conserved and 17 weak conserved residues between residues 521 -560. The alignment of TM4 from both ABCG2 and Wzm showed the best structural alignment compared to the other transmembrane helices alignment between Wzm and ABCG2 as seen in figure 55C; the

structural sequence alignment showed 3 full conserved and 26 weak conserved residues between residues 571-610.

The alignment of ABCG2 TM5a and Wzm TM5 is seen to align to some degree as this was reflected in the structural sequence alignment, where there were 1 fully conserved residue between residues 611-630. ABCG2 TM5a and Wzm TM5 are seen to diverge away from each other. Furthermore, there was no significant alignment of ABCG2 TM5b, TM5c and TM6a to any segments of Wzm, however there was some structural similarity and alignment of ABCG2 TM6b to Wzm TM6 as reflected in the structural sequence alignment where there were 1 fully conserved and 16 weak conserved residues between residues 731-781. The ABCG2 NBD and Wzt showed significant alignment where almost all the α helices and β-pleated sheets were aligned well (appendix 52). Furthermore, the NBD showed there 33 fully conserved residues between residues 1-400. Surprisingly, the alignment of ABCG2 NBD and Wzt had a percentage identity of 21.72% (appendix 53).

Overall, the alignment between WzmWzt and ABCG2 showed ABCG2 TM1-4 and Wzm TM1-4 aligned the best, TM5 and TM6 were also seen to align to some degree but not as well as TM1-4. The results conclude that as almost all the segments of ABCG2 and WzmWzt aligned, it can be postulated that convergence evolution taking place, hence, ABCG families might be evolving to take similar mechanistic features from the WzmWzt family.

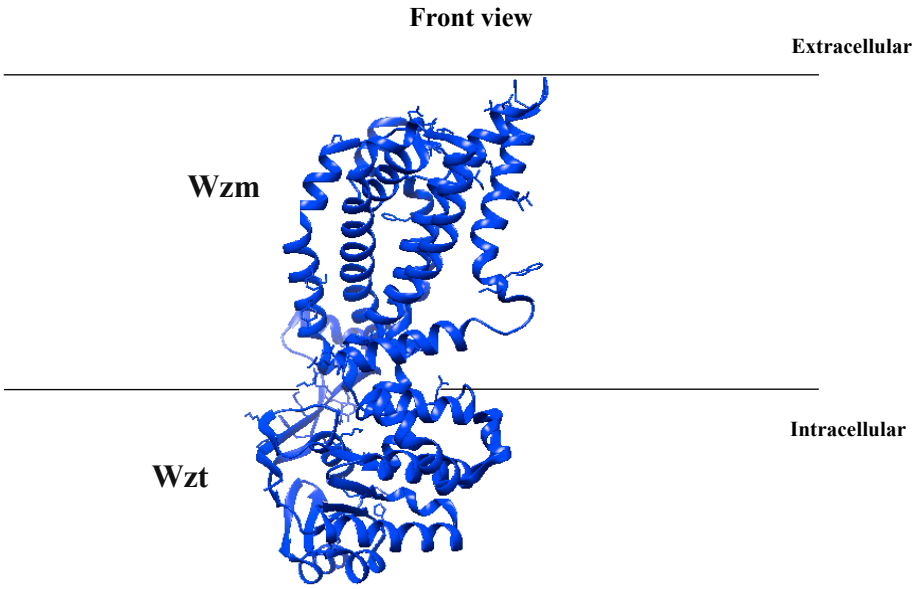


Figure 53:The chimera structure of WzmWzt

The structure of WzmWzt represented in blue ribbons, hence Wzm represents the TMD and Wzt is represented as the NBD. The structure of WzmWzt therefore has the features required for it to be an ABC transporter.

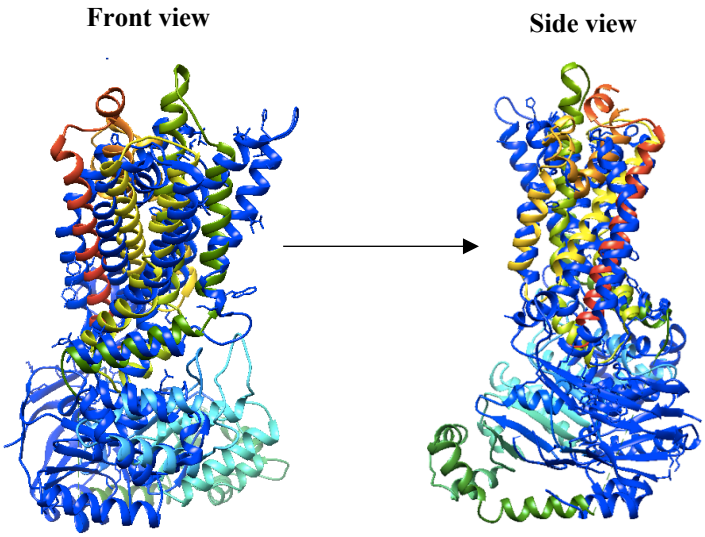


Figure 54:The structural alignment between ABCG2 and WzmWzt

The structure of ABCG2 represented in green, blue, orange and yellow ribbons; the structure of WzmWzt represented in blue ribbons. The structural alignment between ABCG2 and WzmWzt shows high structural similarities of the TMD and NBD.

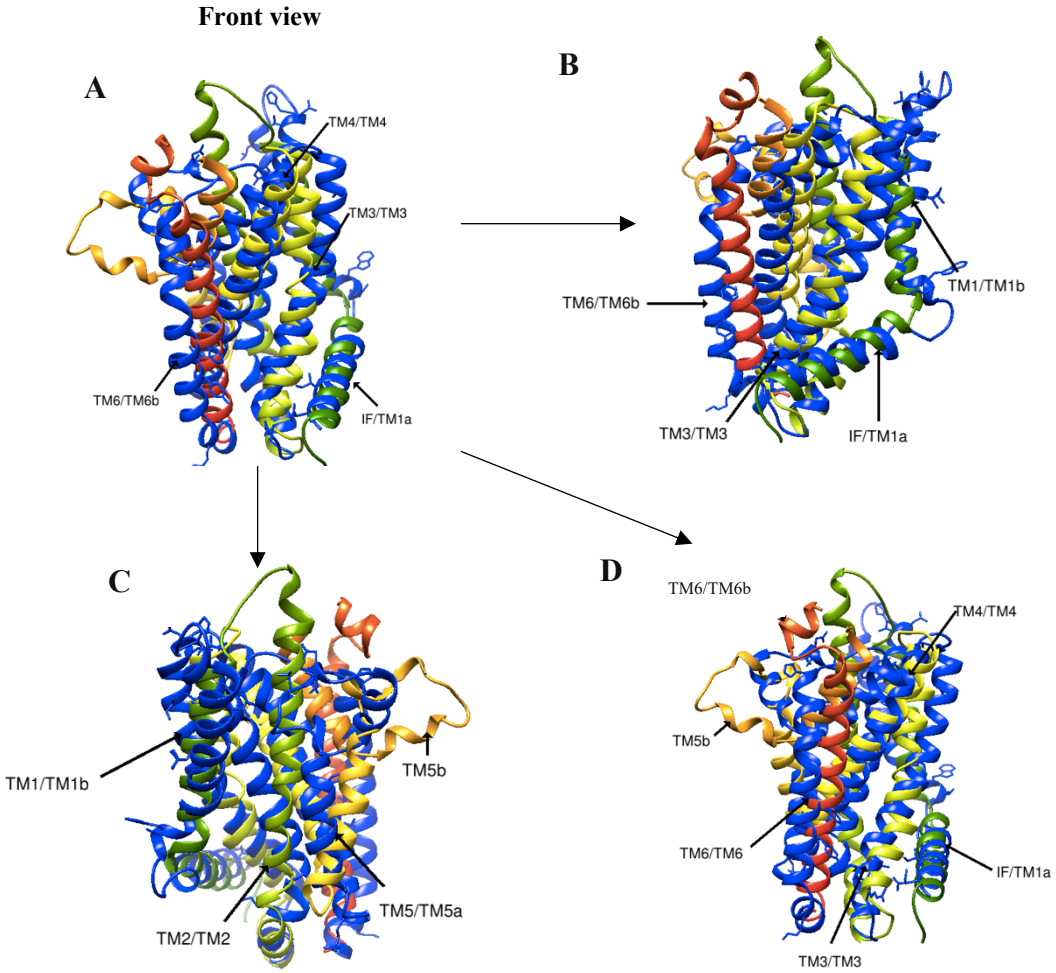


Figure 55: The structural alignment between ABCG2 TMD and Wzm

A) *Front view.* ***B****) Front view rotated by clockwise 90°.* ***C****) Front view rotated by clockwise 180°.* ***D)*** *Front view rotated by 300°.The structure of Wzm represented as blue ribbons in the alignment and ABCG2 is represented as green, yellow, orange and red ribbons.* ***A)*** *The structural alignment between: ABCG2 TM1a and Wzm IF, ABCG2 TM3 and Wzm TM3, ABCG2 TM4 and Wzm TM4 and ABCG2 TM6 and Wzm TM6b.****B)*** *The structural alignment between: ABCG2 TM1a and Wzm IF, ABCG2 TM1b and Wzm TM1, ABCG2 TM3 and Wzm TM3 and ABCG2 TM6b and Wzm TM6.****C)*** *The structural alignment between: ABCG2 TM1b and Wzm TM1, ABCG2 TM2 and Wzm TM2, ABCG2 TM5a and Wzm TM5.* ***D)*** *The structural alignment between: ABCG2 TM1a and Wzm IF, ABCG2 TM3 and Wzm TM3, ABCG2 TM4 and Wzm TM4 and ABCG2 TM6b and Wzm TM6.*

Alignments and residue regions – ABCG2 TMD: Wzm	Fully conserved residues	Weakly conserved residues
TM1a: IF (381-400)	2	16
TM2:TM2 (491-520)	0	23
TM3:TM3 (521-560)	2	17
TM4:TM4 (571-610)	3	26
TM5a:TM5 (611-630)	1	0
TM6b:TM6 (731-781	1	16

Table 21: Quantitative analysis of the alignment between ABCG2 TMD and Wzm

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG2 TMD and Wzm

4.4.2 The structural alignment between ABCG8 and WzmWzt

The WzmWzt used in the alignment of ABCG5/G8 had the additional solvents and ligands removed in order to obtain a better structural alignment of the TMD and NBD. The conserved residues mentioned in the alignment of ABCG8 TMD and Wzm referred from appendix 54; hence table 22 represents the summarised quantitative data from this alignment.

WzmWzt was seen to align with ABCG8 rather than ABCG5.The alignment between ABCG8 TMD and Wzm showed, Wzm IF and ABCG8 TM1 aligned significantly as seen in figure 56A, the corresponding structural sequence alignment showed 4 fully conserved and 11 weak conserved residues. Furthermore, ABCG8 TM1 and Wzm TM1 was seen to align well as it took similar structure and the helices overlapped as seen in figure 56A. The structural sequence alignment showed there to be 4 fully conserved and 11 weak conserved residues. ABCG8 TM2 and Wzm TM2 was also seen to be aligned well, the structural sequence alignment reflects there to be 2 fully conserved residues and 8 weak conserved residues. Thus, so far, the structural alignment of TM1 and TM2 between ABCG8 and Wzm have shown high structural similarity indicating an evolutionary link between the two families.

Through observation it can be seen that the alignment between ABCG8 TM3 and Wzm TM3 to have the best alignment compared to any other transmembrane helices between ABCG8

TMD and Wzm as seen in figure 56A. Henceforth, this was reflected through the structural sequence alignment where there were 4 fully conserved residues and 19 weak conserved residues between residues 531-570. In figure 56D, the alignment of Wzm TM4 and ABCG8 TM4 is seen as highly similar in structure as the helices overlap as seen in figure 56C; this was reflected between residues 579-630 where the structural sequence alignment showed 3 fully conserved and 10 weak conserved residues. The ABCG8 NBD and Wzt showed significant alignment where almost all the α helices and β-pleated sheets were aligned well (appendix 55). Furthermore, the NBD showed there 24 fully conserved residues between residues 1- 450. The alignment of ABCG8 NBD and Wzt had a percentage identity of 20.09% showing high structural similarities (appendix 56)

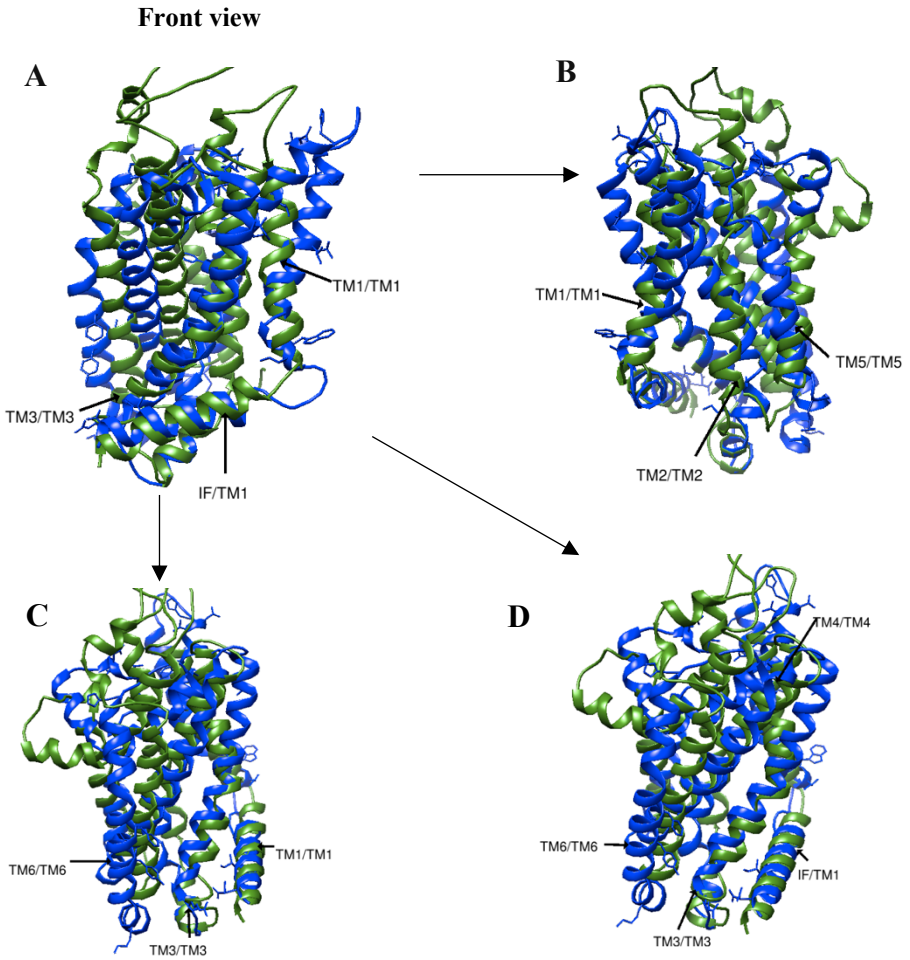


Figure 56:The structural alignment between ABCG8 and Wzm

The structure of Wzm represented as blue ribbons in the alignment and ABCG8 is represented as forest green. ***A)*** *Front view: the structural alignment between: ABCG8 TM1 and Wzm IF, ABCG8 TM1 and Wzm TM1 and ABCG8 TM3 and Wzm TM3.* ***B)*** *Front view rotated clockwise 90°: the structural alignment between: ABCG8 TM1 and Wzm TM1, ABCG8 TM2 and Wzm TM2, ABCG8 TM5 and Wzm TM5.* ***C)*** *Front view rotated anticlockwise 90°: the structural alignment between: ABCG8 TM1 and Wzm TM1, ABCG8 TM3 and Wzm TM3, ABCG8 TM6 and Wzm TM3.* ***D)*** *Front view rotated anticlockwise 45°. The structural alignment between: ABCG8 TM1 and Wzm IF, ABCG8 TM3 and Wzm TM3.*

Alignments and residue regions – ABCG8 TMD: Wzm	Fully conserved residues	Weakly conserved residues
TM1a:IF (1-200)	4	11
TM1:TM1 (200-360)	4	11
TM2:TM2 (361-500)	2	8
TM3:TM3 (531-570)	4	19
TM4:TM4 (579-630)	3	10
TM5:TM5 (700-750)	1	18
TM6:TM6 (780-900)	3	12

Table 22: Quantitative data analysis of the structural alignment between ABCG8 TMD and Wzm

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TMD and Wzm

4.4.3 The structural alignment between ABCA1 and WzmWzt

WzmWzt used in the alignment with ABCA1, had the additional solvents and ligands removed in order to make better transmembrane domain and nucleotide binding domain alignment. The conserved residues mentioned in the alignment of ABCA1 TMD1 and Wzm is referred from appendix 57; table 23 represents the summarised quantitative data for this alignment.

Initial alignment showed WzmWzt aligned to ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 NBD2 as seen from figure 57. The alignment of the transmembrane domain showed ABCA1 IF and Wzm TM1 aligned significantly well as seen in figure 58A.Furthermore, the structural sequence alignment showed 2 fully conserved and 16 weak conserved residues between 381-400. The alignment of ABCA1 TM2 and Wzm TM2 showed to be significantly better compared to the alignment of ABCA1 TM1 and Wzm TM1, however the structural sequence alignment only showed 23 weak residues between residues 491-520. The alignment of ABCA1 TM3 and Wzm TM3 showed to be the best structurally aligned transmembrane helices compared to the other transmembrane helices of the alignments. Henceforth, this was reflected by the structural sequence alignment where it showed 2 fully conserved and 17

weak conserved residues between 521-560. Similar to the alignment of TM3, the alignment of ABCA1 TM4 and Wzm TM4 also showed significant alignment as seen in figure 58A and D, furthermore the structural sequence alignment showed 3 fully conserved and 26 other weak conserved residues.

Once again in this alignment between WzmWzt and ABCA1, ABCA1 TM1-4 and Wzm TM1-4 aligned the best, however ABCA1 TM5 and TM6 did not align as significantly. The alignment between ABCA1 TM5 and Wzm TM5 showed alignment at the beginning of the helices, however both helices eventually diverged away from each other. The structural sequence alignment showed 1 fully conserved and 28 weak conserved residues between residues 611-630. It was noticed only segments of Wzm TM6 aligned with ABCA1 TM6 as seen in figure 58A, thus the structural sequence alignment showed there to be 1 fully conserved and 16 weak conserved residues between residues 731-781. So far, the structural alignment of the transmembrane domain showed ABCA1 and Wzm to have high structural similarity.

The ABCA1 NBD and Wzt showed significant alignment where almost all the α helices and β-pleated sheets were aligned well (appendix 58). Furthermore, the NBD showed 31 fully conserved residues between residues 901-1450 (appendix 59). The alignment of ABCA1 NBD and Wzt had a percentage identity of 20.90% showing high structural similarities.

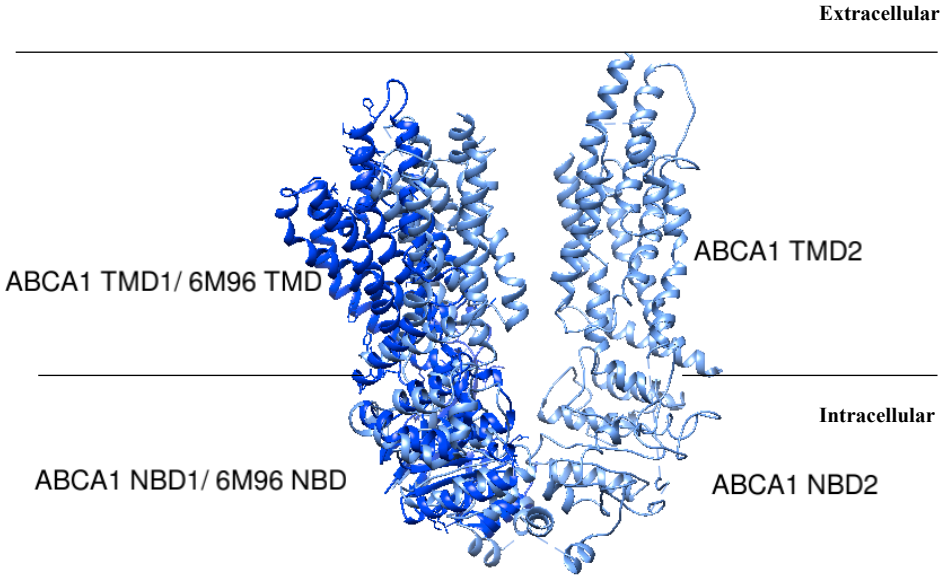


Figure 57: The structural alignment between ABCA1 and WzmWzt

The structure represented in dark blue is WzmWzt, structure ABCA1 is represented in cornflour blue. This figure illustrates WzmWzt shows to have better alignment to ABCA1 TMD1 NBD1 rather than ABCA1 TMD2 NBD2

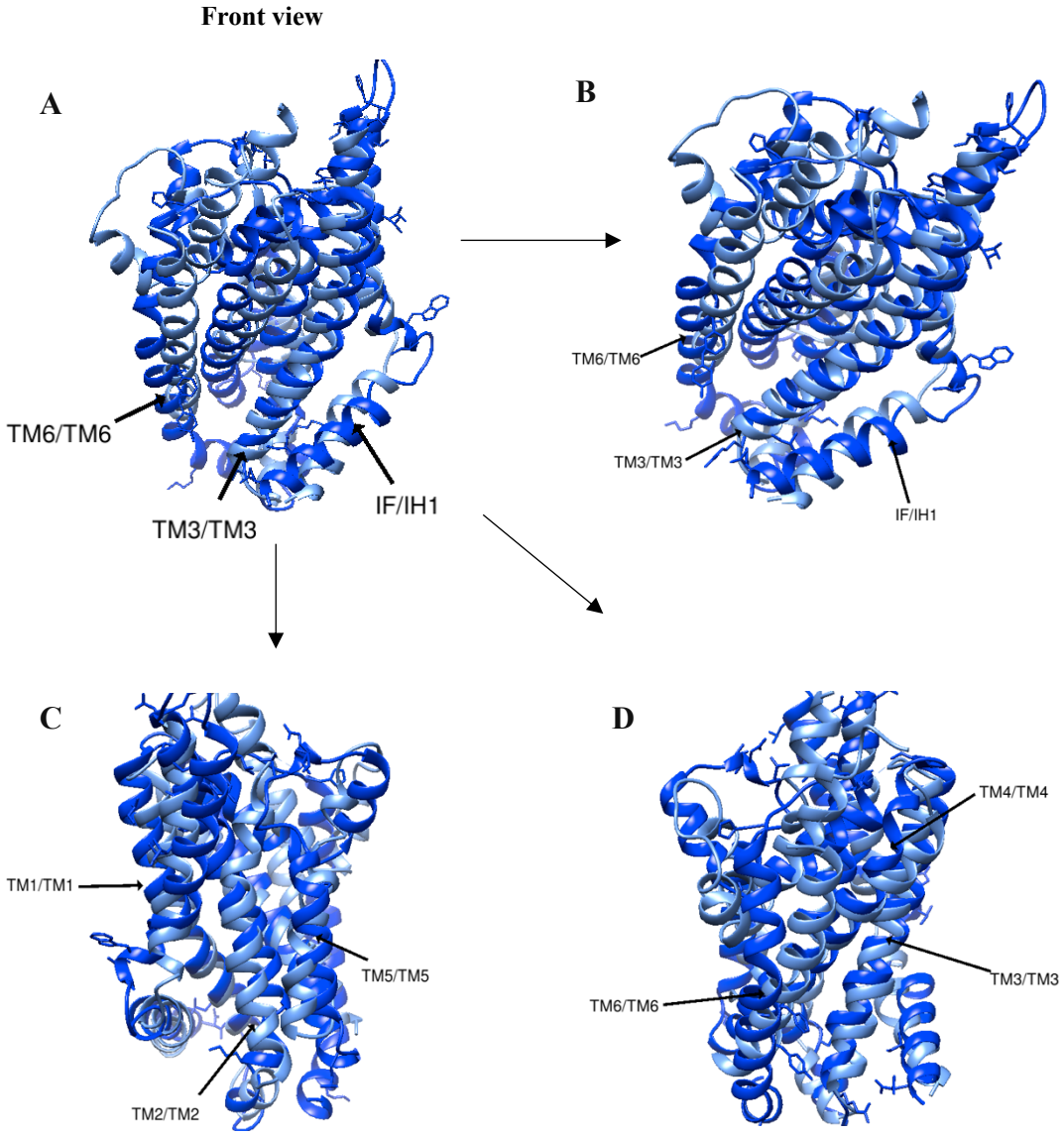


Figure 58: The structural alignment between ABCA1 TMD1 and Wzm

The structure of Wzm represented as blue ribbons in the alignment and ABCG8 is represented as forest green. ***A)*** *Front view.* ***C)*** *Front view rotated clockwise by 90°.* ***A,C)*** *The structural alignment between: ABCA1 TM1 and Wzm TM1, ABCA1 TM2 and Wzm TM2, ABCA1 TM5 and Wzm TM5* ***B)*** *Front view tilted forwards by 45°: the structural alignment between: ABCA1 IH1 and Wzm IF, ABCA1 TM3 and Wzm TM3, ABCA1 TM6 and Wzm TM6* ***D)****Front view rotated by 40° clockwise: the structural alignment between: ABCA1 TM3 and Wzm TM3, ABCA1 TM4 and Wzm TM4, ABCA1 TM6 and Wzm TM6.*

Alignments and residue regions – ABCA1 TMD: Wzm	Fully conserved residues	Weakly conserved residues
TM1:TM1 (381-400)	4	11
TM2:TM2 (491-520)	2	8
TM3:TM3 (521-560)	4	19
TM4:TM4 (561-600)	3	10
TM5:TM5 (611-630)	1	18
TM6:TM6 (731-781)	3	12

Table 23:Quantitative data analysis of the structural alignment between ABCA1 TMD and Wzm

The table represents the quantitative analysis of the fully and weakly conserved regions in the alignment of ABCG8 TM1-4 and Wzm

Chapter 5: Discussion and conclusion

5.1Discussion

Human ABCG2 is a membrane protein and a member of the ATP binding cassette (ABC) transporter superfamily, hence it has been identified from many studies to be associated with multidrug resistance in cancer chemotherapy. Studying the structure, mechanism and thermostability of this ABC transporter is important in developing new venues for therapeutic discoveries in overcoming the multidrug resistance. The first step which is required to study and understand ABCG2, is finding an efficient expression system which has the ability to overexpress ABCG2. Yeast has been known to be an excellent expression system used worldwide by scientist to express recombinant membrane protein. The two yeast species which are commonly used to express recombinant proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*, both these species are very fast growing, however the major advantage of using *Pichia pastoris* over *Saccharomyces cerevisiae* is it produces higher cell density growth and higher secreted protein yield. Furthermore, *Pichia pastoris* has emerged as an alternative host to *Saccharomyces cerevisiae* due to its shorter and less immunogenic glycosylation pattern (Tran *et al*., 2017).

In this thesis, ABCG2 was successfully expressed by *Pichia pastoris*, the *Pichia pastoris* cells were induced at 0.8% of methanol over a 3-day period, the results showed induction of methanol made at 24hr interval gave the highest number of growths hence highest yield of protein. Henceforth, this indicates that the methanol utilising pathway was upregulated to optimal during this period allowing the cells to proliferate at its fastest. Using *Pichia pastoris* as the expression system gave rise to an under glycosylated form of ABCG2 as the molecular weight seemed 10kDa lower in molecular mass compared to ABCG2 expressed in HEK cells. However, this does not cause substantial problems as confirmed by Mao *et al*, where it was reported from the ATPase activity assay, which was conducted, that the functionality of the under glycosylated form of ABCG2 still remained intact.

The mammalian expression systems are generally preferred over other cell lines as the cells are able to produce large and complex proteins with post translational modification. However, major problems faced with mammalian cell lines are that the yield of the recombinant proteins produced are approximately 10-100 folds lower than other expression system due to the slower growth. Furthermore, higher death rates are also a result from the

lower cell-based productivity. The results suggest that ABCG2 expressed using *Pichia pastoris* can be an alternative method and a cost-effective system to produce large quantities of recombinant protein.

Solubilisation of ABCG2 took place with the presence of DDM, in the structural study of ABCG2 conducted by Rosenberg *et al* it was found that DDM successfully solubilised ABCG2; the study had buffer A containing 2% DDM which was used to solubilise the ABCG2. Similarly, in this project ABCG2 was subjected to buffer A containing 2% DDM, and successful solubilisation of ABCG2-GFP was obtained after many attempts. There are many reasons as to why solubilisation may not take place; membrane proteins have a relatively high hydrophobic surface; therefore, membrane proteins have to be extracted from the cell membrane by forming a stable protein- detergent complex, and the protein must remain in its active conformation for it to be used in the experiment. Furthermore, harsh solubilisation conditions are known to cause the loss of protein activity as a result from loss of important lipids. Whether lipids have a major role in ABC transporter stability is yet to be understood, thus biochemical, bioinformatics and biophysical studies will help evaluate lipid's role in stabilising ABC transporter. Redetermining the critical micelle concentration and an alternative detergent such as DTT can be used instead of DDM to achieve successful solubilisation of ABCG2.

Purification of ABC transporters have posed difficulties especially as they are membrane proteins, being a membrane protein results in low abundance and require a detergent to become soluble in an aqueous solution (Lee *et al*., 2008). Once the membrane protein is purified, to resolve the structure of the membrane proteins requires crystallisation and X-ray crystallography to obtain high resolution structures. However, obtaining high resolution structures of ABC transporters have been one of the major challenges faced by structural biologist today. Furthermore, the importance of high-resolution structure of cavity 1 and cavity 2 of ABCG2 would give an opportunity of understanding a better understanding of how substrates and lipids are transported across the membrane, therefore a better delivery of action of drugs to treating cancerous cells or treating dysfunctional ABCG2 leading to various diseases could be determined.

The most common method used to purify ABCG2 so far is by using the Ni-NTA resin; in both studies Rosenberg *et al* showed Ni NTA resins having high affinity to His tag present on

the sequence, as a result purification was successful (Rosenberg *et al*. 2010; Rosenberg *et al*., 2015). Purification attempts were made using Ni NTA agarose beads in a batch process; however, the 8% and10% SDS PAGE gel showed no molecular weight associated to 90kDa. This potentially suggests that the His tag may have been cleaved during the process of purification or hidden within the sequence. To avoid this problem in the future moving the tag position would make it easier for the Ni NTA to bind. The purification of ABCG2 using the GFP trap resin produced a significantly small yield of pure proteins, suggesting that some of the GFP tags may have been cleaved and lost during the batch process. Due to time constraint other methods of purification was not attempted. As there is a FLAG tag present in the sequence, purification methods using FLAG tag affinity purification resin can be attempted, however the FLAG tag affinity purification resins are expensive. Other methods of improving the purification could be by using FPLC or any other AKTA protein purification system rather than the batch method. The advantage of the batch method is its manual handling using the centrifuge and no protein purifying computer systems need to be learnt. However, during each wash and change of emulation, proteins may be lost. As purifying membrane proteins is a very difficult process, as a result only a small concentration is obtained, losing proteins during the batch process can be detrimental to the final concentration of pure proteins obtained.

Understanding the molecular basis of protein thermostability has always been a major challenge in protein biology. There are theoretical implications such as important for the knowledge of how ligands could stabilise proteins, hence understand at which confirmation the protein has the largest stability. CETSA assay also known as cellular thermal shift assay allows to quantify the changes in thermal stability of protein upon ligand binding in intact cells. In this thesis, the microsomes express ABCG2-GFP were subjected to CETSA assay to verify if the thermal shift assay has worked on the protein. Results have shown as the fluorescent band indicating the presence of fused ABCG2-GFP decreases; after 55°C there were no bands seen. The results establishes that ABCG2 expressed by *Pichia pastoris* has a melting temperature of 55°C. It can be expected the melting temperature to slightly vary depending on the expression system used to express ABCG2. Furthermore, as the CETSA assay shows proof of principle, the assay can be performed on ABCG2 with various target drug such as tyrosine kinase inhibitor, thus the performance of each assay can be further subjected to biophysical and biochemical assays to understand further on ligand binding at a certain temperature. In the case of ABCG2 when ligand is bound at 55°C, bioinformatic

simulation should be investigated to see whether there has been a change in position of the ligand at the binding pocket in cavity 1.

Bioinformatics is a useful tool used in structural biology to create an understanding of structure, function and mechanism of the proteins. Understanding structure and mechanism allows one to look at drug interactions with protein of interest. In the last few years new structures have been determined for the ABCG family, Lee *et al* was the first to establish and determine the structure of ABCG5/G8 cholesterol exporter by X-ray crystallography (Ford *et al*., 2019; Lee *et al*., 2016). This epic determination of the structure was useful as later Locher *et al* determined the structure of ABCG2 xenobiotic exporter using single particle cryo-EM (Ford *et al*., 2019). Alongside the ABCG family, in recent years there has been a high-resolution structure of ABCA1 determined. Interestingly, When Lee *et al* first published the structure of ABCG5G8 it was suggested the overall shape of ABCG5/G8 resembled bacterial importer ModBC, MalFGK2 and MetNI (Ford *et al*., 2019), however, this was disagreed due to the radical rearrangement of the topological transmembrane helices would be required to fit this idea. The TMD is a key component in the process of translocation of enormous variety of substrates, ions, small peptides to large toxins out of the cell (Oswald *et al*., 2006). Thus, many TMD have been analysed to lack primary structure conservation, highly heterogenous, hence producing a large diversity of transporters with varying function. Therefore, for Ford *et al* to suggest there is little structural similarity of the TMD of ABCA and ABCG families to bacterial importers ModBC, MalFGK$_2$ and MetNI implicates the idea to be irrelevant and incorrect. As an alternative, Ford *et al* firstly proposed that fold and topology of the transmembrane helices in each ABCA1 TMD resembled similar to that of ABCG2 TMD. The observation of the similarity was interesting as normally in structural studies of ABC transporters, the NBD was the main focus in looking at conservation. Although Ford *et al* suggested the similarity in structure between the two families, there were no detailed structural alignment of the TMD and NBD to support this theory. Secondly it was proposed, that mechanotransducer MacB TMD and NBD had similarity in alignment with TMD and NBD of ABCA and ABCG families, yet no detailed structural alignment once again had been performed to support this theory at the time.

In this thesis, it was found that ABCG2/G8 TMD NBD and ABCA1 TMD1NBD had significant structural alignment reflected by overall RMSD of the alignments were a few Å across the sequence length indicating high degree of conservation. The alignment of the NBD

was as expected as many studies have proved high structural similarity of the motor domain, however the overall similarity of the TMD of ABCG2/G8 and ABCA1 may suggest that at one point in evolution they were exporting similar substrates out of the cells as they belonged from the same family.

The mechanotransducer MacB is present with 4 TMD; the structural alignment between ABCG2/G8 TM1-4 and ABCA1 TM1-4 to MacB showed there to be high structural similarity which was reflected by the overall RMSD of the alignments were a few Å across the sequence length (appendix 1). An evolution of ABCG and ABCA subfamilies from mechanotransducer MacB rather than ModBC-type importer seems to fit the theory as it is easier to rationalise the gene arrangement. Evaluating the similarity of the structural alignment between ABCG and ABCA families to MacB suggest that these ABC proteins may have adopted a more classical transmembrane pumping function earlier in evolution. However, over time as both families became exposed to diverse ranges of substrates, the alternative access mechanism was advantageous therefore was adopted. The structural alignment of TM1-4 of ABCG and ABCA families to MacB suggest that TM1-4 in ABCG2/G8 and ABCA1 are majorly involved in the translocation of substrates out of the cell as it shows high conservation of these transmembrane throughout evolution rather than TM5 and TM6. To further validate this theory, biophysical and biochemical assays should be conducted to look create a simulation of how substrates and lipids interact with each transmembrane helix. Understanding lipids and substrate interaction can impact the understanding of transport function and potentially see if there is a change in membrane localisation whilst translocation of substrates take place. Such that, it is observed that whilst cholesterol molecule is bound in the multidrug binding pocket, the localisation of the molecule is central, hydrophobic and inward facing translocation pathway (Taylor *et al*., 2017). Whether MacB also have similar motion of localisation whilst binding to lipids or the substrates will give an insight into which transmembrane helix is important in the translocation process. Furthermore, studying intensely the interaction of lipids to the specific transmembrane helix will further help understand whether lipids binding are important part of ABC transporter structure and stability.

The motor domain also known as the NBD has the major role in substrate recognition, hence as the results of the structural alignment of ABCG2/G8 NBD to MacB NBD show high

structural similarity it can be postulated that there is a similar pattern in substrate recognition between the families.

The alternating access mechanism is a widely accepted mechanism for ABC proteins; however, a lot of bacterial ABC transporters have shown to not obey this mechanism such as WzmWzt ABC transporter, MlaE and MlaF. The structural alignment of ABCG2 and WzmWzt showed all the helices except TM5b had significant alignment, similarly ABCG8 presented with the same results and within the alignment of ABCA1 TMD1, only EH1 did not align. All the structural alignment shows there to be some evolutionary link and it can be postulated that convergent evolution may take place, therefore this suggest that WzmWzt, MlaE, MlaF may be evolving to adopt the alternative access mechanism however still in the early stages of evolving to adopt this mechanism, hence further bioinformatic, biophysical and biochemical assays should employed to look further into specific transmembrane helix to substrate and lipid interaction to highlight the similarities.

5.2 Conclusion

In this project the conclusion is drawn from the study are as follows:

- Pichia pastoris is an efficient yeast expression system to express ABCG2
- Obtaining the melting temperature of ABCG2 using the CETSA assay shows proof of principle for screening drugs/inhibitors
- The high similarity of the structural alignment between the TMD and NBD of the ABCA and ABCG families shows they most likely share a common ancestor
- The significant similarity of the structural alignment between mechanotransducer MacB to ABCA and ABCG families indicate that MacB may be the common ancestor; this suggest that ABCA and ABCG families may have adopted a more classical transmembrane pumping function
- The significant similarity of the structural alignment between: WzmWzt, MlaE and MlaF to ABCA and ABCG families showed a possibility of convergence evolution taking place between these families.

As a future perspective, screening tyrosine kinase inhibitor like Lapatinib on ABCG2 using the CETSA assay could potentially show whether thermal shift is taking place hence the thermostability of ABCG2 can be studied in the presence of a drug. The bioinformatic studies showed degrees of structural similarity between ABCA and ABCG families bacterial ABC transporters (MacB, WzmWzt, MlaE and MlaF) helps to pinpoint where ABC transporters diverged from prokaryotic structure and mechanism of action. Thus, it gives an indication of how some mammalian ABC transporters may molecularly function based on similarities to bacterial ABC transporters and could be extrapolated to other members of ABC superfamily. However, it is yet to be further investigated how these theories could be applies under physiological and pathophysiological conditions.

References

Aguiar, J. A., Tamminga, A., Lobb, B., Huff, R. D., Nguyen, J. P., Kim, Y., Dvorkin-Gheva, A., Stampfli, M. R., Doxey, A. C. and Hirota, J. A. (2019) 'The impact of cigarette smoke exposure, COPD, or asthma status on ABC transporter gene expression in human airway epithelial cells', *Sci Rep,* 9(1), pp. 153.

Aier, I., Varadwaj, P. and Raj, U., 2016. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Scientific Reports*, 6(1).

Albrecht, C. and Viturro, E., 2006. The ABCA subfamily—gene and protein structures, functions and associated hereditary diseases. *Pflügers Archiv - European Journal of Physiology*, 453(5), pp.581-589.

Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A. and Stevens, R. C. (2008) 'Microscale fluorescent thermal stability assay for membrane proteins', *Structure,* 16(3), pp. 351-9.

Almenara, S., Lozano, B., Gimenez, P., Herrera, I., Miralles, C., Bellot, P., Rodríguez, M., Francés, R., Gonzalez-Navajas, J. M., Pascual, S. and Zapater, P. (2020) 'Functionality of beta-adrenergic receptors in patients with cirrhosis treated chronically with non-selective beta-blockers', *Hepatol Int*.

Andreoletti, P., Raas, Q., Gondcaille, C., Cherkaoui-Malki, M., Trompier, D. and Savary, S. (2017) 'Predictive Structure and Topology of Peroxisomal ATP-Binding Cassette (ABC) Transporters', *Int J Mol Sci,* 18(7).

Anne Morgat, Thierry Lombardot, Elisabeth Coudert, Kristian Axelsen, Teresa Batista Neto, Sebastien Gehant, Parit Bansal, Jerven Bolleman, Elisabeth Gasteiger, Edouard de Castro, Delphine Baratin, Monica Pozzato, Ioannis Xenarios, Sylvain Poux, Nicole Redaschi, Alan Bridge, The UniProt Consortium, Enzyme annotation in UniProtKB using Rhea, *Bioinformatics*, Volume 36, Issue 6, 15 March 2020, Pages 1896–1901, https://doi.org/10.1093/bioinformatics/btz817

Bank, R., 2019. *RCSB PDB*. [online] Rcsb.org. Available at: <https://www.rcsb.org/search> [Accessed 17 March 2021].

Becker, U., Ehrhardt, C., Daum, N., Baldes, C., Schaefer, U. F., Ruprecht, K. W., Kim, K. J. and Lehr, C. M. (2007) 'Expression of ABC-transporters in human corneal tissue and the transformed cell line, HCE-T', *J Ocul Pharmacol Ther,* 23(2), pp. 172-81.

Bickers, S. C., Sayewich, J. S. and Kanelis, V. (2020) 'Intrinsically disordered regions regulate the activities of ATP binding cassette transporters', *Biochim Biophys Acta Biomembr,* 1862(6), pp. 183202.

Burgard, J., Grünwald-Gruber, C., Altmann, F., Zanghellini, J., Valli, M., Mattanovich, D. and Gasser, B., 2019. The secretome of Pichia pastoris in fed-batch cultivations is largely independent of the carbon source but changes quantitatively over cultivation time *Microbial Biotechnology*, 13(2), pp.479-494.

Canet, M. J., Merrell, M. D., Harder, B. G., Maher, J. M., Wu, T., Lickteig, A. J., Jackson, J. P., Zhang, D. D., Yamamoto, M. and Cherrington, N. J. (2015) 'Identification of a functional antioxidant response element within the eighth intron of the human ABCC3 gene', *Drug Metab Dispos,* 43(1), pp. 93-9.

Caffalette, C. A., Corey, R. A., Sansom, M. S. P., Stansfeld, P. J. and Zimmer, J. (2019) 'A lipid gating mechanism for the channel-forming O antigen ABC transporter', *Nat Commun,* 10(1), pp. 824.

Chai, A., Ammit, A. and Gelissen, I., 2017. Examining the role of ABC lipid transporters in pulmonary lipid homeostasis and inflammation. *Respiratory Research*, 18(1).

Chi, X., Fan, Q., Zhang, Y., Liang, K., Wan, L., Zhou, Q. and Li, Y. (2020) 'Structural mechanism of phospholipids translocation by MlaFEDB complex', *Cell Res*.

Chen, Z., Shi, T., Zhang, L., Zhu, P., Deng, M., Huang, C., Hu, T., Jiang, L. and Li, J. (2016) 'Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade', *Cancer Lett,* 370(1), pp. 153-64.

Cho, K. H., Ribeiro, O., Du, Y., Tikhonova, E., Mortensen, J. S., Markham, K., Hariharan, P., Loland, C. J., Guan, L., Kobilka, B. K., Byrne, B. and Chae, P. S. (2016) 'Mesitylene-Cored Glucoside Amphiphiles (MGAs) for Membrane Protein Studies: Importance of Alkyl Chain Density in Detergent Efficacy', *Chemistry,* 22(52), pp. 18833-18839.

Consortium, U. (2019b) 'UniProt: a worldwide hub of protein knowledge', *Nucleic Acids Res,* 47(D1), pp. D506-D515.

Couch, G.S., Hendrix, D.K., and Ferrin, T.E. "Nucleic acid visualization with UCSF Chimera." *Nucleic Acids Res.* **34**(4):e29 (2006).

Cox, M. H., Kapoor, P., Briggs, D. A. and Kerr, I. D. (2018) 'Residues contributing to drug transport by ABCG2 are localised to multiple drug-binding pockets', *Biochem J,* 475(9), pp. 1553-1567.

Crow, A., Greene, N. P., Kaplan, E. and Koronakis, V. (2017) 'Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily', *Proc Natl Acad Sci U S A,* 114(47), pp. 12572-12577.

Dai, C. L., Tiwari, A. K., Wu, C. P., Su, X. D., Wang, S. R., Liu, D. G., Ashby, C. R., Huang, Y., Robey, R. W., Liang, Y. J., Chen, L. M., Shi, C. J., Ambudkar, S. V., Chen, Z. S. and Fu, L. W. (2008) 'Lapatinib (Tykerb, GW572016) reverses multidrug resistance in cancer cells by inhibiting the activity of ATP-binding cassette subfamily B member 1 and G member 2', *Cancer Res,* 68(19), pp. 7905-14.

Dean, M., Hamon, Y. & Chimini, G., 2001. The Human ATP-Binding Cassette transporter superfamily. *Journal of Lipid Research*, 42, pp.1007–1017. Available at: http://www.jlr.org/content/42/7/1007.full.pdf+html.

Domanitskaya, N., Wangari-Talbot, J., Jacobs, J., Peiffer, E., Mahdaviyeh, Y., Paulose, C., Malofeeva, E., Foster, K., Cai, K. Q., Zhou, Y., Egleston, B. and Hopper-Borge, E. (2014)

'Abcc10 status affects mammary tumour growth, metastasis, and docetaxel treatment response', *Br J Cancer,* 111(4), pp. 696-707.

Domenichini, A., Adamska, A. and Falasca, M., 2019. ABC transporters as cancer drivers: Potential functions in cancer development. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1863(1), pp.52-60.

El-Awady, R., Saleh, E., Hashim, A., Soliman, N., Dallah, A., Elrasheed, A. and Elakraa, G. (2016) 'The Role of Eukaryotic and Prokaryotic ABC Transporter Family in Failure of Chemotherapy', *Front Pharmacol,* 7, pp. 535.

Fitzpatrick, A. W. P., Llabrés, S., Neuberger, A., Blaza, J. N., Bai, X. C., Okada, U., Murakami, S., van Veen, H. W., Zachariae, U., Scheres, S. H. W., Luisi, B. F. and Du, D. (2017) 'Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump', *Nat Microbiol,* 2, pp. 17070.

Ford, R. C., Marshall-Sabey, D. and Schuetz, J. (2020) 'Linker Domains: Why ABC Transporters 'Live in Fragments no Longer'', *Trends Biochem Sci,* 45(2), pp. 137-148.

Ford, R. C. and Beis, K. (2019) 'Learning the ABCs one at a time: structure and mechanism of ABC transporters', *Biochem Soc Trans,* 47(1), pp. 23-36.

Forrest, L., Krämer, R. and Ziegler, C., 2011. The structural basis of secondary active transport mechanisms. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(2), pp.167-188.

Frick, M. and Schmidt, C. (2019) 'Mass spectrometry-A versatile tool for characterising the lipid environment of membrane protein assemblies', *Chem Phys Lipids,* 221, pp. 145-157.

Friman, T. (2020) 'Mass spectrometry-based Cellular Thermal Shift Assay (CETSA®) for target deconvolution in phenotypic drug discovery', *Bioorg Med Chem,* 28(1), pp. 115174.

Galkin, M., Russell, A., Vik, S., Berry, R. and Ishmukhametov, R., 2018. Detergent-free Ultrafast Reconstitution of Membrane Proteins into Lipid Bilayers Using Fusogenic complementary-charged Proteoliposomes. *Journal of Visualized Experiments*, (134).

Gil-Martins, E., Barbosa, D. J., Silva, V., Remião, F. and Silva, R. (2020) 'Dysfunction of ABC transporters at the blood-brain barrier: Role in neurological disorders', *Pharmacol Ther,* 213, pp. 107554.

Greene, N. P., Kaplan, E., Crow, A. and Koronakis, V. (2018) 'Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective', *Front Microbiol,* 9, pp. 950.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. "Visualizing density maps with UCSF Chimera." *J. Struct. Biol.* **157**(1):281-287 (2007).

Goddard, T.D., Huang, C.C., and Ferrin, T.E. "Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies." *Structure* **13**(3):473-482 (2005).

Goh, B. C., Chua, Y. K., Qian, X., Lin, J., Savko, M., Dedon, P. C. and Lescar, J. (2020) 'Crystal structure of the periplasmic sensor domain of histidine kinase VbrK suggests indirect sensing of β-lactam antibiotics', *J Struct Biol,* 212(2), pp. 107610

Guo, Z., Song, T., Xue, Z., Liu, P., Zhang, M., Zhang, X. and Zhang, Z. (2020) 'Using CETSA assay and a mathematical model to reveal dual Bcl-2/Mcl-1 inhibition and on-target mechanism for ABT-199 and S1', *Eur J Pharm Sci,* 142, pp. 105105.

Harwood, M. D., Neuhoff, S., Rostami-Hodjegan, A. and Warhurst, G. (2016) 'Breast Cancer Resistance Protein Abundance, but Not mRNA Expression, Correlates With Estrone-3-Sulfate Transport in Caco-2', *J Pharm Sci,* 105(4), pp. 1370-5.

He, P., Gelissen, I. and Ammit, A., 2020. Regulation of ATP binding cassette transporter A1 (ABCA1) expression: cholesterol-dependent and – independent signaling pathways with relevance to inflammatory lung disease. *Respiratory Research*, 21(1).

Hertig S, Goddard TD, Johnson GT, Ferrin TE. *Biophys J.* 2015 May 5;108(9):2097-102.

Higgins, C. F. (1992) 'ABC transporters: from microorganisms to man', *Annu Rev Cell Biol,* 8, pp. 67-113.

Hinz, A. and Tampé, R. (2012) 'ABC transporters and immunity: mechanism of self-defense', *Biochemistry,* 51(25), pp. 4981-9.

Hodges, L. M., Markova, S. M., Chinn, L. W., Gow, J. M., Kroetz, D. L., Klein, T. E. and Altman, R. B. (2011) 'Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein)', *Pharmacogenet Genomics,*21(3), pp. 152-61.

Hollenstein, K., Dawson, R. J. and Locher, K. P. (2007) 'Structure and mechanism of ABC transporter proteins', *Curr Opin Struct Biol,* 17(4), pp. 412-8.

Hulme, E. C. and Trevethick, M. A. (2010) 'Ligand binding assays at equilibrium: validation and interpretation', *Br J Pharmacol,* 161(6), pp. 1219-37.

Ivashov, V. A., Grillitsch, K., Koefeler, H., Leitner, E., Baeumlisberger, D., Karas, M. and Daum, G. (2013) 'Lipidome and proteome of lipid droplets from the methylotrophic yeast Pichia pastoris', *Biochim Biophys Acta,* 1831(2), pp. 282-90.

Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundbäck, T., Nordlund, P. and Martinez Molina, D. (2014) 'The cellular thermal shift assay for evaluating drug target interactions in cells', *Nat Protoc,* 9(9), pp. 2100-22.

Jasinski, M., Ducos, E., Martinoia, E. and Boutry, M., 2003. The ATP-Binding Cassette Transporters: Structure, Function, and Gene Family Comparison between Rice and Arabidopsis. *Plant Physiology*, 131(3), pp.1169-1177.

Johnson, M., Sutcliffe, M. and Blundell, T., 1990. Molecular anatomy: Phyletic relationships derived from three-dimensional structures of proteins. *Journal of Molecular Evolution*, 30(1), pp.43-59.

Karbanova, S., Cerveny, L., Jiraskova, L., Karahoda, R., Ceckova, M., Ptackova, Z. and Staud, F. (2019) 'Transport of ribavirin across the rat and human placental barrier: Roles of nucleoside and ATP-binding cassette drug efflux transporters', *Biochem Pharmacol,*163, pp. 60-70

Kopecka, J., Trouillas, P., Gašparović, A., Gazzano, E., Assaraf, Y. G. and Riganti, C. (2020) 'Phospholipids and cholesterol: Inducers of cancer multidrug resistance and therapeutic targets', *Drug Resist Updat,* 49, pp. 100670.

Kos, V. & Ford, R.C., 2009. The ATP-binding cassette family: A structural perspective. *Cellular and Molecular Life Sciences*, 66(19), pp.3111–3126.

Krissinel, E. and Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp.2256-2268.

Kwan, T. O. C., Reis, R., Siligardi, G., Hussain, R., Cheruvara, H. and Moraes, I. (2019) 'Selection of Biophysical Methods for Characterisation of Membrane Proteins', *Int J Mol Sci,* 20(10).

Lee, J. Y., Kinch, L. N., Borek, D. M., Wang, J., Urbatsch, I. L., Xie, X. S., Grishin, N. V., Cohen, J. C., Otwinowski, Z., Hobbs, H. H. and Rosenbaum, D. M. (2016) 'Crystal structure of the human sterol transporter ABCG5/ABCG8', *Nature,* 533(7604), pp. 561-4.

Lee, Y., Block, G., Chen, H., Folch-Puy, E., Foronjy, R., Jalili, R., Jendresen, C., Kimura, M., Kraft, E., Lindemose, S., Lu, J., McLain, T., Nutt, L., Ramon-Garcia, S., Smith, J., Spivak, A., Wang, M., Zanic, M. and Lin, S., 2008. One-step isolation of plasma membrane proteins using magnetic beads with immobilized concanavalin A. *Protein Expression and Purification*, 62(2), pp.223-229.

Lengger, B. and Jensen, M. K. (2020) 'Engineering G protein-coupled receptor signalling in yeast for biotechnological and medical purposes', *FEMS Yeast Res,* 20(1).

Lewinson, O. and Livnat-Levanon, N. (2017) 'Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations', *J Mol Biol,* 429(5), pp. 606-619.

Linton, K. J. (2007) 'Structure and function of ABC transporters', *Physiology (Bethesda),* 22, pp. 122-30.

Lin, J., Yu, Y., Wang, X., Ke, Y., Sun, C., Yue, L., Xu, G., Xu, B., Xu, L., Cao, H., Xu, D., Olsen, N. and Chen, W. (2019) 'Iguratimod Inhibits the Aggressiveness of Rheumatoid Fibroblast-Like Synoviocytes', *J Immunol Res,* 2019, pp. 6929286.

Liu, X. (2019) 'ABC Family Transporters', *Adv Exp Med Biol,* 1141, pp. 13-100

Magrone, T., Magrone, M., Russo, M. A. and Jirillo, E. (2020) 'Taking Advantage of Plant Defense Mechanisms to Promote Human Health. The Plant Immune System. First of Two Parts', *Endocr Metab Immune Disord Drug Targets*.

Maqbool, A., Horler, R. S., Muller, A., Wilkinson, A. J., Wilson, K. S. and Thomas, G. H. (2015) 'The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity', *Biochem Soc Trans,* 43(5), pp. 1011-7.

Mao, Q., Conseil, G., Gupta, A., Cole, S. P. and Unadkat, J. D. (2004) 'Functional expression of the human breast cancer resistance protein in Pichia pastoris', *Biochem Biophys Res Commun,* 320(3), pp. 730-7.

Martinez, N. J., Asawa, R. R., Cyr, M. G., Zakharov, A., Urban, D. J., Roth, J. S., Wallgren, E., Klumpp-Thomas, C., Coussens, N. P., Rai, G., Yang, S. M., Hall, M. D., Marugan, J. J., Simeonov, A. and Henderson, M. J. (2018b) 'A widely-applicable high-throughput cellular thermal shift assay (CETSA) using split Nano Luciferase', *Sci Rep,* 8(1), pp. 9472.

Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C., and Ferrin, T.E. "Tools for integrated sequence-structure analysis with UCSF Chimera." *BMC Bioinformatics* **7**(1):339 (2006).

Mishra, N. K., Chang, J. and Zhao, P. X. (2014) 'Prediction of membrane transport proteins and their substrate specificities using primary sequence information', *PLoS One,* 9(6), pp. e100278.

Mo, W. & Zhang, J.T., 2012. Human ABCG2: Structure, function, and its role in multidrug resistance. *International Journal of Biochemistry and Molecular Biology*, 3(1), pp.1– 27

Moitra, K. and Dean, M. (2011) 'Evolution of ABC transporters by gene duplication and their role in human disease', *Biol Chem,* 392(1-2), pp. 29-37.

Morita, M. and Imanaka, T. (2012) 'Peroxisomal ABC transporters: structure, function and role in disease', *Biochim Biophys Acta,* 1822(9), pp. 1387-96.

Morris, J.H., Huang, C.C., Babbitt, P.C., and Ferrin, T.E. "structureViz: Linking Cytoscape and UCSF Chimera." *Bioinformatics* **23**(17):2345-2347 (2007).

Morrison, K. A., Akram, A., Mathews, A., Khan, Z. A., Patel, J. H., Zhou, C., Hardy, D. J., Moore-Kelly, C., Patel, R., Odiba, V., Knowles, T. J., Javed, M. U., Chmel, N. P., Dafforn, T. R. and Rothnie, A. J. (2016) 'Membrane protein extraction and purification using styrene-maleic acid (SMA) copolymer: effect of variations in polymer structure', *Biochem J,*473(23), pp. 4349-4360.

Nakagawa, H., Wakabayashi-Nakao, K., Tamura, A., Toyoda, Y., Koshiba, S. and Ishikawa, T. (2009a) 'Disruption of N-linked glycosylation enhances ubiquitin-mediated proteasomal degradation of the human ATP-binding cassette transporter ABCG2', *FEBS J,* 276(24), pp. 7237-52.

Nakagawa, H., Wakabayashi-Nakao, K., Tamura, A., Toyoda, Y., Koshiba, S. and Ishikawa, T. (2009b) 'Disruption of N-linked glycosylation enhances ubiquitin-mediated proteasomal degradation of the human ATP-binding cassette transporter ABCG2', *FEBS J,* 276(24), pp. 7237-52.

Nasim, F., Schmid, D., Szakács, G., Sohail, A., Sitte, H. H., Chiba, P. and Stockner, T. (2020) 'Active transport of rhodamine 123 by the human multidrug transporter P-glycoprotein involves two independent outer gates', *Pharmacol Res Perspect,* 8(2), pp. e00572.

Neumann, J., Rose-Sperling, D. and Hellmich, U., 2017. Diverse relations between ABC transporters and lipids: An overview. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1859(4), pp.605-618.

Nucleic Acids Research, 2018. UniProt: a worldwide hub of protein knowledge. 47(D1), pp.D506-D515.

Okada, U., Yamashita, E., Neuberger, A., Morimoto, M., van Veen, H. W. and Murakami, S. (2017) 'Crystal structure of tripartite-type ABC transporter MacB from Acinetobacter baumannii', *Nat Commun,* 8(1), pp. 1336.

Oliveira, A. S., Baptista, A. M. and Soares, C. M. (2010) 'Insights into the molecular mechanism of an ABC transporter: conformational changes in the NBD dimer of MJ0796', *J Phys Chem B,* 114(16), pp. 5486-96.

Oswald, C., Holland, I. and Schmitt, L., 2006. The motor domains of ABC-transporters. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 372(6), pp.385-399

Petravicius, P. O., Costa-Martins, A. G., Silva, M. N., Reis-Cunha, J. L., Bartholomeu, D. C., Teixeira, M. M. G. and Zingales, B. (2019) 'Mapping benznidazole resistance in trypanosomatids and exploring evolutionary histories of nitroreductases and ABCG transporter protein sequences', *Acta Trop,* 200, pp. 105161.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *J. Comput. Chem.* **25**(13):1605-1612 (2004)

Pollock, N. L., Rimington, T. L. and Ford, R. C. (2015) 'Characterizing diverse orthologues of the cystic fibrosis transmembrane conductance regulator protein for structural studies', *Biochem Soc Trans,* 43(5), pp. 894-900.

Rajesh, S., Overduin, M. and Bonev, B. B. (2016) 'NMR of Membrane Proteins: Beyond Crystals', *Adv Exp Med Biol,* 922, pp. 29-42.

Rice, A. J., Park, A. and Pinkett, H. W. (2014) 'Diversity in ABC transporters: type I, II and III importers', *Crit Rev Biochem Mol Biol,* 49(5), pp. 426-37.

Rees, D. C., Johnson, E. and Lewinson, O. (2009) 'ABC transporters: the power to change', *Nat Rev Mol Cell Biol,* 10(3), pp. 218-27.

Robajac, D., Zámorová, M., Katrlík, J., Miković, Ž. and Nedić, O. (2017) 'Screening for the best detergent for the isolation of placental membrane proteins', *Int J Biol Macromol,*102, pp. 431-437.

Robey, R. W., Pluchino, K. M., Hall, M. D., Fojo, A. T., Bates, S. E. and Gottesman, M. M. (2018) 'Revisiting the role of ABC transporters in multidrug-resistant cancer', *Nat Rev Cancer,* 18(7), pp. 452-464.

Rosenberg, M. F., Bikadi, Z., Chan, J., Liu, X., Ni, Z., Cai, X., Ford, R. C. and Mao, Q. (2010) 'The human breast cancer resistance protein (BCRP/ABCG2) shows conformational changes with mitoxantrone', *Structure,* 18(4), pp. 482-93.

Rosenberg, M. F., Bikadi, Z., Hazai, E., Starborg, T., Kelley, L., Chayen, N. E., Ford, R. C. and Mao, Q. (2015) 'Three-dimensional structure of the human breast cancer resistance protein (BCRP/ABCG2) in an inward-facing conformation', *Acta Crystallogr D Biol Crystallogr,* 71(Pt 8), pp. 1725-35.

Shafi, T. and Jabeen, I. (2017) 'Grid-independent Descriptors (GRIND) Analysis and SAR Guided Molecular Docking Studies to Probe Selectivity Profiles of Inhibitors of Multidrug Resistance Transporters ABCB1 and ABCG2', Curr Cancer Drug Targets, 17(2), pp. 177-190.

Sheps, J. A., Ralph, S., Zhao, Z., Baillie, D. L. and Ling, V. (2004) 'The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes', *Genome Biol,* 5(3), pp. R15.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol Syst Biol,* 7, pp. 539

Sikosek, T. and Chan, H., 2015. Correction to 'Biophysics of protein evolution and evolutionary protein biophysics'. *Journal of The Royal Society Interface*, 12(112), p.20150915.

Sonett, J., Goldklang, M., Sklepkiewicz, P., Gerber, A., Trischler, J., Zelonina, T., Westerterp, M., Lemaître, V., Okada, Y. and Armiento, J., 2018. A critical role for ABC transporters in persistent lung inflammation in the development of emphysema after smoke exposure. *The FASEB Journal*, 32(12), pp.6724-6736.

Sridharan, S., Robeson, M., Bastihalli-Tukaramrao, D., Howard, C. M., Subramaniyan, B., Tilley, A. M. C., Tiwari, A. K. and Raman, D. (2019) 'Targeting of the Eukaryotic Translation Initiation Factor 4A Against Breast Cancer Stemness', *Front Oncol,* 9, pp. 1311.

Sun, D., Zhang, X., Li, S., Jiang, C. Z., Zhang, Y. and Niu, L. (2016) 'LrABCF1, a GCN-type ATP-binding cassette transporter from Lilium regale, is involved in defense responses against viral and fungal pathogens', *Planta,* 244(6), pp. 1185-1199.

Swainsbury, D. J. K., Scheidelaar, S., Foster, N., van Grondelle, R., Killian, J. A. and Jones, M. R. (2017) 'The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation

of integral membrane proteins from SMA-accessible and SMA-resistant membranes', *Biochim Biophys Acta Biomembr,* 1859(10), pp. 2133-2143.

Tarling, E., Vallim, T. and Edwards, P., 2013. Role of ABC transporters in lipid transport and human disease. *Trends in Endocrinology & Metabolism*, 24(7), pp.342-350.

Taylor, N., Manolaridis, I., Jackson, S., Kowal, J., Stahlberg, H. and Locher, K., 2017. Structure of the human multidrug transporter ABCG2. *Nature*, 546(7659), pp.504-509.

ter Beek, J., Guskov, A. and Slotboom, D. J. (2014) 'Structural diversity of ABC transporters', *J Gen Physiol,* 143(4), pp. 419-35.

The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D506–D515, https://doi.org/10.1093/nar/gky1049

Theodoulou, F.L. & Kerr, I.D., 2015. ABC transporter research: going strong 40 years on. *Biochemical Society transactions*, 43(5), pp.1033–40. Available at: http://www.biochemsoctrans.org/content/43/5/1033.abstract

Tomaszowski, K. H., Schirrmacher, R. and Kaina, B. (2015) 'Multidrug Efflux Pumps Attenuate the Effect of MGMT Inhibitors', *Mol Pharm,* 12(11), pp. 3924-34.

Toyoda, Y., Takada, T. and Suzuki, H. (2019) 'Inhibitors of Human ABCG2: From Technical Background to Recent Updates with Clinical Implications', *Front Pharmacol,* 10, pp. 208.

Wakabayashi, K., Nakagawa, H., Adachi, T., Kii, I., Kobatake, E., Kudo, A. and Ishikawa, T. (2006) 'Identification of cysteine residues critically involved in homodimer formation and protein expression of human ATP-binding cassette transporter ABCG2: a new approach using the flp recombinase system', *J Exp Ther Oncol,* 5(3), pp. 205-22.

Wang, H. et al., 2008. Membrane Topology of the Human Breast Cancer Resistance Protein (BCRP/ABCG2) Determined by Epitope Insertion and Immunofluorescence. , 47(52), pp.13778–13787

Woebking, B., Reuter, G., Shilling, R., Velamakanni, S., Shahi, S., Venter, H., Balakrishnan, L. and van Veen, H., 2005. Drug-Lipid A Interactions on the Escherichia coli ABC Transporter MsbA. *Journal of Bacteriology*, 187(18), pp.6363-6369

Wilkens, S. (2015) 'Structure and mechanism of ABC transporters', *F1000Prime Rep,* 7, pp. 14.

Xavier, B. M., Jennings, W. J., Zein, A. A., Wang, J. and Lee, J. Y. (2019) 'Structural snapshot of the cholesterol-transport ATP-binding cassette proteins', *Biochem Cell Biol,*97(3), pp. 224-233.

Xiong, J., Feng, J., Yuan, D., Zhou, J. and Miao, W. (2015) 'Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily', *Sci Rep,* 5, pp. 16724.

Xu, L., Shi, Y., Zhuang, S. and Liu, N. (2017) 'Recent advances on uric acid transporters', *Oncotarget,* 8(59), pp. 100852-100862

Yang, N. J. and Hinner, M. J. (2015) 'Getting across the cell membrane: an overview for small molecules, peptides, and proteins', *Methods Mol Biol,* 1266, pp. 29-53.

Zein, A. A., Kaur, R., Hussein, T. O. K., Graf, G. A. and Lee, J. Y. (2019) 'ABCG5/G8: a structural view to pathophysiology of the hepatobiliary cholesterol secretion', *Biochem Soc Trans,* 47(5), pp. 1259-1268.

Zhang, W. et al., 2003. The expression and functional characterization of ABCG2 in brain endothelial cells and vessels. *The FASEB journal*, 17, pp.2085–2087. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12958161.

Zor, T. and Selinger, Z. (1996) 'Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies', *Anal Biochem,* 236(2), pp. 302-8.

Appendix- Supplementary figures

Appendix 1: Reagents which have been used to execute the expression, purification and characterisation of ABCG2

- 30% Acrylamide/Bis solution 19:1- *Biorad*
- Adenosine 5'- triphosphate magnesium salt (Mg-ATP) – *Sigma Aldrich*
- Agar – *Formedium*
- Aminoethylbenzenesulfonyl fluoride (AEBSF) – *Sigma Aldrich*
- Ammonium persulphate (APS) – *Fisher Scientific*
- Benzamidine Hydrochloride – *Sigma Aldrich*
- Bestatin – *Sigma-Aldrich*
- Bradford Reagent Concentrated – *Bio-Rad*
- Coomassie Reagent: *instant blueTM - Expedeon*
- Chymostatin – *Sigma-Aldrich*
- Dithiothreitol (DTT) – *Sigma Aldrich*
- Dimethylsulfoxide (DMSO) – *Sigma Aldrich*
- n-Dodecyl-β-D-maltopyranoside (DDM) - *Anatrace*
- Diaminoethanetetra-acetic acid disodium salt dehydrate (EDTA) - *Fisher Scientific*
- Ethanol, Absolute – *Fisher Scientific*
- Epoxysuccinyl-Leucylamido-Butane (E-64) – *Sigma Aldrich*
- GFP-Trap resin - *Chromotek*
- Glass beads acid-washed (426-600 μm - diameter) - *Sigma Aldrich*
- D-Glucose (Glucose) - *Fisher Scientific*
- Glycerol ≥ 99.5% - *Fisher Scientific*
- Glycine – *Fisher scientific*
- Hydrochloric acid (HCl) – *Fisher scientific*
- Imidazole – *Sigma Aldrich*
- K_2HPO_4: di-Potassium hydrogen orthophosphate 3-hydrate - *BDH Laboratory Supplies*
- KH_2PO_4: Potassium dihydrogen orthophosphate - *BDH Laboratory Supplies*
- LB Broth Lennox – *Formedium™*
- Lapatinib Ditosylate – *Provided by Ford's Lab*
- Leupeptin Hydrochloride – *Sigma Aldrich, 1553*
- Methanol - *Fisher Scientific*
- Molecular markers (SDS-PAGE): Pre-stained SDS-PAGE Standards, Broad Range - *BioRad*
- MQ H_2O – *Provided by Ford's Lab* (Millipore Q 0.22 uM filter system)
- Pepstatin A – *Sigma Aldrich*
- PBS: Dublecco's phosphate buffered saline 10x - *Sigma Aldrich*
- Tryptone - *Formedium*
- Sodium chloride (NaCl) - *Fisher Scientific*
- TEMED: N,N,N',N'-Tetramethyl ethylenediamine - *Merck*
- TEV protease - *Sigma Aldrich*
- Tris -Base - *Fisher Scientific*
- Tris -HCL - *Fisher Scientific*
- Yeast Extract powder - *Formedium™*

- Yeast nitrogen base without amino acid - *Formedium™*
- Zeocin™ 100 mg/mL – *Invitrogen*

Appendix 2: Reagents to make the protease inhibitors (PIs):

PIs were added to buffers and made to a 10x dilution when used and stored in aliquots at -20°C.

To obtain a 100x stock solution:

- 144mg of Aminoethylbenzenesulfonyl fluoride (AEBSF)
- 5mg Chymostatin
- 7.5mg E64
- 25mg Leupeptin
- 25mg Pepstatin
- 522mg PMSF
- 6mg Bestatin
- All the above reagents dissolved in 30ml DMSO

To obtain a 1000 x Benzamidine stock solution:

- 360mg Benzamidine was dissolved in 1ml MQ H_2O

Appendix 3: Buffers

Buffer for Cell Harvest and breakage:

- mPiB buffer: 0.25 M sucrose, 0.25 M Tris-HCL pH = 8.0, 1 mM EDTA, 2 mM DTT, cocktail of protease inhibitors (PI) - 10µl of 100 x stock solution
- High Salt Buffer: 50 mM Tris-HCL pH 8.0, 500 mM NaCl, 10% glycerol, cocktail of PI – 10µl of 100 x stock solution.
- Buffer A: 50 mM Tris- HCL pH 8.0, 50 mM NaCl, 10% glycerol, 1mM β-mercaptoethanol. Just before solubilisation add 2% DDM

Buffers for solubilisation:

- Buffer A: 50 mM Tris-HCL pH 8.0, 50 mM NaCl, 10% glycerol, 1mM β-mercaptoethanol. Just before solubilisation add 2% DDM
- Buffer AA: 50 mM Tris-HCL pH = 8.0, 50 mM NaCl, 10% glycerol, 0.1% DDM.

Buffers for SDS PAGE:

- Separating SDS-PAGE gel buffer: 1.5 M Tris-HCL pH = 8.8, 0.4% SDS
- Stacking SDS-PAGE gel buffer: 0.5 M Tris-HCL pH = 6.8, 0.4% SDS

Buffers for Nickle NTA and GFP trap resin purification:

- Buffer A: 50 mM Tris-HCL pH 8.0, 50 mM NaCl, 10% glycerol, 1mM β-mercaptoethanol. Just before solubilisation add 2% DDM
- Buffer AA: 50 mM Tris-HCL pH = 8.0, 50 mM NaCl, 10% glycerol, 0.1% DDM

Buffers for CETSA assay:

- Buffer A (assay buffer): 50 mM Tris-HCL pH 8.0, 50 mM NaCl, 10% glycerol, 1mM β-mercaptoethanol. Just before solubilisation add 2% DDM
- Ice-cold PBS (1L): 0.137M NaCl, 2.7mM KCl, 0.01M Na_2HPO_4, 1.8mM KH_2PO_4 Adjusting the pH to 7.4

Appendix 4: Stock solution

- 2x SDS Loading dye (10ml): 50 mM Tris-HCl pH = 7.6, 5% glycerol, 5mM EDTA pH = 8.0, 0.02% bromophenol blue, 4% SDS, 50mM DTT
- DTT: 0.5M stocks stored at -20 ºC
- SDS: 20% stock stored on the shelf at room temperature

Appendix 5: SDS-PAGE gels

2 X 1.5mm – 8% Gel

Separating	Reagents	Stacking
8.3ml	**H_2O**	6.2ml
5ml	**Buffer**	2.5ml
6.7ml	**30% Acrylamide**	1.3ml
40ul	**TEMED**	20ul
100ul	**20% APS**	50ul

2 x 1.5mm – 10% Gel

Separating	Reagents	Stacking
9.6ml	**H_2O**	6.2ml
5ml	**Buffer**	2.5ml
5.4ml	**30% Acrylamide**	1.3ml
40ul	**TEMED**	20ul
100ul	**20% APS**	50ul

Separating Buffer Gel	Stacking Buffer Gel
1.5M Tris-HCl pH 8.8	0.5M Tris-HCl pH 6.8
0.4% SDS	0.4% SDS

Appendix 6: Equipment

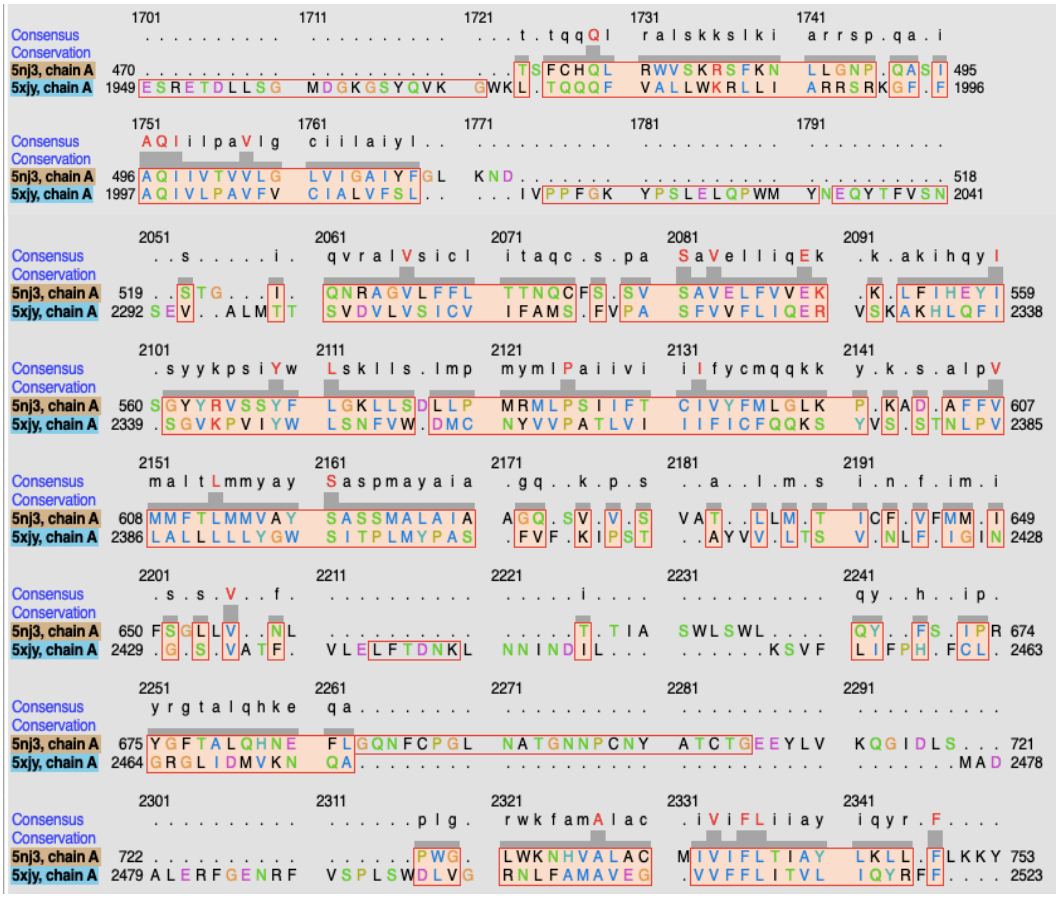
- Autoclave
- BeadBeater cell disrupter
- Centrifuges and centrifuge tubes
- ChemiDoc MP gel imager
- Computer with eclipse programme – used for detecting the intensity of GFP fluorescence
- Fluorimeter
- 50ml Falcon tubes
- 1L Flasks
- pH meter
- Orbital Shaking incubator (for fermentation of *Pichia Pastoris*)
- Scanner
- 1cm Quartz Cuvette

Appendix 7: The quantitative data for the overall structural alignments between the different ABC transporters.

Alignment	RMSD (Å)	SDM	Q score
ABCA1 TMD1: ABCG2 TMD	2.201	44.827	0.401
5WS4: ABCG2 TM1-4	2.316	51.918	0.364
5LIL:ABCG2 TM1-4	2.309	50.907	0.227
5NIK: ABCG2 TM1-4	2.476	56.224	0.388
5WS4:ABCG8 TM1-4	2.408	47.909	0.289
5LIL:ABCG8 TM1-4	2.445	56.907	0.286
5NIK:ABCG8 TM1-4	2.452	58.425	0.156
5WS4:ABCA1 TM1-4	2.330	58.922	0.264
5NIK:ABCA1 TM1-4	2.534	56.931	0.302
MlaE: ABCG2 TMD	2.570	59.362	0.302
MlaE: ABCG8 TMD	2.585	51.375	0.325
MlaE: ABCA1 TMD	2.687	65.460	0.268
Wzm:ABCG2 TMD	2.438	51.994	0.390
Wzm: ABCG8 TMD	2.643	52.728	0.230
Wzm: ABCA1 TMD	2.386	51.624	0.326

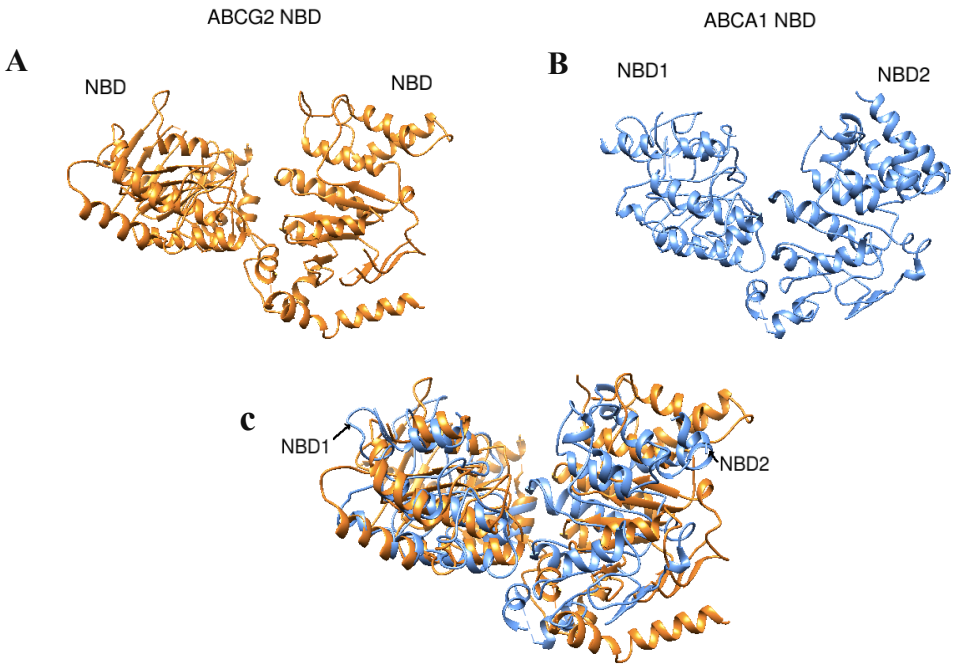
Appendix 8: The structural sequence alignment for the alignment between ABCA1 TMD1 and ABCG2 TMD

This structural superimposed sequence alignment generated using Match and Align showed overall RMSD as 2.201 across the sequence length 266 246, SDM score 44.635 and a Q score of 0.401. The residues associated with the transmembrane structural alignment between ABCA1 TMD2 and ABCG2 TMD are as followed: ABCA1 IH3/ ABCG2 TM1a = 1721-1744, ABCA1 TM7/ ABCG2 TM1b = 1746-1770, ABCA1 TM8/ ABCG2 TM2 =2041-2091, ABCA1 TM9/ ABCG2 TM3 = 2101-2140 ABCA1 TM10/ ABCG2 TM4 =2141-2201, ABCA1 TM11/ ABCG2 TM5a =2176-2200, ABCA1 EH4/ ABCG2 TM5c =2245-2307, ABCA1 TM12/ ABCG2 TM6b =2311-2350.



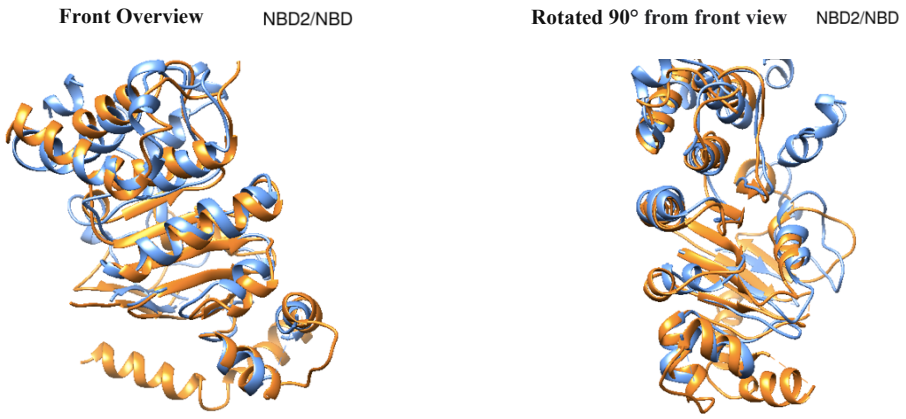
Appendix 9: The structural alignment between ABCG2 NBD and ABCA1 NBD1

Making an observation with the visible eye, ABCG2 NBD would have been expected to align significantly to ABCA1 NBD1 rather than NBD2.Furthermore the structural alignment showed that the alignment between ABCA1 NBD1 and ABCG2 NBD gave an overall RMSD being 1.994, SDM resulted as 48.441, Q score as 0.363.



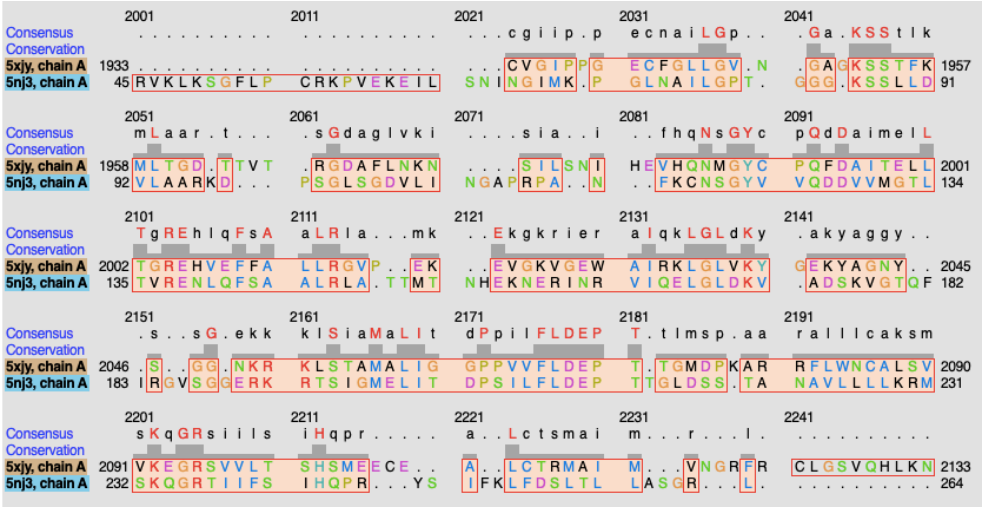
Appendix 10: **The structural alignment between ABCG2 NBD and ABCA1 NBD2**

The superimposed sequence alignment between ABCA1 NBD2 and ABCG2 NBD showed the overall RMSD being 2.271, SDM resulted as 43.903, Q score as 0.303. Overall, the results showed that ABCA1 NBD1 aligned better structurally and through sequence alignment with ABCG2 NBD, compared to the alignment between ABCA1 NBD2 and ABCG2 NBD. This significant alignment of the ABCA1 NBD and ABCG2 NBD validates Xiong et al and Ford et al theory of having highly similar NBD structure.



Appendix 11: The structural sequence alignment between ABCA1 NBD1 and ABCG2 NBD

The structural sequence alignment between ABCA1 NBD1 and ABCG2 NBD gave 44 fully conserved residues alongside 127 weak conserved groups between 2001-2250 residues, overall RMSD as 1.994, SDM resulted as 48.441, Q score as 0.363. These results showed high structural similarity, hence supporting the structural alignment.

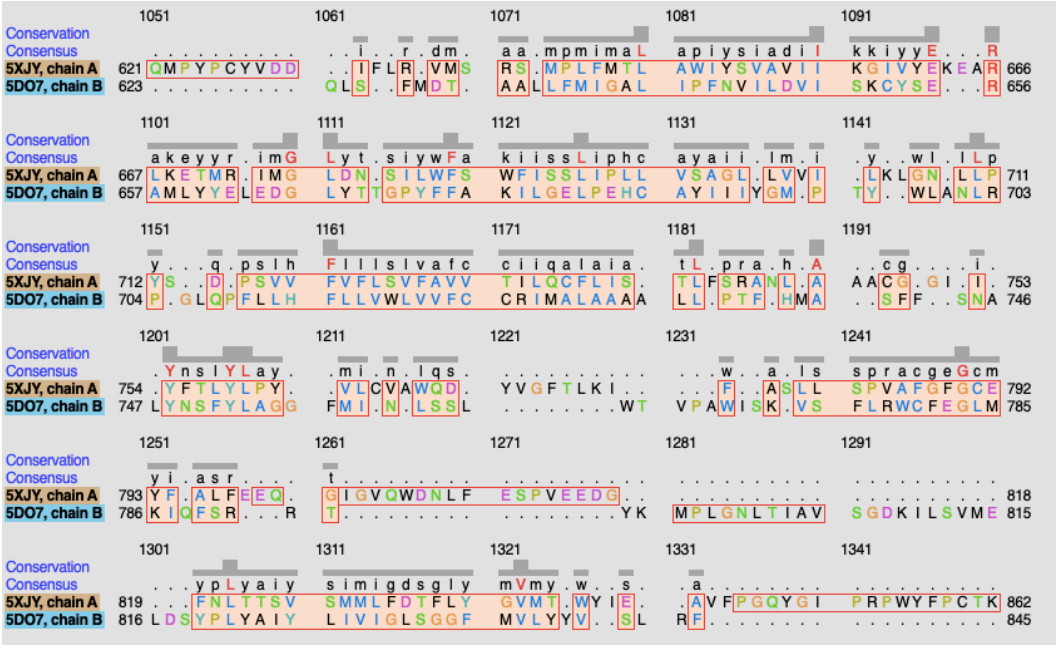


Appendix 12: The structural sequence alignment between ABCA1 NBD2 and ABCA1 NBD

The structural sequence alignment between ABCA1 NBD2 and ABCG2 NBD displayed 20 fully conserved residues alongside 160 weak conserved groups between 951-1300 residues. These results showed high structural similarity hence supporting the structural alignment

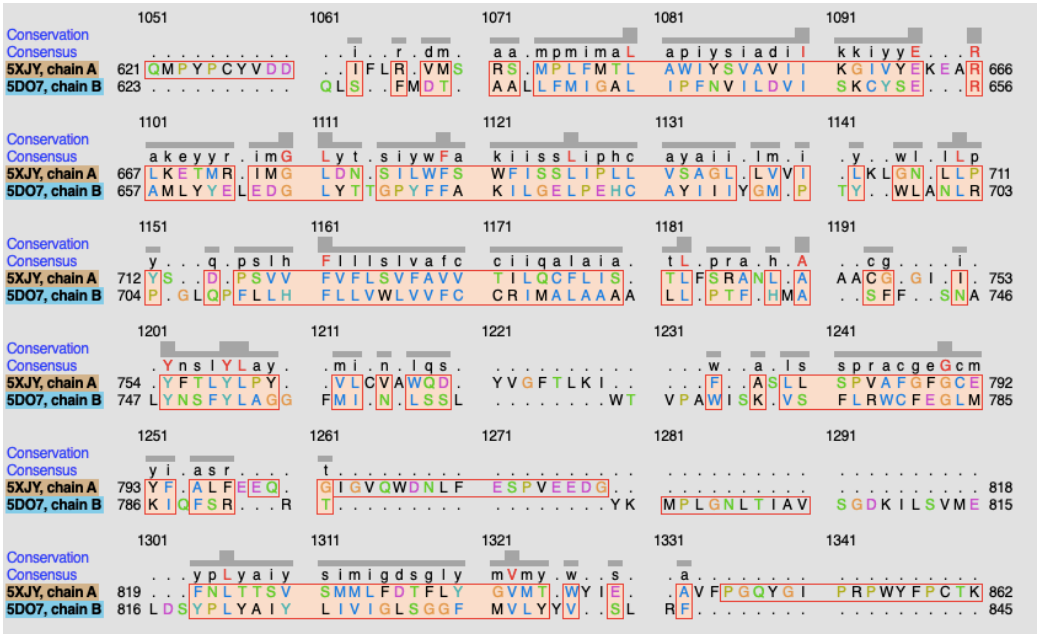
Appendix 13: The structural sequence alignment between ABCG8 TMD and ABCA1 TMD1

The structural sequence alignment between ABCA1 TMD1 and ABCG8 TMD gave an the overall RMSD was 2.264 across sequence length of 266248, the SDM score as 46.253 and the Q score as 0.453Q These results showed high structural similarity, hence supporting the structural alignment.



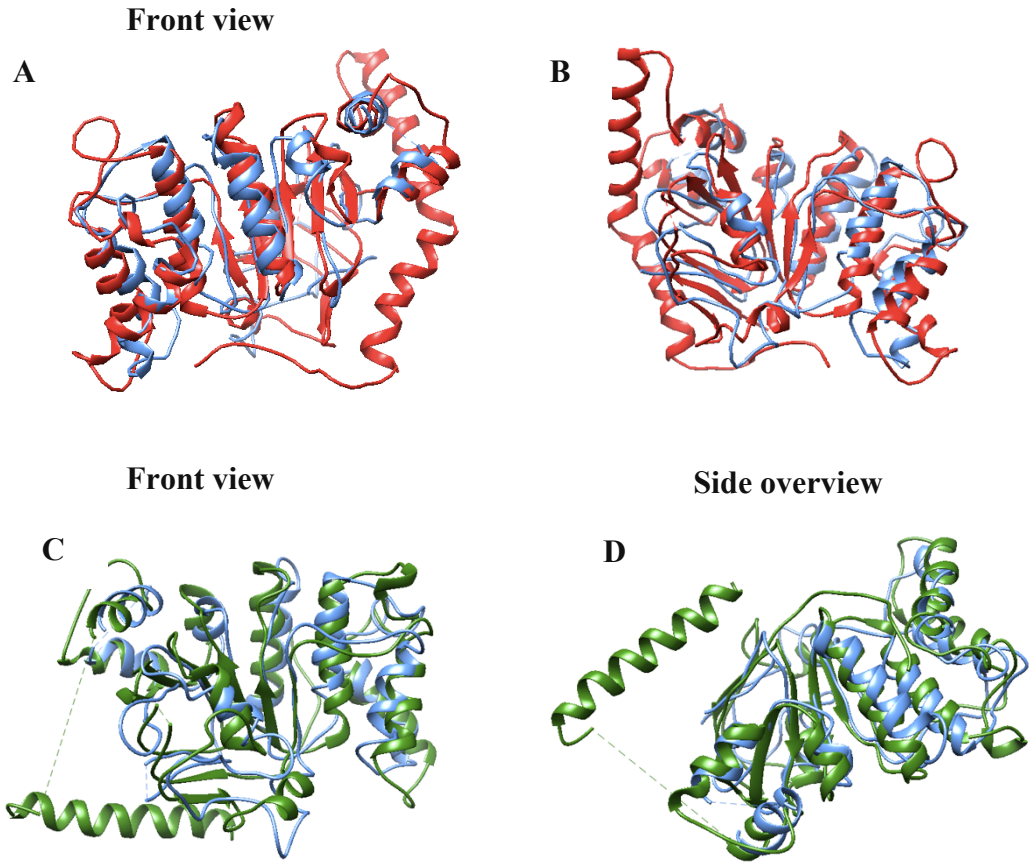
Appendix 14: The structural sequence alignment ABCG5 TMD and ABCA1 TMD1

As for the alignment of the TM helices of ABCG5 TMD and ABCA1 TMD1 the overall RMSD was 2.428 Å; SDM value was 41.883 and Q score being 0.299



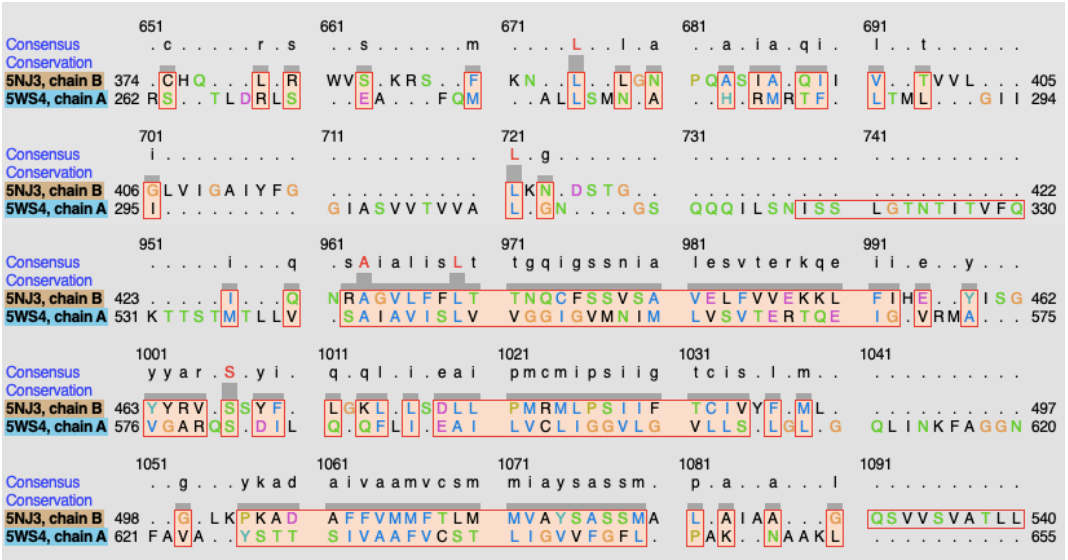
Appendix 15: The structural alignment between ABCA1 NBD1 and ABCG5/G8 NBD

The structure of ABCG5 NBD is represented in red., The structure of ABCG8 TMD is represented in forest green, ABCA1 TMD1 is represented in cornflour blue. **A,B**) The structural alignment of ABCG5 NBD and ABCA1 NBD1. **C,D)** The structural alignment of ABCG8 NBD and ABCA1 NBD1. This figure illustrates the similarity in the structure of ABCA1 NBD1 and ABCG5G8 NBDs, as the α helices and β pleated sheets show significant alignments.



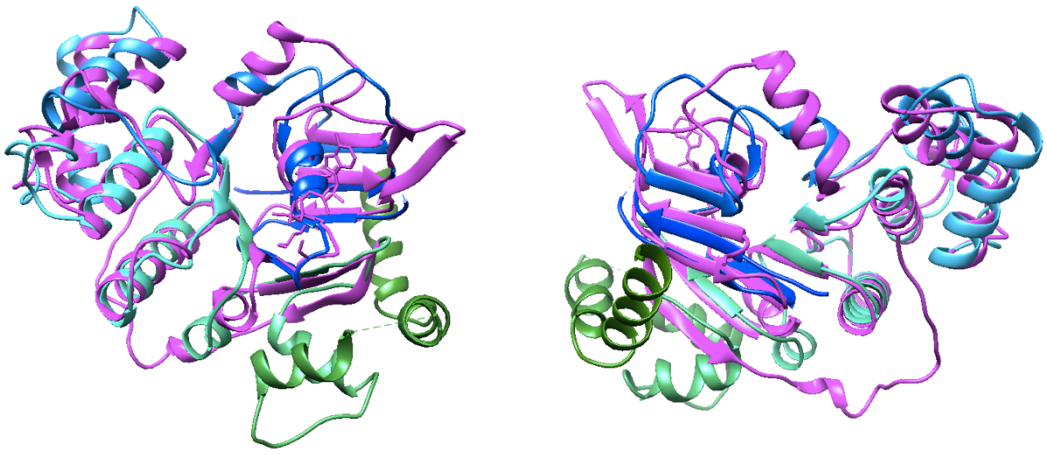
Appendix 16: The structural sequence alignment between ABCG2 TM1-4 and 5WS4 TMD

The structural sequence alignment between ABCG2 TM 1-4 and 5WS4 showed overall RMSD of 2.316, SDM of 51.918 and Q score of 0.248, using Clustal omega alignment the percent identity worked out to be 19.28%.



Appendix 17: The structural alignment of ABCG2 NBD and 5WS4

The structure of ABCG2 NBD is represented in green, blue, cyan and cornflour blue ribbons. The structure of 5WS4 NBD is represented in magenta. The structural alignments show significant overlap of the α helices and β pleated sheets show significant alignments.



Front view **Back view**

Appendix 18:The structural sequence alignment between 5WS4 NBD and ABCG2 NBD

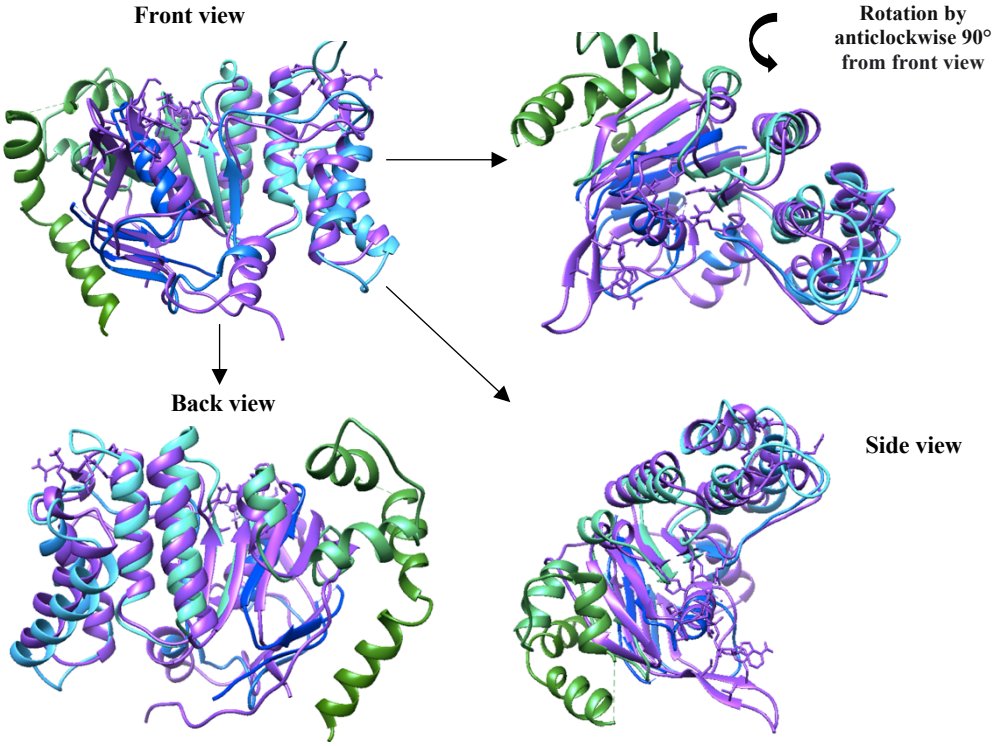
The structural sequence alignment between 5WS4 NBD and ABCG2 NBD overall RMSD as 1.930, SDM as 40.556, Q score as 0.394 and the percentage identity worked out to be 20.01%.

Appendix 19:The structural sequence alignment between 5LIL TMD and ABCG2 TM1-4

The structural sequence alignment between 5LIL TMD and ABCG2 TM1-4 showed overall RMSD as 2.309, SDM as 50.907, Q score as 0.364 and the percentage identity worked out to be 15.36%.

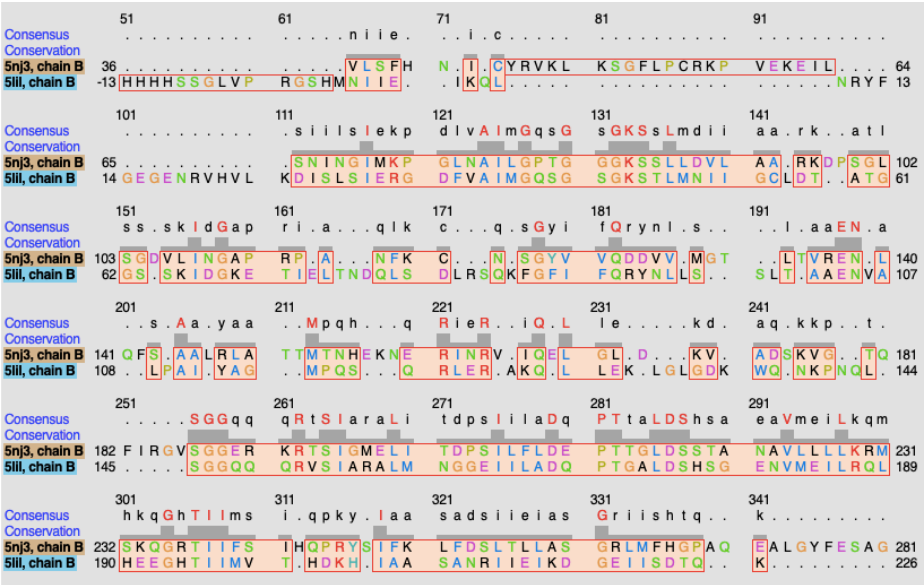
Appendix 20:The structural alignment of 5LIL NBD and ABCG2 NBD

The structure of ABCG2 NBD is represented in green, blue, cyan and cornflour blue ribbons. The structure of 5LIL NBD is represented in purple. The structural alignments show significant overlap of the α helices and β pleated sheets indicating high structural similarity.



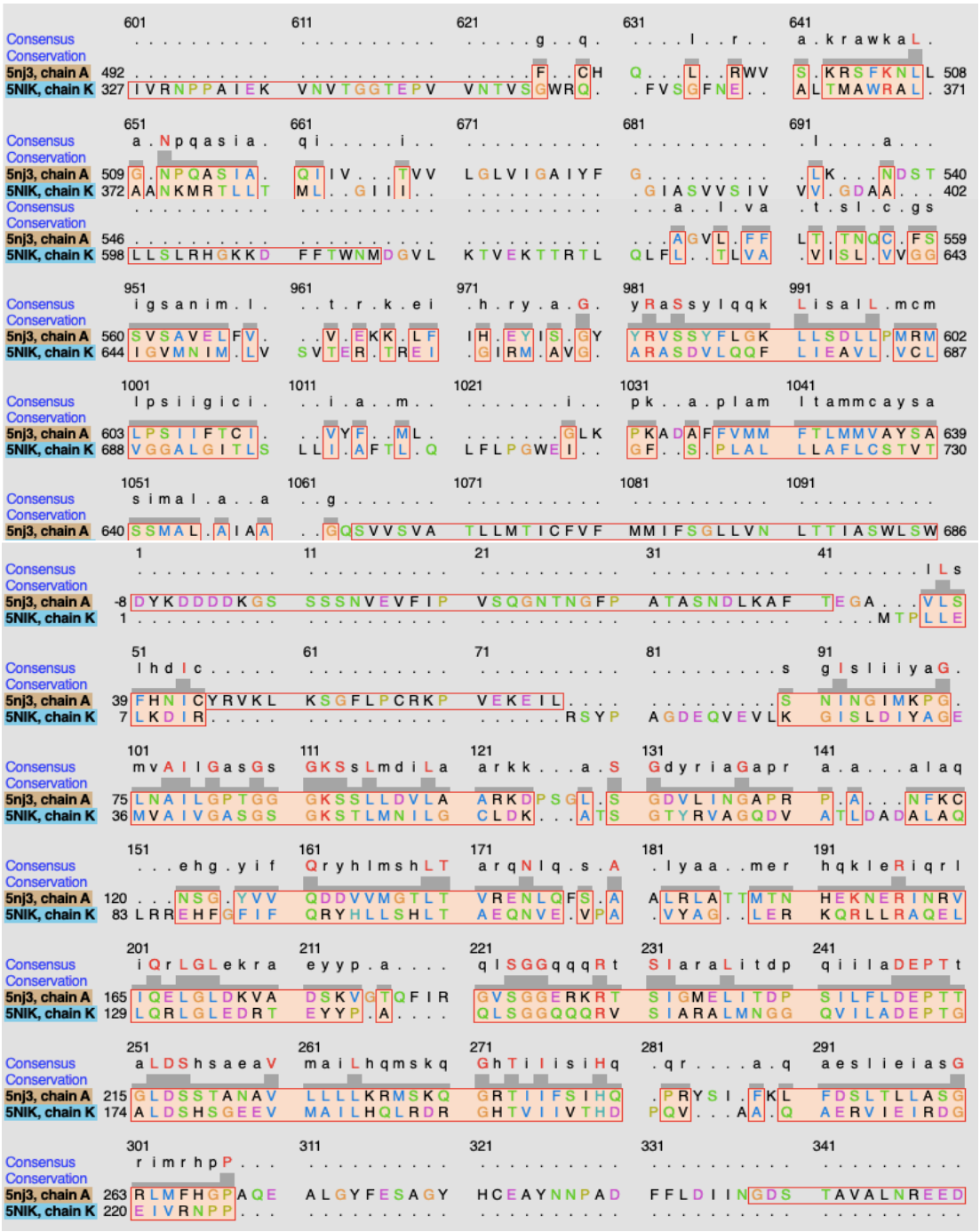
Appendix 21: The structural sequence alignment of 5LIL NBD and ABCG2 NBD

The structural sequence alignment between 5LIL NBD and ABCG2 NBD showed overall RMSD as 2.413, SDM as 49.304, Q score as 0.329.



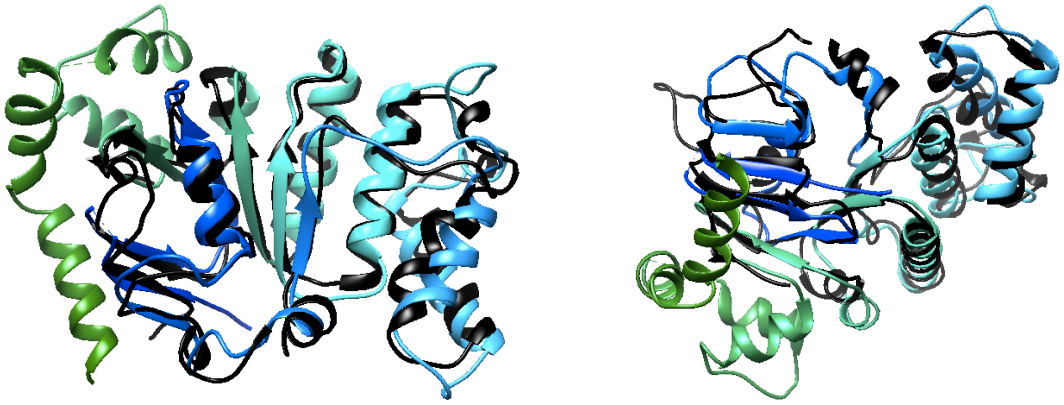
Appendix 22:The structural sequence alignment between 5NIK TMD and ABCG2 TM1-4

The structural sequence alignment between 5NIK TMD and ABCG2 TM1-4 showed overall RMSD as 2.476, SDM as 56.229, Q score as 0.227 and the percentage identity as 16.82%.



Appendix 23: The structural alignment between 5NIK NBD and ABCG2 NBD

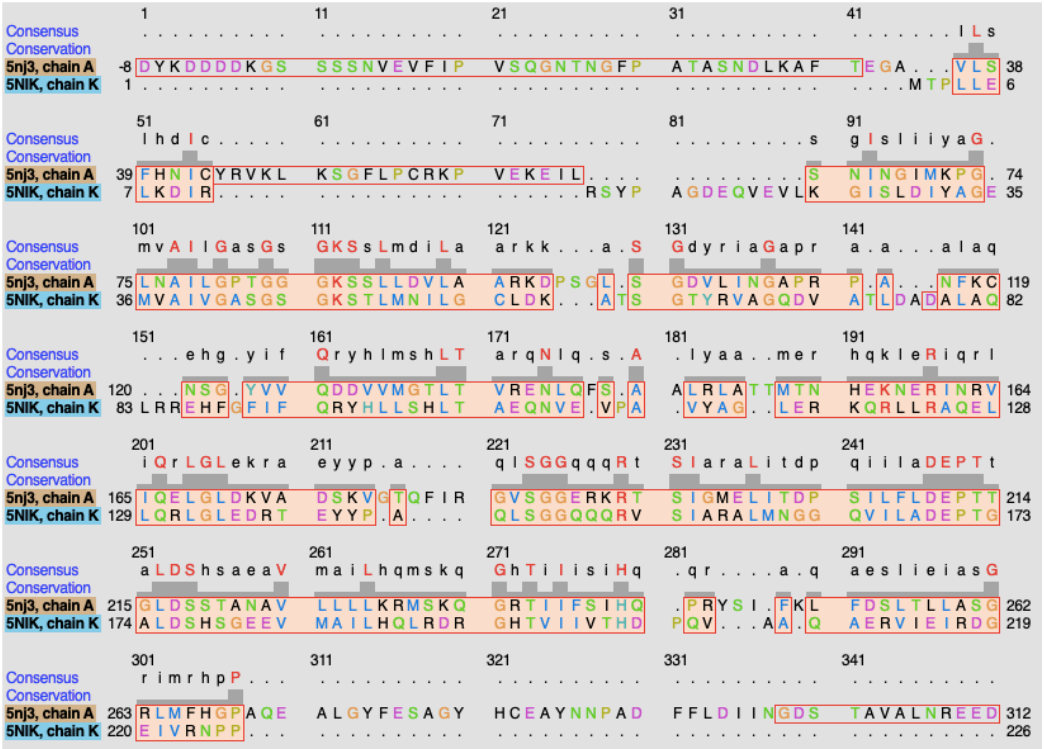
The structure of ABCG2 NBD is represented in green, blue, cyan and cornflour blue ribbons. The structure of 5NIK NBD is represented in black. The structural alignments show significant overlap of the α helices and β pleated sheets indicating high structural similarity.



Front view **Side overview**

Appendix 24: The structural alignment between 5NIK NBD and ABCG2 NBD

The structural sequence alignment between 5NiK NBD and ABCG2 NBD showed overall RMSD was 1.921, SDM as 38.435, Q score as 0.438 and percentage identity of 16.82%.

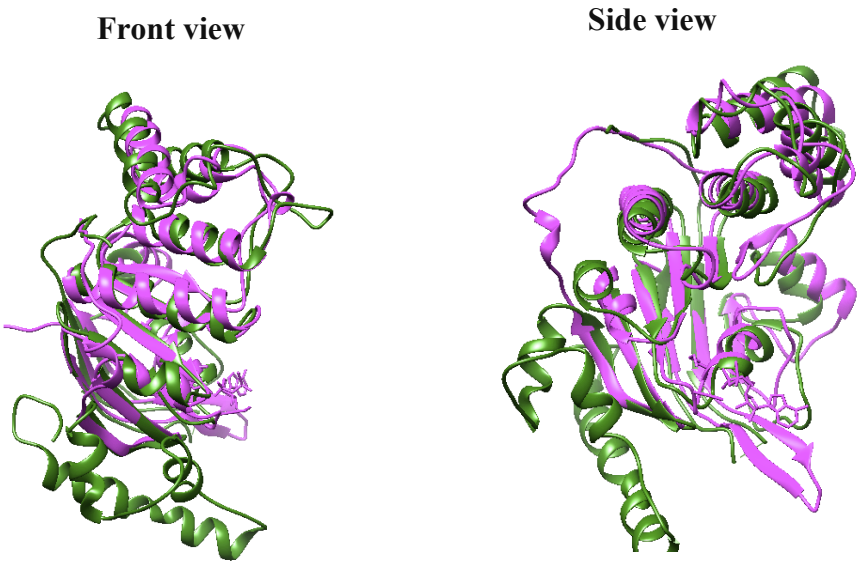


Appendix 25:The structural sequence alignment of ABCG8 TM1-4 and 5WS4

The structural sequence alignment between 5WS4 TMD and ABCG8 TM1-4 showed overall RMSD as 2.408, SDM as 57.909, Q score as 0.388 and the percentage identity as 15.95%.

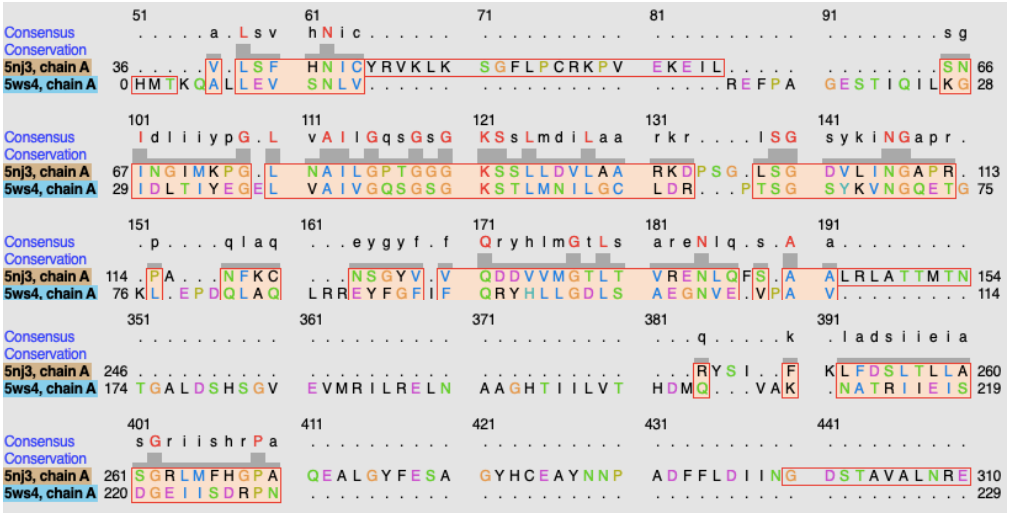
Appendix 26 The structural alignment between ABCG8 NBD and 5WS4 NBD

The structure of ABCG8 NBD is represented in forest green and the structure of 5WS4 NBD is represented in magenta. The structural alignments show significant overlap of the α helices and β pleated sheets indicating high structural similarity



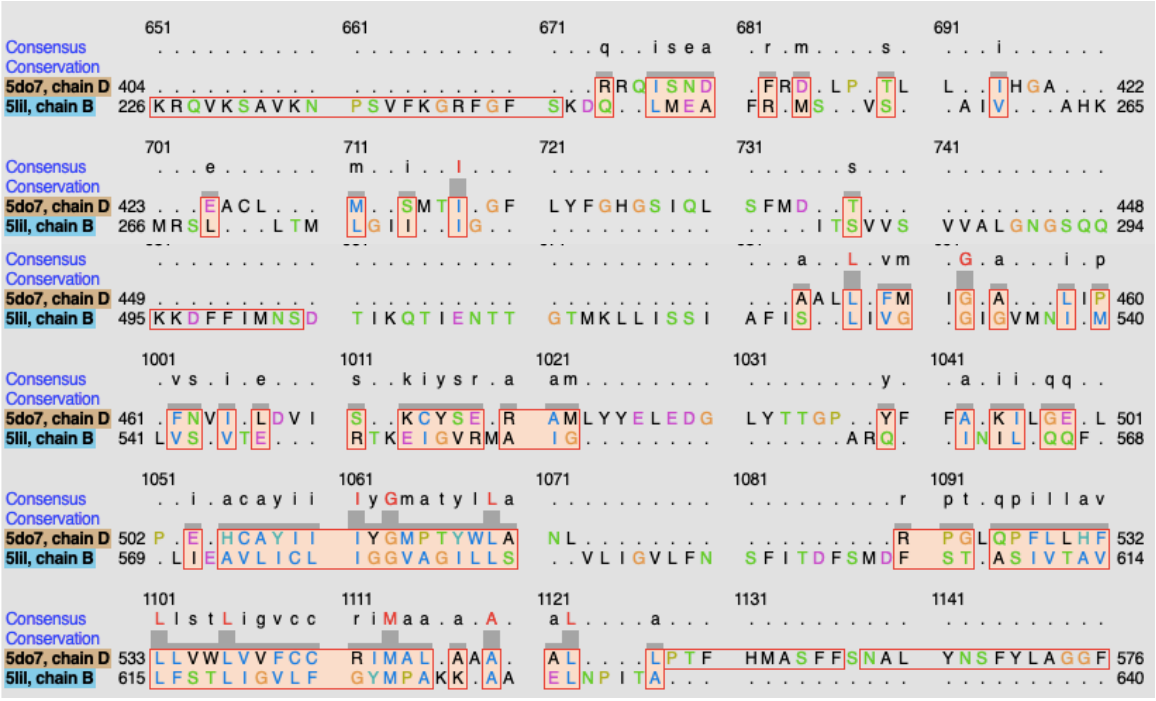
Appendix 27 The structural sequence alignment between ABCG8 NBD and 5WS4 NBD

The structural sequence alignment between 5WS4 NBD and ABCG8 NBD showed overall RMSD as 2.268, SDM as 48.933, Q score as 0.381 and the percentage identity as 16.82%.



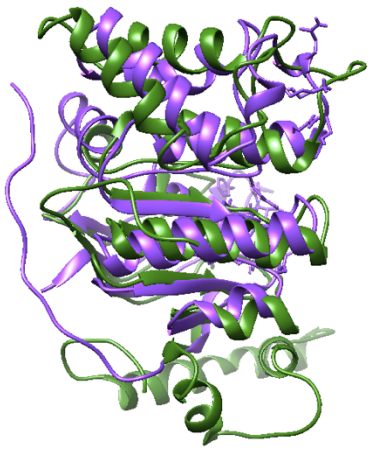
Appendix 28:The structural sequence alignment between ABCG8 TM1-4 and 5LIL TMD

The structural sequence alignment between 5LIL TMD and ABCG8 TM1-4 overall RMSD 2.445, SDM was 56.907, Q score was 0.289

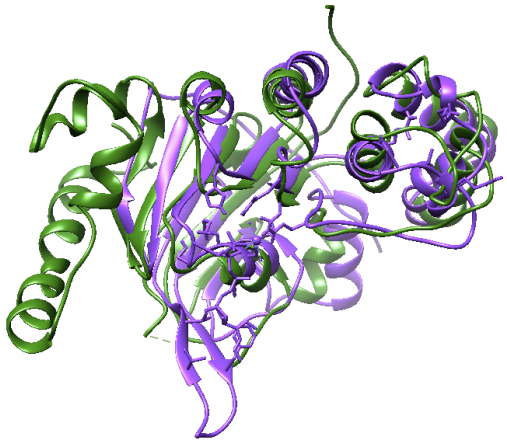


Appendix 29: The structural alignment between ABCG8 NBD and 5LIL NBD

The structure of ABCG8 NBD is represented in forest green and the structure of 5LIL NBD is represented in purple. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.



Front view



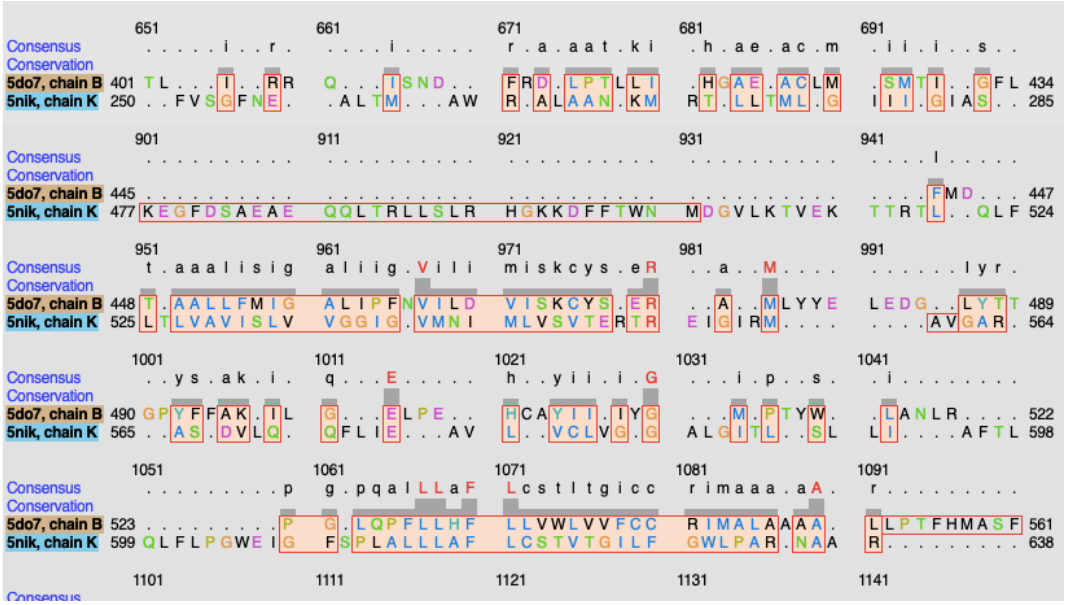
Side view

Appendix 30: The structural sequence alignment between ABCG8 NBD and 5LIL NBD

The structural sequence alignment between 5LIL NBD and ABCG8 NBD overall RMSD of 2.167, SDM was 47.210 and Q score of 0.289.

Appendix 31:The structural alignment of ABCG8 TM1-4 and 5NIK TMD

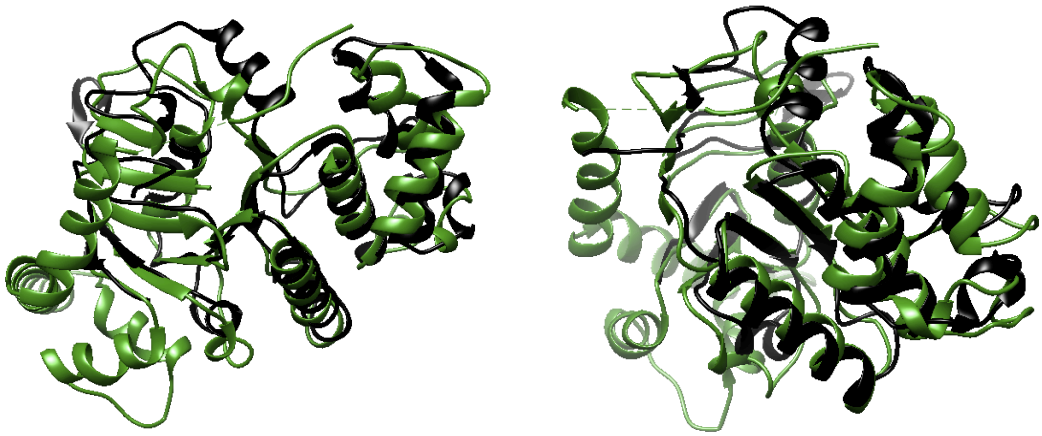
The structural sequence alignment between 5NIK TMD and ABCG8 TM1-4 overall RMSD was 2.452, SDM was 58.425, Q score was 0.286 and percentage identity was 16.82%.



Appendix 32:The structural alignment between ABCG8 NBD and 5NIK NBD

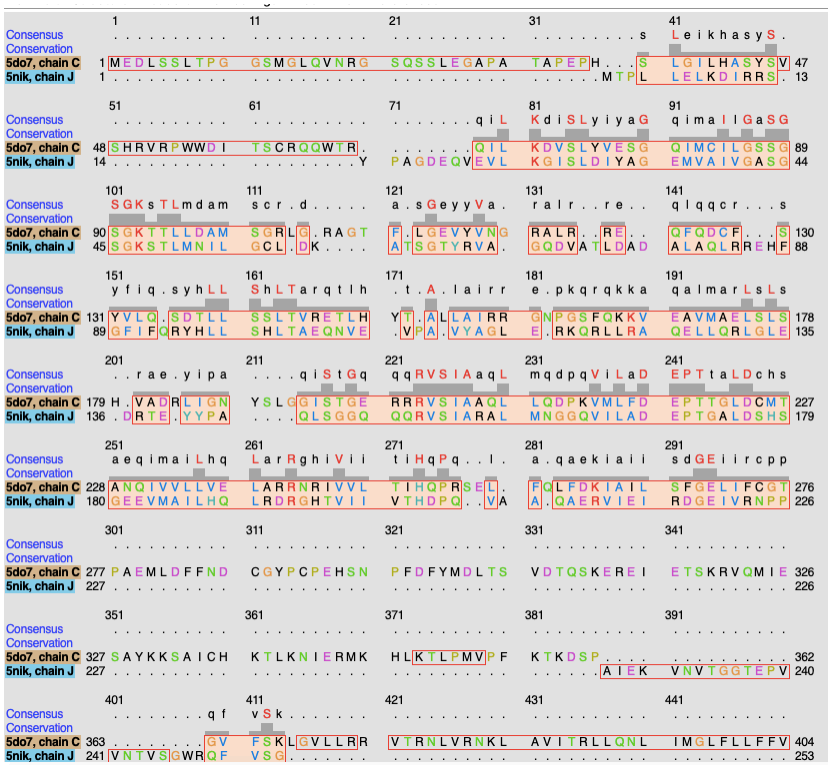
The structure of ABCG8 NBD is represented in forest green and the structure of 5NIK NBD is represented in black. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front view **Side view**



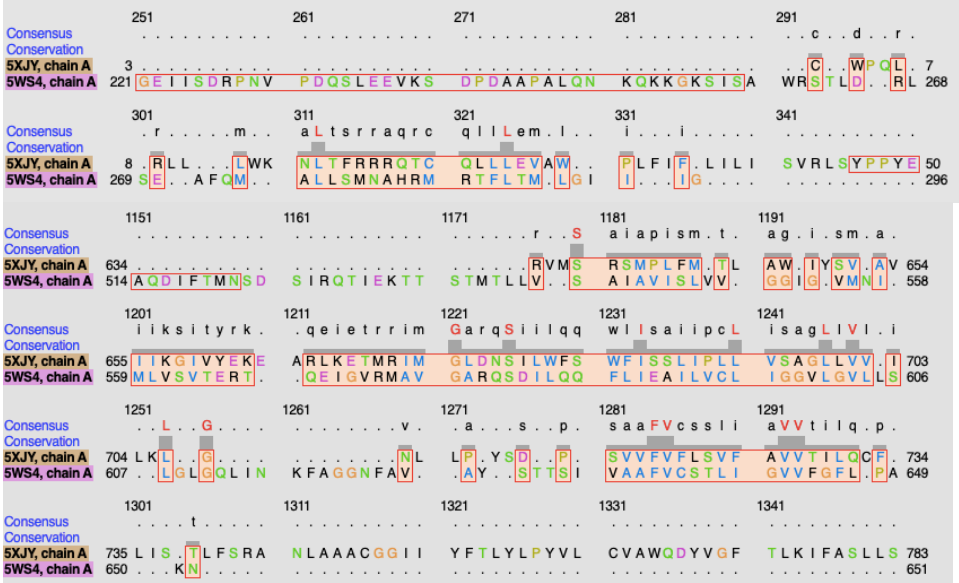
Appendix 33: The structural sequence alignment between ABCG8 NBD and 5NIK NBD

The structural sequence alignment between 5NIK NBD and ABCG8 NBD overall RMSD of 2.196, SDM was 42.937, Q score of 0.398 and percentage identity of 17.13%.



Appendix 34: The structural sequence alignment between ABCA1 TMD1 and 5WS4 TMD

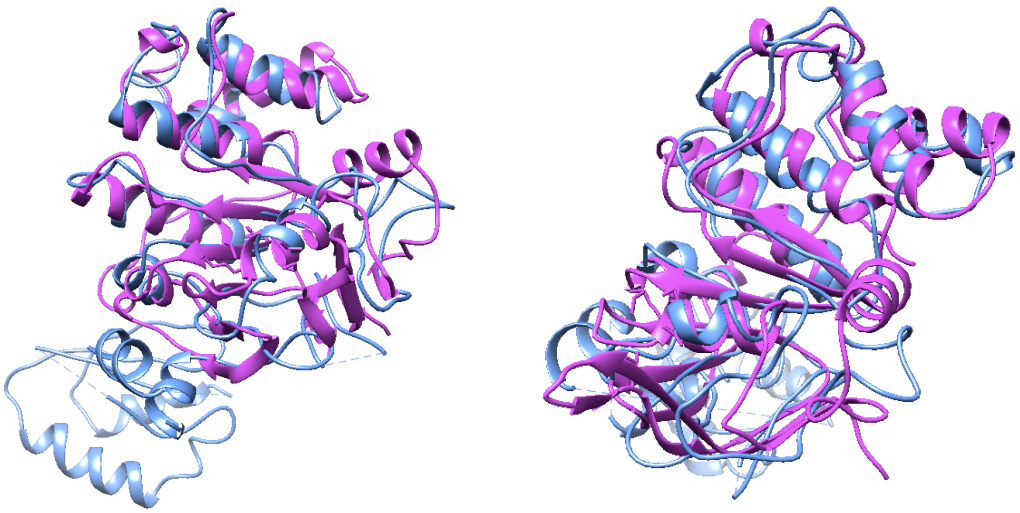
The structural sequence alignment between ABCA1 TMD1 and 5WS4 TMD overall RMSD was 2.330, SDM was 58.922, Q score of 0.156.



Appendix 35:The structural alignment between ABCA1 NBD1 and 5WS4 NBD

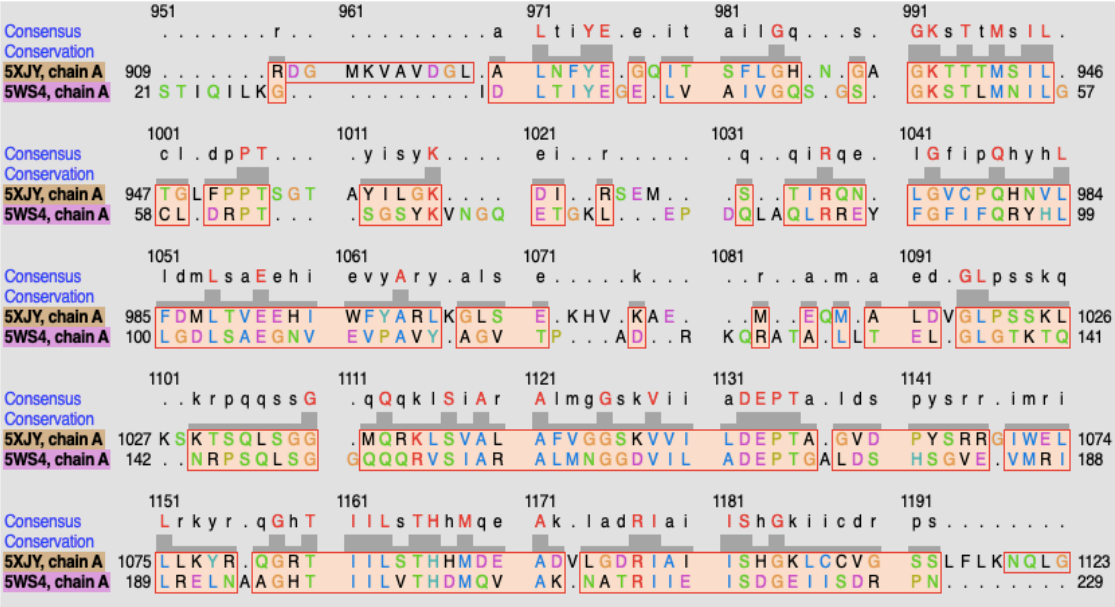
The structure of ABCA1 NBD is represented in cornflour blue and the structure of 5WS4 NBD is represented in magenta. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front view **Side view**



Appendix 36:The structural sequence alignment between ABCA1 TMD1 and 5WS4 TMD

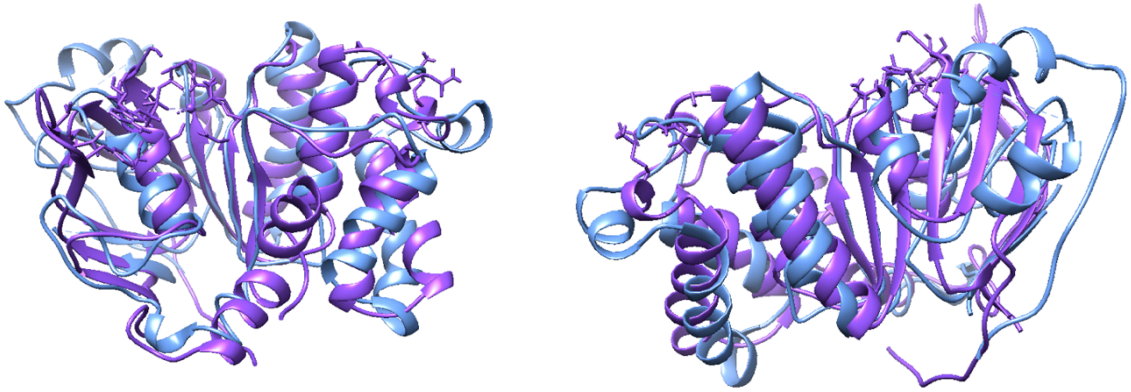
The structural sequence alignment between ABCA1 TMD1 and 5WS4 TMD overall RMSD was 2.186, SDM was 47.986, Q score was 0.390.



Appendix 37: The structural alignment between ABCA1 NBD1 and 5LIL NBD

The structure of ABCA1 NBD is represented in cornflour blue and the structure of 5LIL NBD is represented in purple. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front view **Side view**

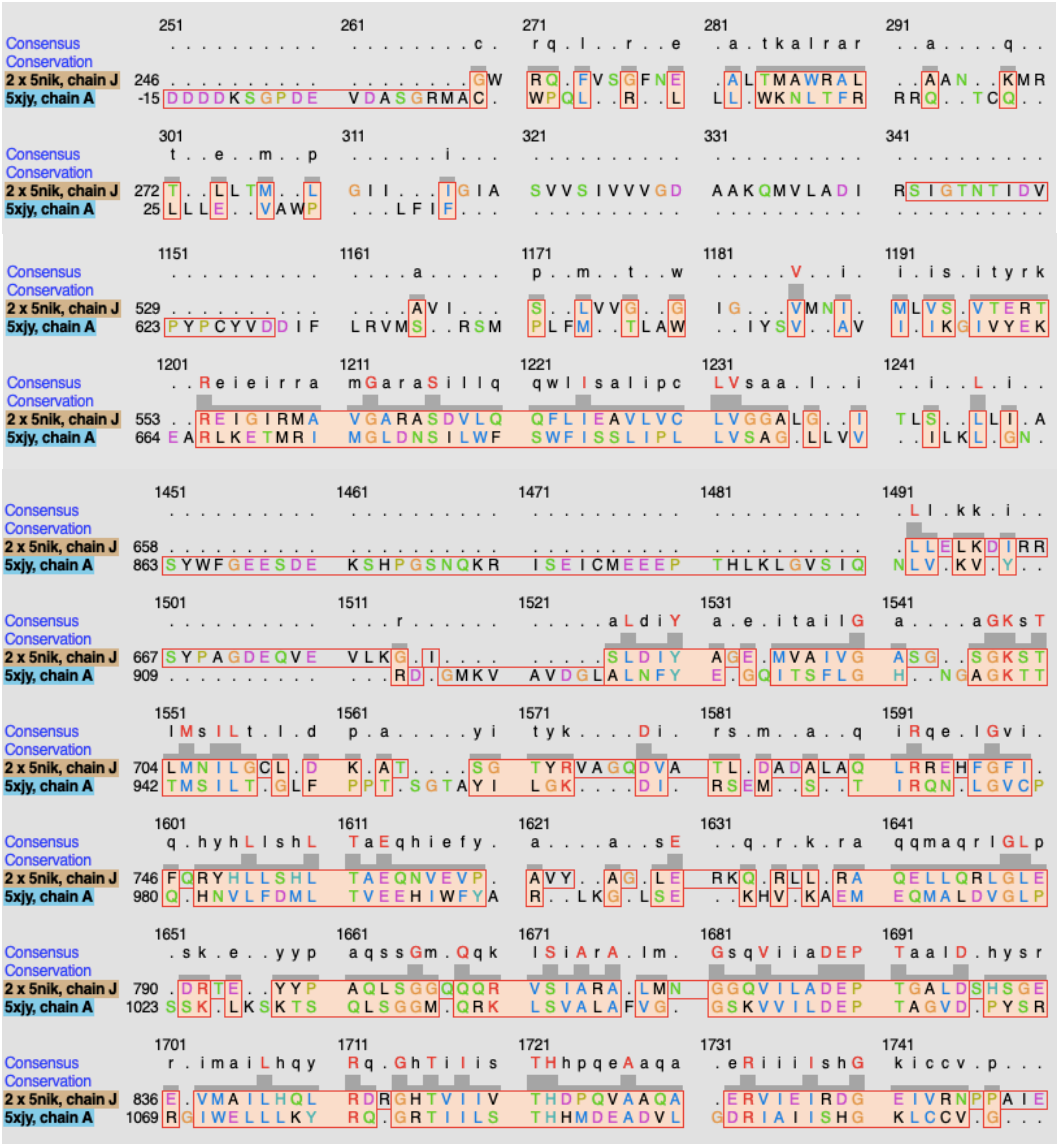


Appendix 38: The structural alignment between ABCA1 NBD1 and 5LIL NBD

The structural sequence alignment between ABCA1 NBD1 and 5WS4 NBD overall RMSD 2.004, SDM was 41.460 and Q score was 0.435.

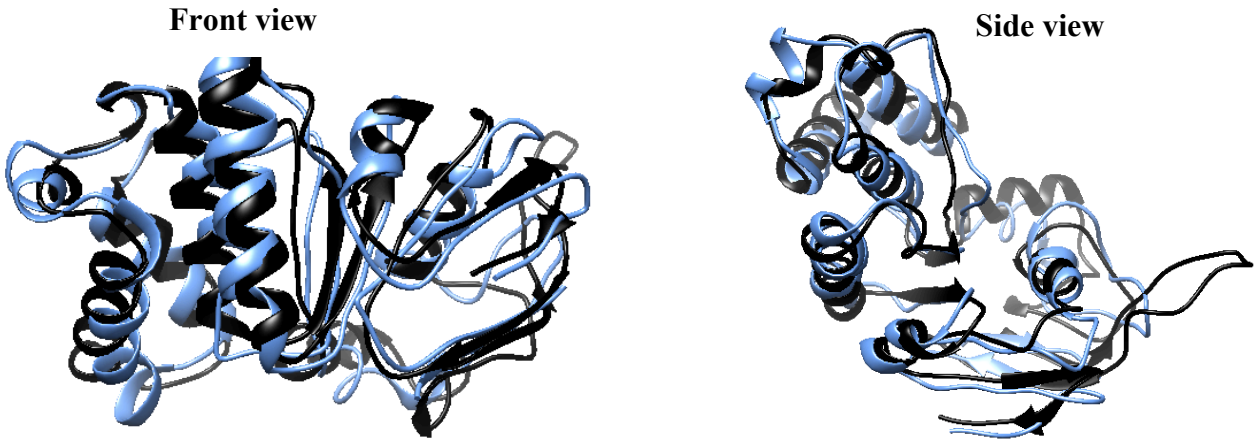
Appendix 39:The structural sequence alignment between ABCA1 TMD1 and 5NIK TMD

The structural sequence alignment between ABCA1 TMD1 and 5NIK TMD overall RMSD was 2.387, SDM was 50.264, Q score of 0.146.



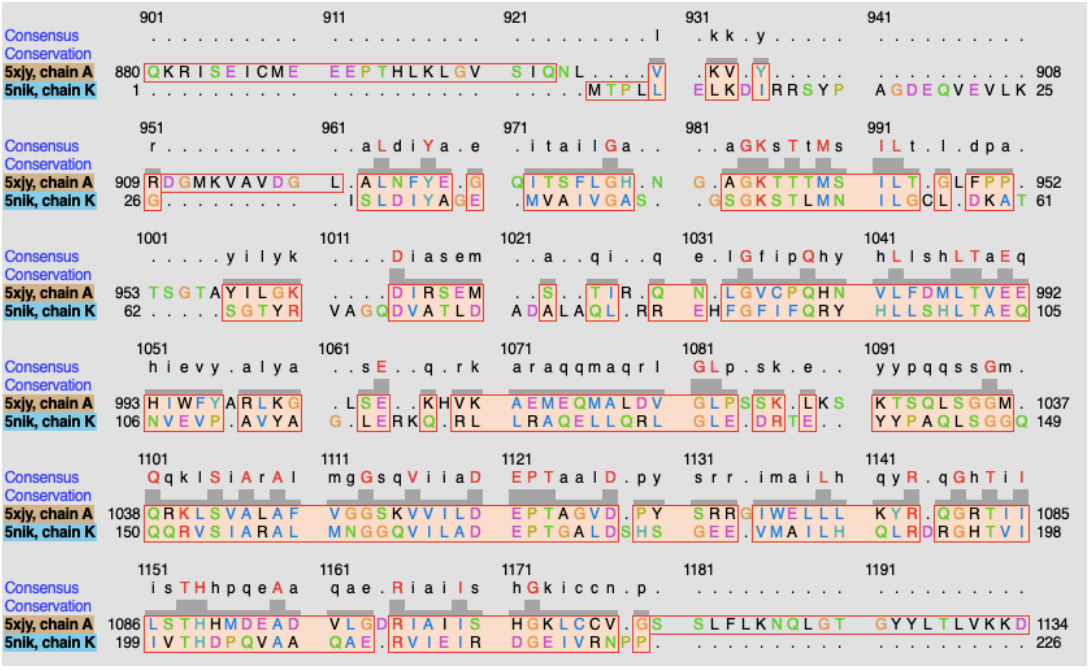
Appendix 40:The structural alignment between ABCA1 NBD1 and 5NIK NBD

The structure of ABCA1 NBD is represented in cornflour blue and the structure of 5NIK NBD is represented in black. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity



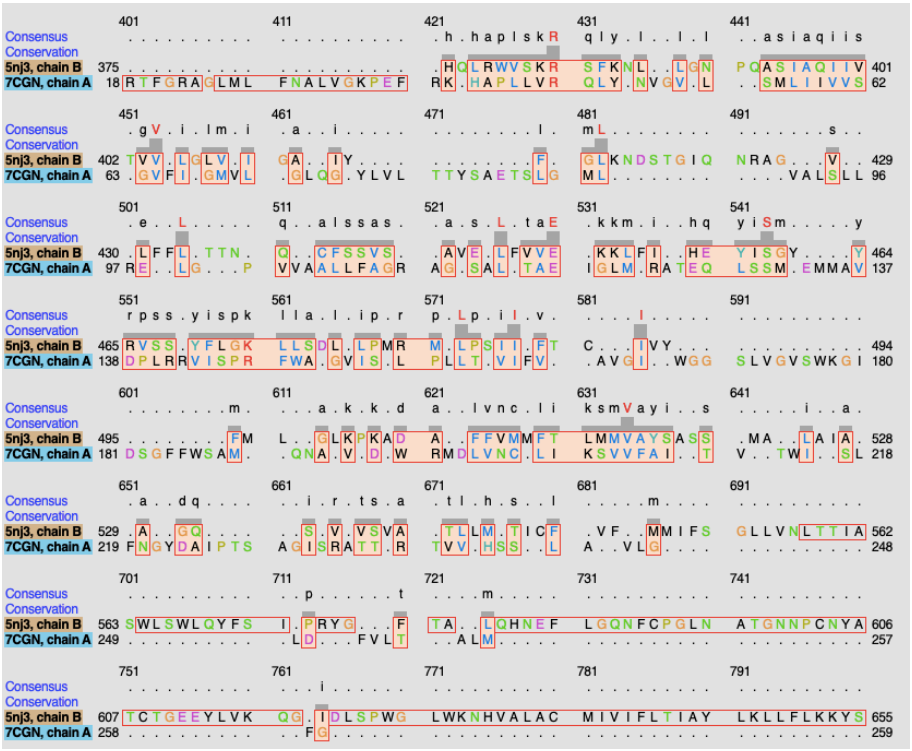
Appendix 41:The structural sequence alignment between ABCA1 NBD1 and 5NIK NBD

The structural sequence alignment between ABCA1 NBD1 and 5NIK NBD overall RMSD was 2.004, SDM was 41.460, Q score of 0.435.



Appendix 42: The structural sequence alignment between ABCG2 TM1-5 and MlaE.

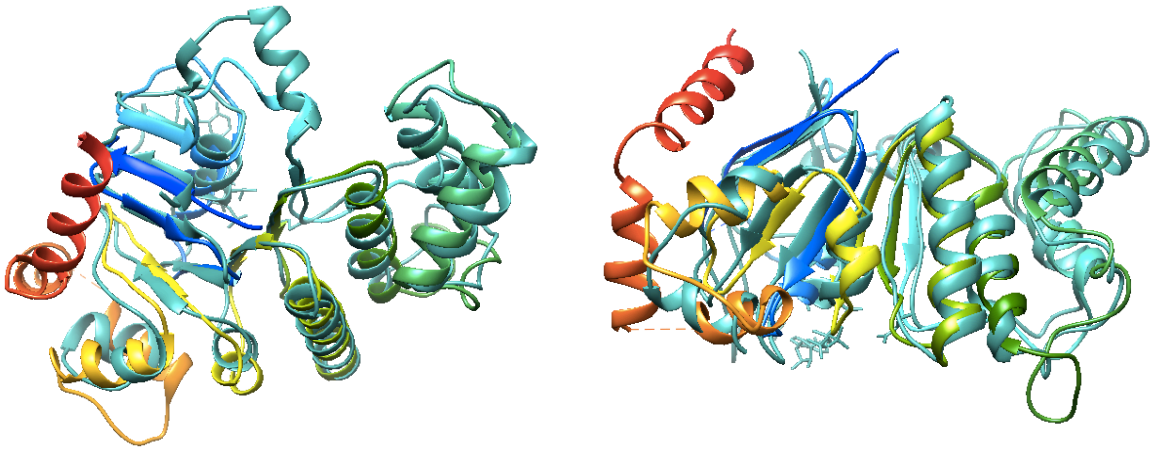
The structural sequence alignment between ABCG2 TMD1-5 and MlaE overall RMSD was 2.570, SDM was 59.362, Q score of 0.302.



Appendix 43:The structural alignment between ABCG2 NBD and MlaF

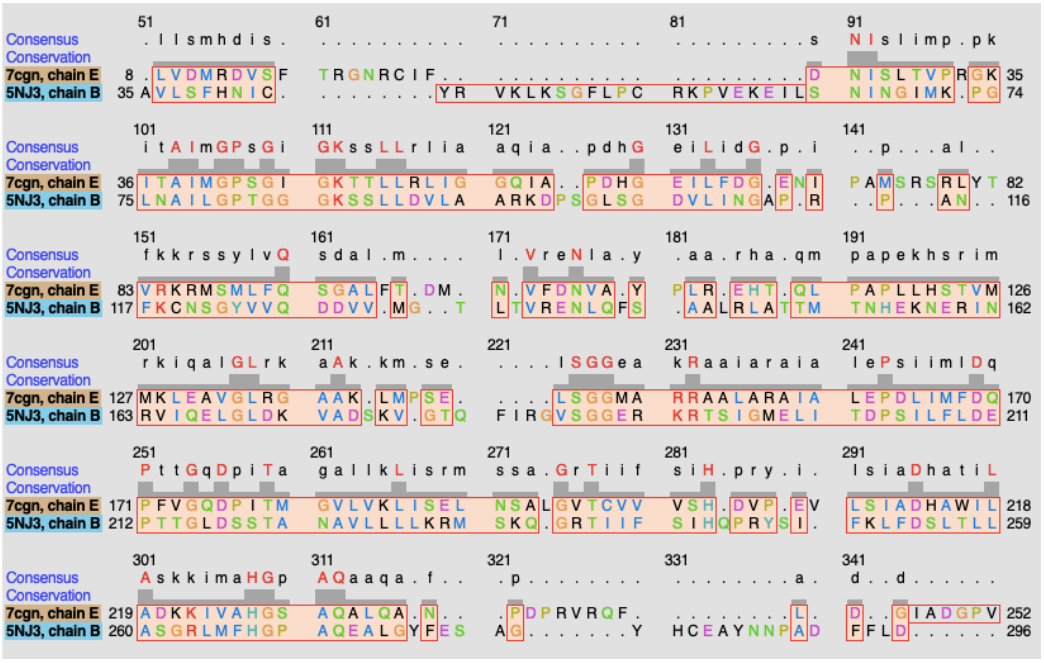
The structure of MlaF is represented in light sea green. ABCG2 NBD is represent in green, yellow, red and blue ribbons. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front overview **Side overview**



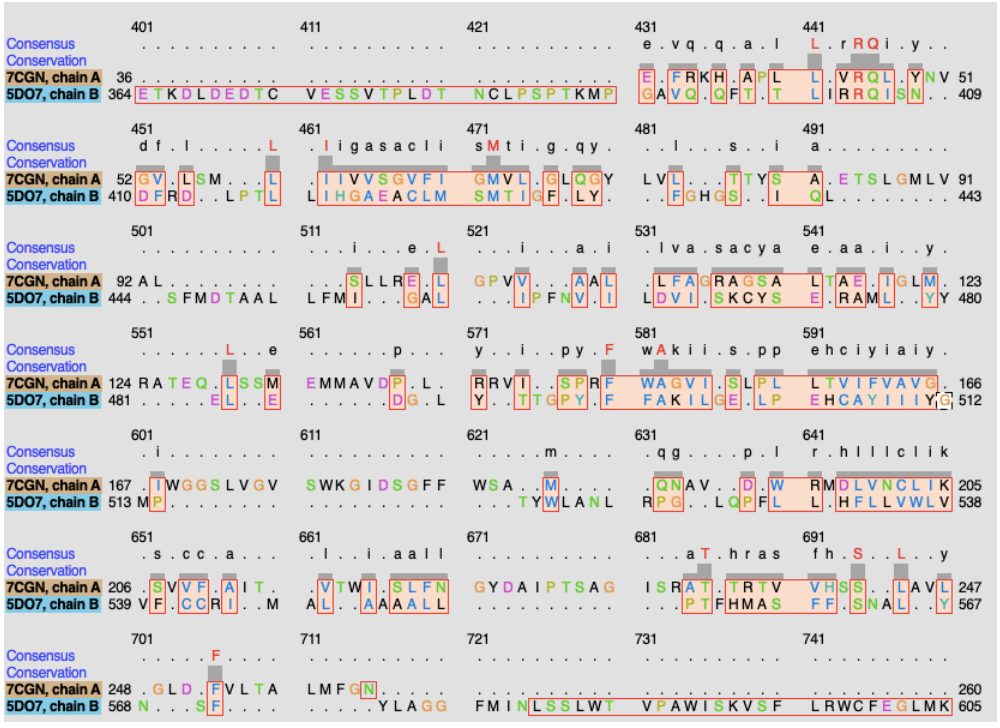
Appendix 44:The structural sequence alignment between ABCG2 NBD and MlaF

The structural sequence alignment between ABCG2 NBD and MlaF overall RMSD was2.034, SDM was 41.093, Q score of 0.420.



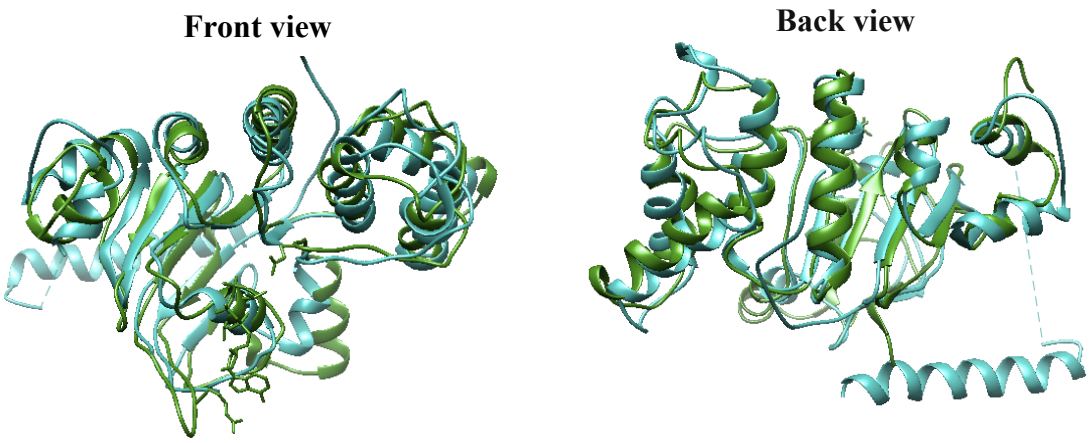
Appendix 45:The structural sequence alignment between ABCG8 TM1-5 and MlaE

The structural sequence alignment between ABCG8 TMD1-5 and MlaE overall RMSD was 2.585, SDM was 51.375, Q score of 0.325.



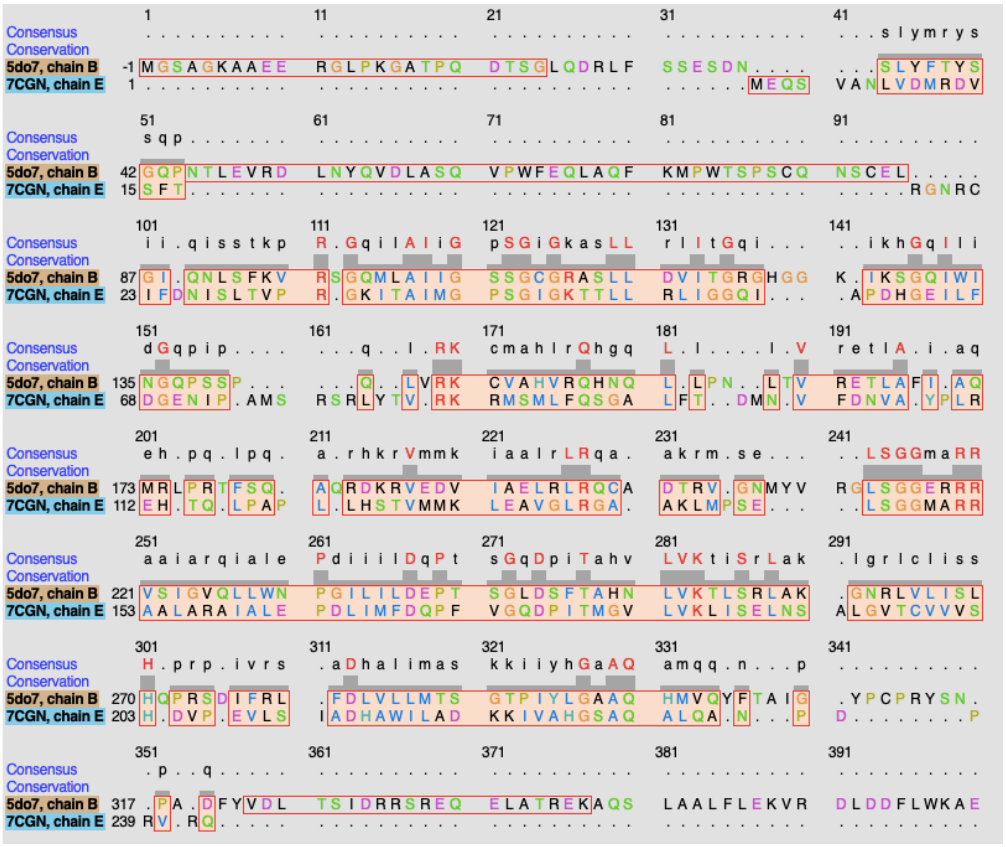
Appendix 46:The structural alignment between ABCG8 NBD and MlaF

The structure of MlaF is represented in light sea green. ABCG8 NBD represented in forest green. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.



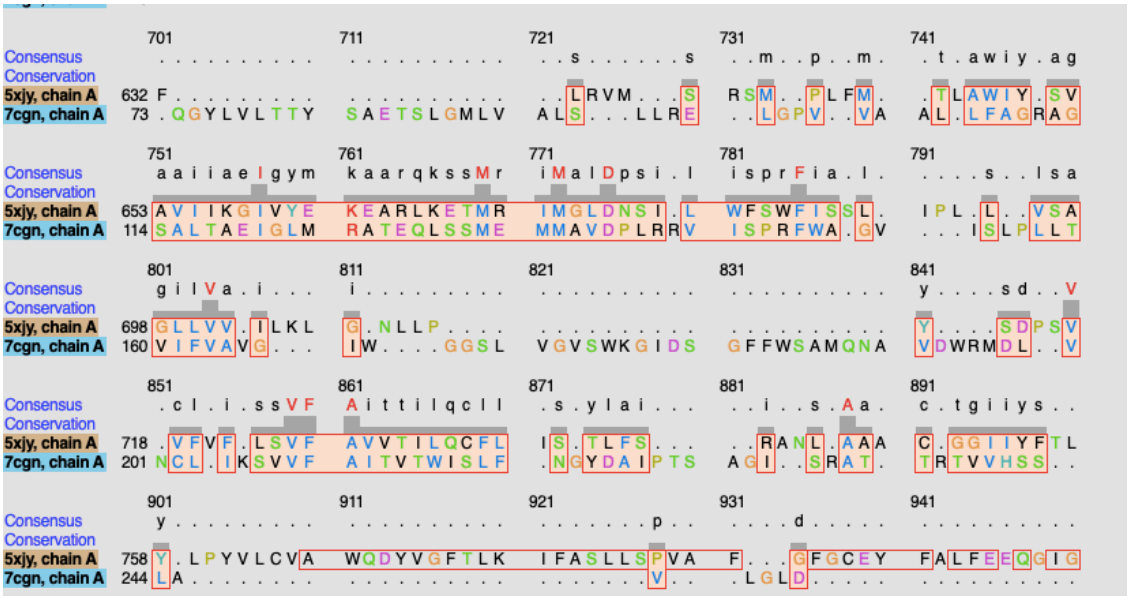
Appendix 47: The structural sequence alignment between ABCG8 NBD and MlaF

The structural sequence alignment between ABCG8 NBD and MlaF overall RMSD was 2.041, SDM was 42.235, Q score of 0.436.



Appendix 48:The structural sequence alignment between ABCA1 TM1-5 and MlaE

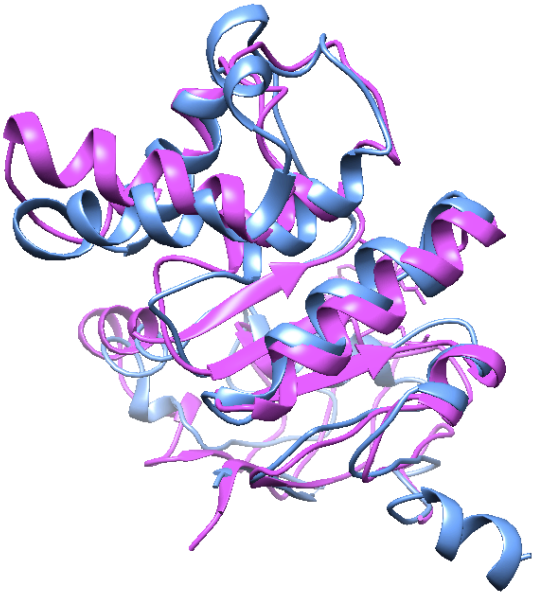
The structural sequence alignment between ABCA1 TMD1-5 and MlaE overall RMSD was 2.687, SDM was 65.460 and Q score was 0.268.



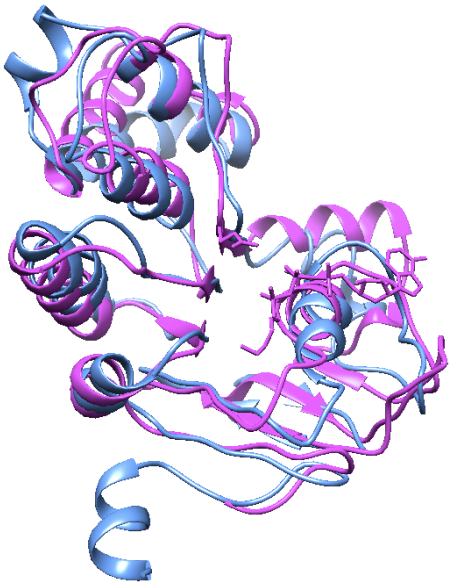
Appendix 49:The structural alignment between ABCA1 NBD and MlaF

The structure of MlaE is represented in magenta and ABCA1 is represented in cornflour blue ribbons. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front view

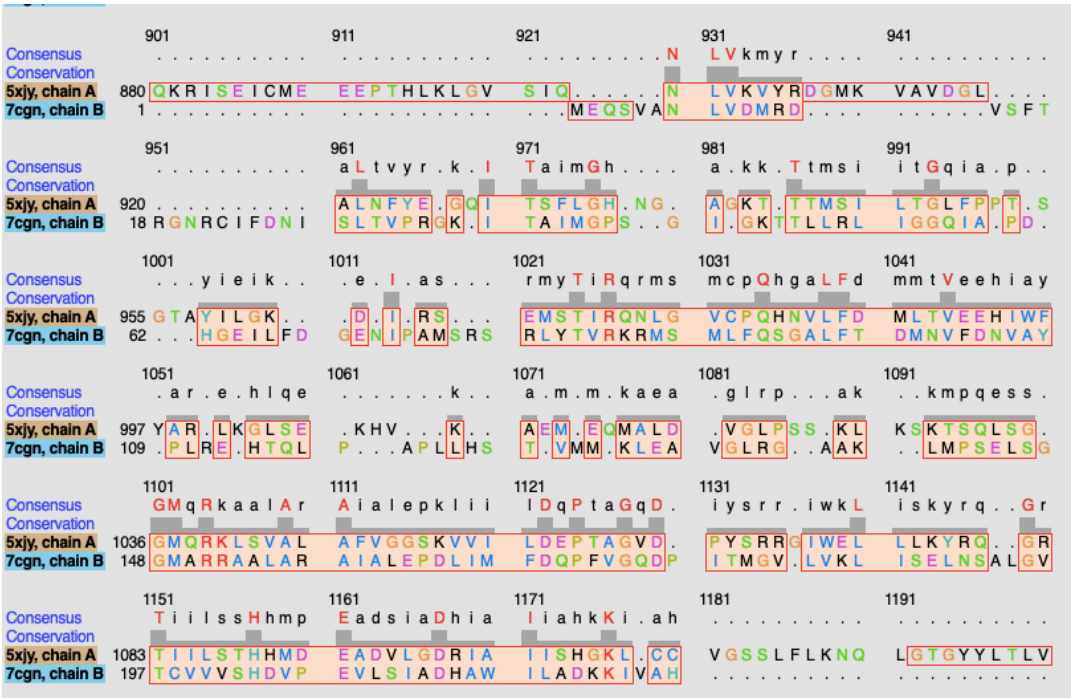


Overview



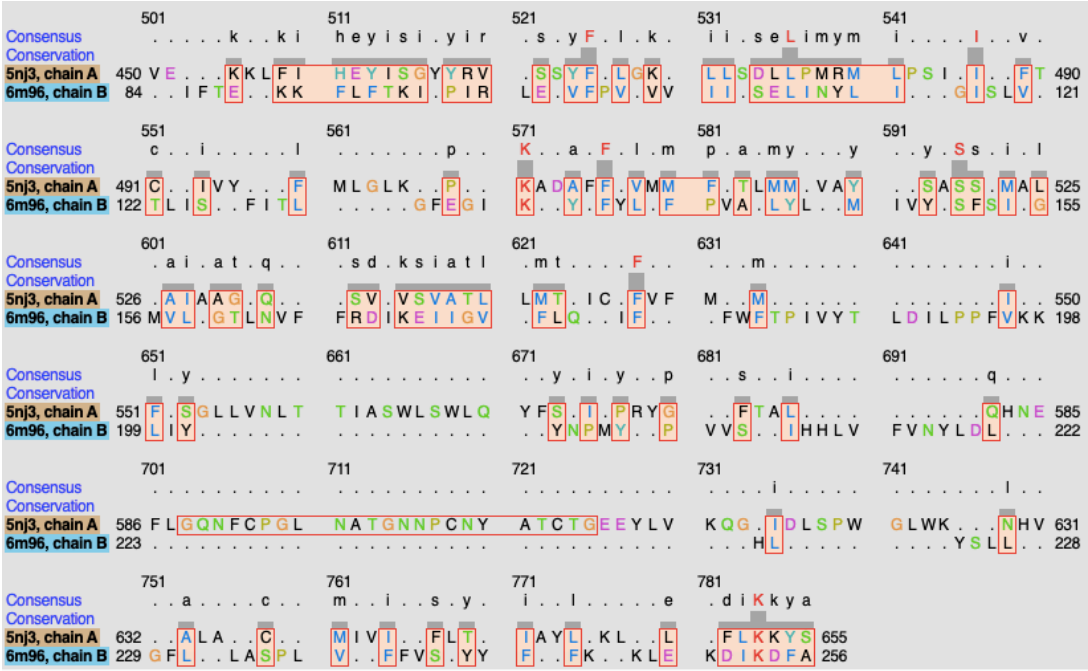
Appendix 50: The structural alignment of ABCA1 NBD and MlaF

The structural sequence alignment between ABCA1 NBD and MlaF; overall RMSD was 2.057, SDM was 41.434, Q score of 0.421.



Appendix 51: The structural sequence alignment between ABCG2 TMD and Wzm

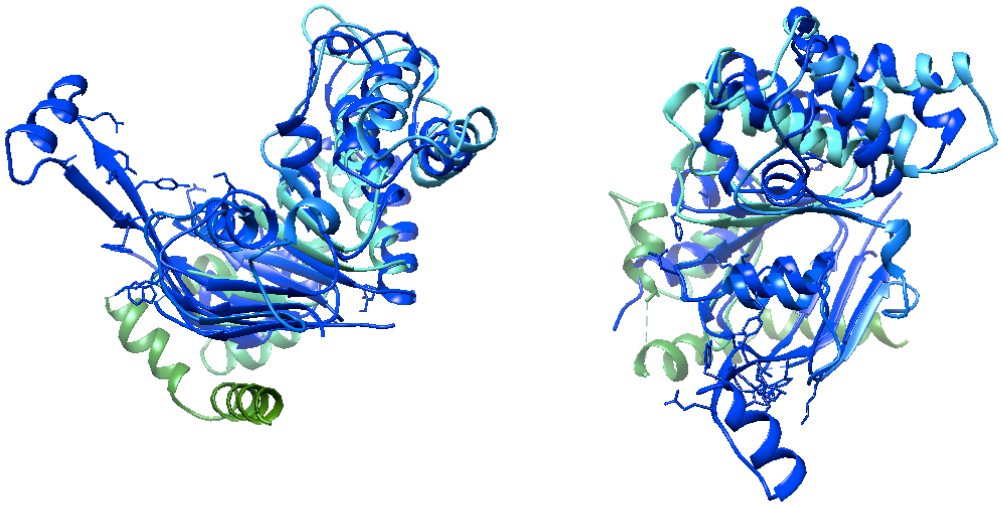
The structural sequence alignment between ABCG2 TMD and Wzm; overall RMSD was 2.438, SDM was 51.994, Q score of 0.390.



Appendix 52:The structural sequence alignment between ABCG2 NBD and Wzt

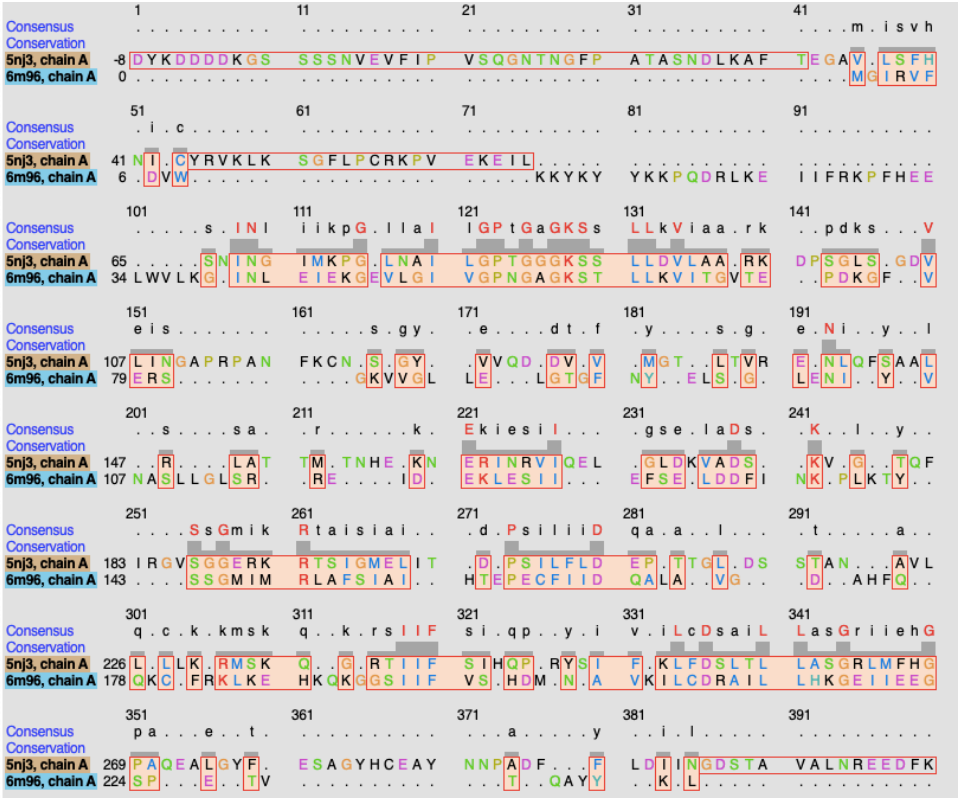
The structure of Wzt represented in blue and ABCG2 NBD represented green, cyan and blue ribbons. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.

Front view **Side view**



Appendix 53:The structural sequence alignment between ABCG2 NBD and Wzt

The structural sequence alignment between ABCG2 NBD and Wzt; overall RMSD was 2.435, SDM was 55.082, Q score of 0.231.

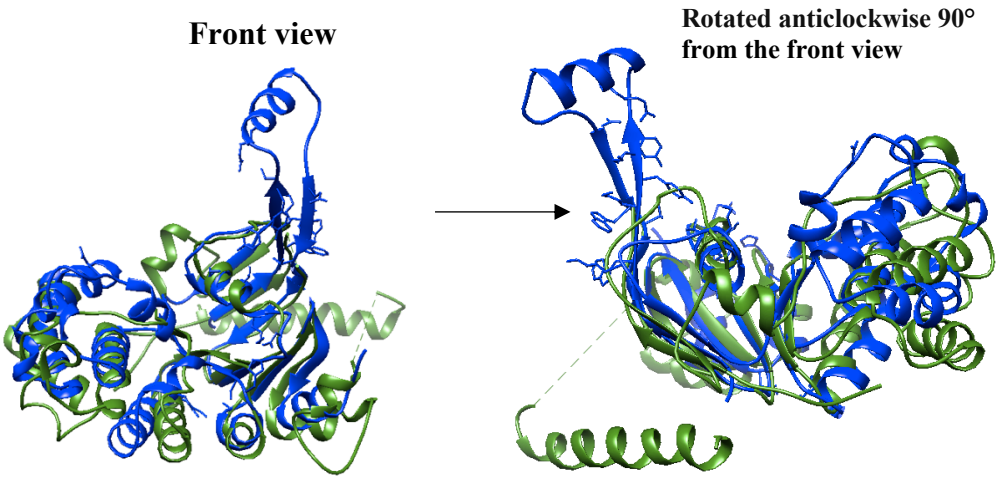


Appendix 54:The structural sequence alignment between ABCG8 TMD and Wzm

The structural sequence alignment between ABCG8 TMD and Wzm; overall RMSD was 2.643, SDM was 52.728, Q score of 0.230.

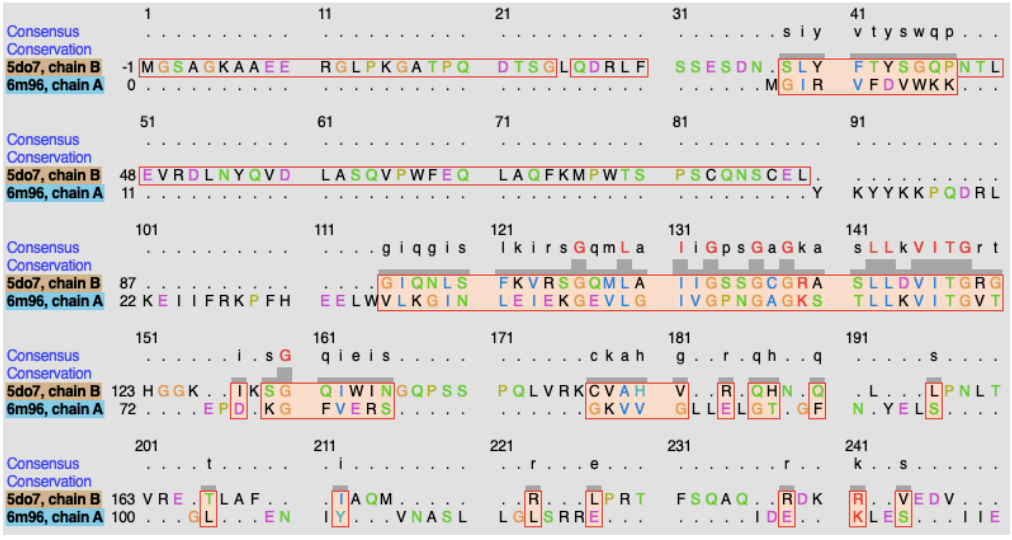
Appendix 55:The structural alignment between ABCG8 NBD and Wzt

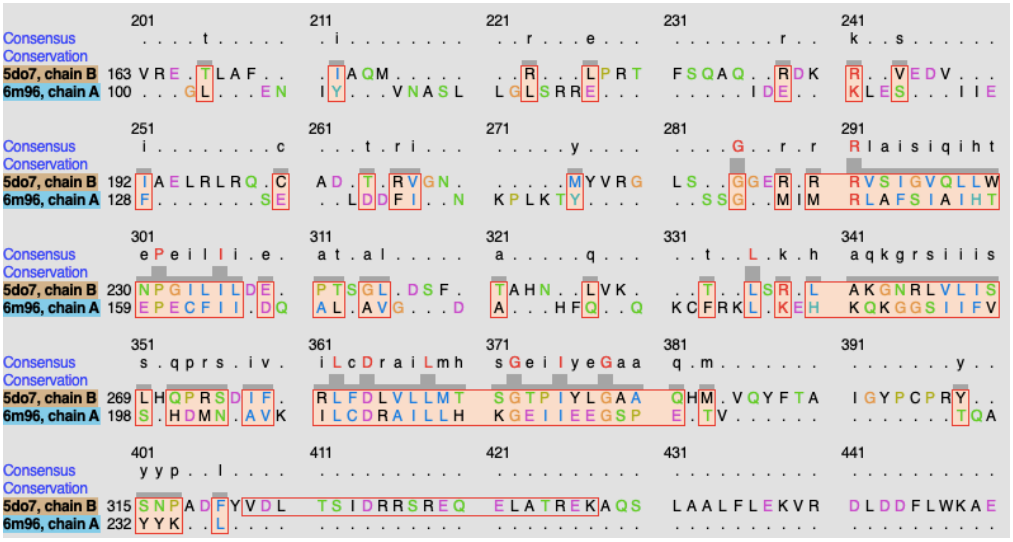
The structure of Wzm represented as blue ribbons in the alignment and ABCG8 is represented as forest green. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.



Appendix 56: The structural sequence alignment for ABCG8 NBD and Wzt

The structural sequence alignment between ABCG8 NBD and Wzt; overall RMSD was 2.161, SDM was 52.816, Q score of 0.213.



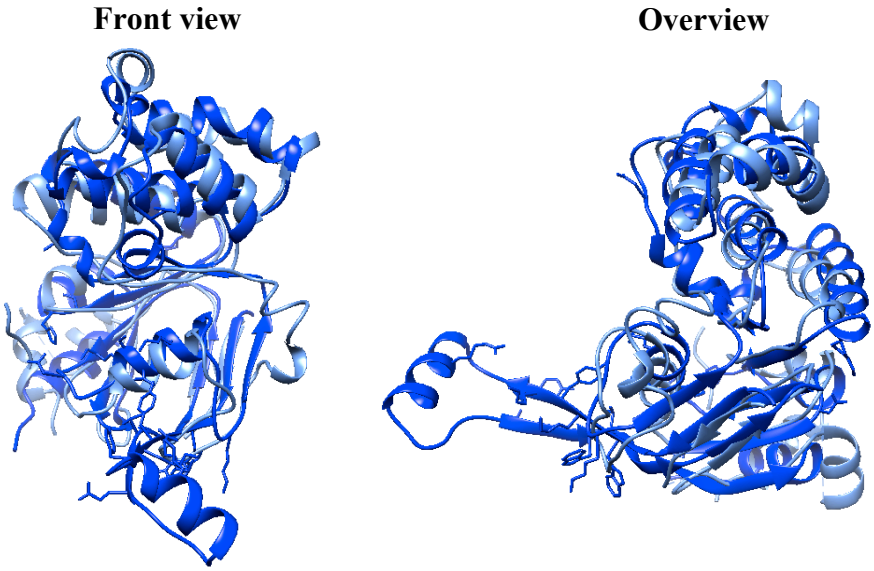


Appendix 57: The structural sequence alignment between ABCA1 TMD1 and Wzm

The structural sequence alignment between ABCA1 TMD1 and Wzm; overall RMSD was 2.389, SDM was 51.624, Q score of 0.326.

Appendix 58:The structural alignment between ABCA1 NBD1 and Wzm

The structure represented in dark blue is WzmWzt, structure ABCA1 is represented in cornflour blue. The structural alignments showed significant overlap of the α helices and β pleated sheets indicating high structural similarity.



Appendix 59: The structural sequence alignment between ABCA1 NBD1 and Wzt

The structural sequence alignment between ABCA1 NBD and Wzt; overall RMSD was 2.230, SDM was 56.342, Q score of 0.250.